social-biofeedback model proposed by Gergely and Watson (1996; 1999; Fonagy et al. 2002; Gergely & Unoka 2008). Currently, this model assumes that in repetitive episodes of (mostly) nonverbal communication (Csibra & Gergely 2006) mothers provide marked emotional "mirroring" displays which are highly (but inevitably imperfectly) contingent on the emotional displays of the infant. By doing so, mothers provide specific forms of biofeedback, allowing infants to parse their affective experience, form separate categories of their affective states, and form associations between these categories and their developing knowledge of the causal roles of emotions in other people's behaviour.

It is important to note that socio-constructivist theory is an essential complement to Carruthers' model 4, bridging a potentially fatal gap in his argument. People do *attribute* propositional emotional states to the self, and it seems reasonable to assume that their *actual* emotional states (propositional or not) play a role in generating such attributions. Carruthers' current proposal under-specifies how the mindreading system, which evolved for the purpose of interpreting others' behaviour, comes to be capable of interpreting primary somatic data specific to categories of affective states and of attributing them to the self. Furthermore, according to Carruthers, when the mindreading system does its standard job of third-person mental-state attribution, this sort of data "play little or no role" (target article, sect. 2, para. 8). Presumably, they can contribute, for example, by biasing the outcome of the mindreading processes (like when negative affect leads one to attribute malicious rather than friendly intentions). However, in first-person attributions, their function is quite different. They are the main source of input, providing the mindreading system with cues on the basis of which it can recognize current emotional attitude-states. The social-biofeedback model assumes that the mindreading system is *not readily* capable of doing this job and spells out the mechanism facilitating *development* of this ability. Putting it in terms of Carruthers' model 4: it explains how primary intra- and proprioceptive stimulation gains attentional focus to become globally accessible and how the mindreading system becomes able to win competition for these data.

Research on borderline personality disorder further illuminates the value of the socio-constructivist model (Fonagy & Bateman 2008). The primary deficit in borderline personality disorder (BPD) is often assumed to be a deficit in affect self-regulation (e.g., Linehan 1993; Schmideberg 1947; Siever et al. 2002). We have evidence of structural and functional deficits in brain areas of patients with BPD normally considered central in affect regulation (Putnam & Silk 2005). Accumulating empirical evidence suggests that patients with BPD have characteristic limitations in their self-reflective (metacognitive) capacities (Diamond et al. 2003; Fonagy et al. 1996; Levy et al. 2006) that compromise their ability to represent their own subjective experience (Fonagy & Bateman 2007). There is less evidence for a primary deficit of mindreading (Choi-Kain & Gunderson 2008). Evidence from longitudinal investigations suggests that neglect of a child's emotional responses (the absence of mirroring interactions) may be critical in the aetiology of BPD (Lyons-Ruth et al. 2005), more so even than frank maltreatment (Johnson et al. 2006). We think that the BPD model may become an important source of new data that could illuminate relationships between mindreading and self-awareness and their developmental antecedents. We suggest that children who experience adverse rearing conditions may be at risk of developing compromised second-order representations of self-states because they are not afforded the opportunity to create the necessary mappings between the emerging causal representations of emotional states in others and emerging distinct emotional self-states.

ACKNOWLEDGMENTS

The work of the authors was supported by a Marie Curie Research Training Network grant 35975 (DISCOS). We are grateful for the help and suggestions made by Liz Allison and Tarik Bel-Bahar.

Banishing "I" and "we" from accounts of metacognition

doi:10.1017/S0140525X09000661

Bryce Huebner[a,b] and Daniel C. Dennett[a]

[a]*Center for Cognitive Studies, Tufts University, Medford, MA 02155; and* [b]*Cognitive Evolution Laboratory, Harvard University, Cambridge, MA 02138.*
huebner@wjh.harvard.edu
http://www.wjh.harvard.edu/~huebner
daniel.dennett@tufts.edu
http://ase.tufts.edu/cogstud/incbios/dennettd/dennettd.htm

Abstract: Carruthers offers a promising model for how "we" know the propositional contents of "our" own minds. Unfortunately, in retaining talk of first-person access to mental states, his suggestions assume that a higher-order self is already "in the loop." We invite Carruthers to eliminate the first-person from his model and to develop a more thoroughly third-person model of metacognition.

Human beings habitually, effortlessly, and for the most part unconsciously represent one another *as persons*. Adopting this personal stance facilitates representing others as unified entities with (relatively) stable psychological dispositions and (relatively) coherent strategies for practical deliberation. While the personal stance is not necessary for every social interaction, it plays an important role in intuitive judgments about which entities count as objects of moral concern (Dennett 1978; Robbins & Jack 2006); indeed, recent data suggest that when psychological unity and practical coherence are called into question, this often leads to the removal of an entity from our moral community (Bloom 2005; Haslam 2006).

Human beings also reflexively represent themselves as persons through a process of self-narration operating over System 1 processes. However, in this context the personal stance has deleterious consequences for the scientific study of the mind. Specifically, the personal stance invites the assumption that every (properly functioning) human being is a *person* who has access to *her own* mental states. Admirably, Carruthers goes further than many philosophers in recognizing that the mind is a distributed computational structure; however, things become murky when he turns to the sort of access that we find in the case of metacognition.

At points, Carruthers notes that the "mindreading system has access to perceptual states" (sect. 2, para. 6), and with this in mind he claims that in "virtue of receiving globally broadcast perceptual states as input, the mindreading system should be capable of self-attributing those percepts in an 'encapsulated' way, without requiring any other input" (sect. 2, para. 4). Here, Carruthers offers a model of metacognition that relies exclusively on computations carried out by subpersonal mechanisms. However, Carruthers makes it equally clear that "*I* never have the sort of direct access that my mindreading system has to *my own* visual images and bodily feelings" (sect. 2, para. 8; emphasis added). Moreover, although "*we do* have introspective access to some forms of thinking . . . *we* don't have such access to any propositional *attitudes*" (sect. 7, para. 11; emphasis over "we" added). Finally, his discussion of split-brain patients makes it clear that Carruthers thinks that these data "force us to recognize that *sometimes* people's access to their own judgments and intentions can be interpretative" (sect. 3.1, para. 3, emphasis in original).

Carruthers, thus, relies on two conceptually distinct accounts of cognitive access to metarepresentations. First, he relies on an account of *subpersonal access*, according to which metacognitive representations are accessed by systems dedicated to belief fixation. Beliefs, in turn, are accessed by systems dedicated to the production of linguistic representations; which are accessed by systems dedicated to syntax, vocalization, sub-vocalization, and so on. Second, he relies on an account of *personal access*, according to which *I* have access to the metacognitive representations that allow me to interpret *myself* and form person-level beliefs about *my own* mental states.

The former view that treats the mind as a distributed computational system with no central controller seems to be integral to

Carruthers' (2009) current thinking about cognitive architecture. However, this insight seems not to have permeated Carruthers' thinking about metacognition. Unless the "I" can be laundered from this otherwise promising account of "self-knowledge," the assumption of personal access threatens to require an irreducible Cartesian *res cogitans* with access to computations carried out at the subpersonal level. With these considerations in mind, we offer what we see as a friendly suggestion: translate all the talk of personal access into subpersonal terms.

Of course, the failure to translate personal access into the idiom of subpersonal computations may be the result of the relatively rough sketch of the subpersonal mechanisms that are responsible for metarepresentation. No doubt, a complete account of metarepresentation would require an appeal to a more intricate set of mechanisms to explain how subpersonal mechanisms can construct "the self" that is represented by the personal stance (Metzinger 2004). As Carruthers notes, the mindreading system must contain a model of *what minds are* and of "the access that agents have to their own mental states" (sect. 3.2, para. 2). He also notes that the mindreading system is likely to treat minds as having direct introspective access to themselves, despite the fact that the mode of access is inherently interpretative (sect. 3.2). However, merely adding these details to the model is insufficient for avoiding the presumption that there must ("also") be *first*-person access to the outputs of metacognition. After all, even with a complete account of the subpersonal systems responsible for the production and comprehension of linguistic utterances, the fixation and updating of beliefs, and the construction and consumption of metarepresentations, it may still seem perfectly natural to ask, "But how do *I know my own mental states*?"

The banality that *I* have access to *my own* thoughts is a consequence of adopting the personal stance. However, at the subpersonal level it is possible to explain how various subsystems access representations without requiring an appeal to a centralized *res cogitans*. The key insight is that a module "dumbly, obsessively converts thoughts into linguistic form and vice versa" (Jackendoff 1996). Schematically, a conceptualized thought triggers the production of a linguistic representation that approximates the content of that thought, yielding a reflexive *blurt*. Such linguistic *blurts* are protospeech acts, issuing subpersonally, not yet from or by the person, and they are either sent to exogenous broadcast systems (where they become the raw material for personal speech acts), or are endogenously broadcast to language comprehension systems which feed directly to the mindreading system. Here, *blurts* are tested to see whether they should be uttered overtly, as the mindreading system accesses the content of the *blurt* and reflexively generates a belief that approximates the content of that *blurt*. Systems dedicated to belief fixation are then recruited, beliefs are updated, the *blurt* is accepted or rejected, and the process repeats. Proto-linguistic *blurts*, thus, dress System 1 outputs in mentalistic clothes, facilitating system-level metacognition.

Carruthers (2009) acknowledges that System 2 thinking is realized in the cyclical activity of reflexive System 1 subroutines. This allows for a model of metacognition that makes no appeal to a pre-existing *I*, a far more plausible account of self-knowledge in the absence of a *res cogitans*.

Unsymbolized thinking, sensory awareness, and mindreading

doi:10.1017/S0140525X09000673

Russell T. Hurlburt

Department of Psychology, University of Nevada, Las Vegas, Las Vegas, NV 89154-5030.

russ@unlv.nevada.edu

http://www.nevada.edu/~russ

Abstract: Carruthers views unsymbolized thinking as "purely propositional" and, therefore, as a potential threat to his mindreading-is-prior position. I argue that unsymbolized thinking may involve (non-symbolic) sensory aspects; it is therefore not purely propositional, and therefore poses no threat to mindreading-is-prior. Furthermore, Descriptive Experience Sampling lends empirical support to the view that access to our own propositional attitudes is interpretative, not introspective.

Section 8 of Carruthers' target article considers my Descriptive Experience Sampling (DES) work, particularly its finding of unsymbolized thinking (Hurlburt 1990; 1993; 1997; Hurlburt & Akhter 2008; Hurlburt & Heavey 2006). Carruthers implies that I characterize unsymbolized thinking as being purely propositional: "many subjects also report the presence of 'purely propositional,' unsymbolized thoughts at the moment of the beep" (sect. 8, para. 2). As a result, he supposes that my claim that unsymbolized thoughts are introspected (Hurlburt 1990; 1993) might present a difficulty for his mindreading-is-prior view, which holds that purely propositional events are not introspected but are, instead, interpreted.

Against this supposition, Carruthers argues that the introspection of unsymbolized thinking is an illusion; what is mistaken for introspection is a swift but unconscious interpretation of external events (Carruthers 1996b) and/or internal events such as images (present target article). As a result, he concludes in the target article that DES is neutral regarding Carruthers' mindreading view: "although there is no *support* to be derived for a 'mindreading is prior' account from the introspection-sampling data, neither is there, as yet, any evidence to count against it" (sect. 8, para. 5, emphasis in original).

I think Hurlburt and Akhter (2008) successfully rebutted Carruthers (1996b), and the target article does not change my mind. But I agree that unsymbolized thinking does not threaten Carruthers' mindreading-is-prior position, not because unsymbolized thinking is an unconscious interpretation but because it is not "purely propositional." Unsymbolized thinking is a directly apprehendable experience that may well have some kind of (probably subtle) sensory presentation, is therefore not purely propositional, and for that reason is not at odds with the mindreading-is-prior view.

In seeking to discover why Carruthers might hold, mistakenly, that I believe that unsymbolized thinking is "purely propositional," I reviewed what I have written on unsymbolized thinking and discovered this sentence:

> Unsymbolized Thinking is the experience of an inner process which is clearly a thought and which has a clear meaning, but which seems to take place without symbols of any kind, that is, *without* words, images, *bodily sensations*, etc. (Hurlburt 1993, p. 5; emphasis added)

"Without ... bodily sensations" might be understood to mean "purely propositional," but that is not at all what I intended. I should have written "without ... bodily *sensory awareness*" instead of "without ... bodily *sensations*."

"Sensory awareness" is a term of art in DES: "A sensory awareness is a sensory experience (itch, visual taking-in, hotness, pressure, hearing) that is in itself a primary theme or focus for the subject" (Hurlburt & Heavey 2006, p. 223). That is, sensory awareness is not merely a bodily or external sensation, but is a sensation that is itself a main thematic focus of experience. Thus, for example, Jack picks up a can of Coke, and, while preparing to drink, particularly notices the cold, slippery moistness against his fingertips. Jill picks up a can of Coke, and, while preparing to drink, says to herself in inner speech, "Carruthers is right!" Both Jack and Jill are having bodily sensations of the coldness, the moistness, and the slipperiness of the can (neither drops it). Jack's central focus is on the cold, slippery moistness; therefore, he is experiencing a sensory awareness as DES defines it. Jill's central focus is on her inner speech, not on the can; therefore she is *not* experiencing a sensory awareness as defined by DES (see Hurlburt & Heavey, in preparation).

Thus, unsymbolized thinking, as I and my DES colleagues describe the phenomenon, is an experience that is directly apprehended at the moment of the DES beep but which does not