

of the concept Ω when S_A was stimulated. (This point will be somewhere in the zone corresponding to A's very early learning of the concept, before the representation would be conscious. P&V explicitly concede that there is such a period.) At this point, we now have A relearn Ω and B learn Ω for the first time, employing exactly the same procedure originally used when A first learned Ω . Surely, P&V would agree that A would relearn the concept Ω faster than B because, as we have set things up, A will have a representational "head-start" over B. We thus have a very simple hypothetical case of how an unconscious representation could significantly affect the conscious experience of concept acquisition. Further, the SOC account, relying as it does only on conscious representations, would be at a loss in explaining this learning-time difference, unless they took the unfalsifiable position that A's more rapid learning of Ω simply demonstrated that the decayed representation with which A started prior to relearning Ω must, in fact, have been conscious all along.

It may well be that there is, indeed, some sort of "connectivity phase change" when a neural representation has the possibility of becoming conscious when activated. This could be the point described by Hebb as when "reverberation in the structure might be possible . . . reverberation which might frequently last for periods of time as great as half a second or a second, [this being] the best estimate one can make of the duration of a single 'conscious content'" (Hebb 1949, p. 74). But if one is to present a coherent picture of cognition that takes into account neural, representational, and cognitive phenomena, one must not neglect the representational stages leading up to this creation of cell-assemblies or, in the language of P&V, up to the emergence of fully conscious representations.

In conclusion, we suggest that the SOC model might do well to turn to basic neural network principles that would allow it, without difficulty, to encompass unconscious representations, as described above. (See, e.g., Cleeremans & Jiménez 2002; Mathis & Mozer 1996.) These "unconscious" representations – some of which may evolve into representations that, when activated, would be conscious – can affect consciousness processing, but do so via the same basic associative, excitatory, and inhibitory mechanisms that we observe in conscious representations. The inclusion of this type of representation in no way requires the authors to also posit sophisticated unconscious computational mechanisms.

Unconscious semantic access: A case against a hyperpowerful unconscious

Daniel Holender^a and Katia Duscherer^b

^aLaboratoire de Psychologie Expérimentale, Université Libre de Bruxelles 50, Brussels, Belgium; ^bLaboratoire de Psychologie Expérimentale, Université René Descartes (Paris V), Boulogne-Billancourt, France.
holender@ulb.ac.be kduscher@ulb.ac.be

Abstract: We analyze some of the recent evidence for unconscious semantic access stemming from tasks that, although based on a priming procedure, generate semantic congruity effects because of response competition, not semantic priming effects. We argue that such effects cannot occur without at least some glimpses of awareness about the identity and the meaning of a significant proportion of the primes.

Like Perruchet & Vinter (P&V), we fully endorse a mentalistic perspective, which implies that we do not posit the existence of a "powerful," or more precisely, an *intentional* cognitive unconscious. Thus, we basically share the view of Searle (1990; 1992) and Dulany (1997) that the unconscious is intentional in a dispositional way. In this commentary, we expand on the claim made by P&V in section 8.2 that the available data on unconscious semantic access do not constitute a challenge to the mentalistic framework.

In assessing the plausibility of the evidence for unconscious se-

matic access, a distinction must be made between tasks generating semantic priming effects and tasks generating other effects based on stimulus meaning, such as Stroop and Stroop-like congruity effects. This distinction has been somewhat blurred in recent work, maybe partly because of the multiple meanings of the term *priming*, which can designate an experimental procedure, an observed effect, and a hypothetical causal process, such as automatic spreading activation in semantic memory (e.g., Neely 1991). Much of the early evidence for unconscious semantic access under masking, criticized by Holender (1986), was based on a semantic priming paradigm yielding bona fide semantic priming effects. Much of the recent evidence for unconscious semantic access discussed by P&V does not qualify as priming because it rests on tasks that, although based on a priming procedure, are functionally equivalent to Stroop-like tasks. These tasks are generally assumed to generate congruity effects because of response competition (e.g., Eriksen 1995; Holender 1992; MacLeod 1991), not priming effects.

The studies of Greenwald et al. (1996; Draine & Greenwald 1998) are based on prime and target words with strong positive and negative affective connotations. The SOA between the prime and the target is very brief (under 100 msec), and the prime is interleaved between two masks consisting of random letters strings. Even though the primes could not be discriminated above chance, the binary classification of the target words in terms of their pleasantness is more accurate in congruent trials, in which the polarity of the prime and the target words are the same, than in incongruent trials, in which the polarities are opposite. Similarly, in the studies of Dehaene et al. (1998; Naccache & Dehaene 2001), which are based on a comparable procedure, the speed of classification of a single-digit target number in terms of whether it is larger or smaller than five is affected by the congruency of the unconscious prime number.

Initially, Greenwald et al. (1996; Draine & Greenwald 1998) interpreted their finding as reflecting semantic priming based on spreading activation. Then, Klinger et al. (2000) demonstrated that this effect does not depend at all on spreading activation but on response competition. This was taken as evidence that the unconscious primes must be covertly classified according to the same rule as the one applied to the visible target (see also Dehaene et al. 1998). Next, it was shown that the congruity effect only appears with primes that have been used repeatedly as targets (Abrams & Greenwald 2000; Damian 2001), which prompted a reinterpretation of the effects in terms of the formation through learning of a direct stimulus-response link based on superficial features of the stimuli. However, Abrams et al. (2002) argued that this link must rather be established between the stimuli and the semantic categories, as the learning effect resisted a change in response assignment. Nevertheless, Naccache and Dehaene (2001) persisted in their account in terms of unconscious semantic classification, because the congruity effect still occurs with unconscious primes, which have not been seen before as targets.

All these interpretations of unconscious congruity effects rest on the assumption that the primes are completely unavailable to awareness. If correct, they imply a hyperpowerful unconscious, that is, an unconscious even more powerful than the one already required to explain unconscious semantic priming effects. We contend that this conception is profoundly mistaken because, as was pointed out by Prinz (1997), a stimulus has no inherent information sufficient to specify a response outside the context of a goal-directed task imposed by the instructions. Besides, the primary source of response conflicts underlying the congruity effects described above must lie in conscious mental representations (cf. Holender 1992), because there is no stored information, and hence no information that can be automatically activated, about whether a number is smaller or larger than five or about whether the concept denoted by a word has a pleasant or unpleasant connotation. Therefore, the only possible source of conflict lies in the fact that most participants think about the irrelevant information in terms similar to those used by the instructions to describe how

the relevant information has to be related to the responses. Actually, we have evidence stemming from a Stroop-like task with visible distractors that participants not noticing the critical relation between the irrelevant information and the responses do not show any congruity effect (Duscherer & Holender, submitted).

To conclude, we submit that none of the congruity effects discussed above could occur without at least some glimpses of awareness about the identity and the meaning of a significant proportion of the primes. However, not all the apparent evidence of unconscious semantic access is flawed. There are also replicable effects that, although puzzling at first sight, can now be elegantly accommodated within the mentalistic framework. Let us take three examples. First, there is no need to postulate a center-surround attentional process taking place in semantic memory, to account for the reversal in the semantic priming effect (faster responses in unrelated than related trials) when participants make a semantic similarity judgment on masked primes in a preliminary task (Carr & Dagenbach 1990; Dagenbach et al. 1989). Kahan (2000) accounts for this effect by a retrospective prime clarification process stemming from participants' deliberate attempt to extract the meaning of the masked word, with this strategy being itself promoted by the preliminary task. Second, Duscherer and Holender (2002) explain the unconscious negative priming effects found by Allport et al. (1985) by the fact that a distractor potentially available to awareness can be made unconscious by the synergy between an act of selective inhibition taking place during prime processing and attentional diversion elicited by the onset of a mask coming too late to cause much perceptual degradation. Third, it is not necessary to postulate a powerful unconscious that can make affective preference judgments on stimuli that cannot be recognized (Zajonc 1980). This effect can simply result from applying a more effective nonanalytic strategy in preference judgments, and a less effective analytic strategy in recognition judgments (Whitelsea & Price 2001).

ACKNOWLEDGMENT

Katia Duscherer was supported by a postdoctoral grant from the Fyssen Foundation.

Consciousness and unconsciousness of logical reasoning errors in the human brain

Olivier Houdé

Groupe d'Imagerie Neurofonctionnelle, UMR 6095, CNRS, CEA, Université de Caen (Centre Cyceron) and Université Paris 5, Institut Universitaire de France, Sorbonne, 75005 Paris, France. olivier.houde@paris5.sorbonne.fr
www.cyceron.fr

Abstract: I challenge here the concept of SOC in regard to the question of the consciousness or unconsciousness of logical errors. My commentary offers support for the demonstration of how neuroimaging techniques might be used in the psychology of reasoning to test hypotheses about a potential hierarchy of levels of consciousness (and thus of partial unconsciousness) implemented in different brain networks.

Since Aristotle, we have known that the essence of the human mind is the *logos*, that is, both reason (logic) and language. But the seventeenth-century French philosopher Descartes (1628/1961) also showed with his method that an important challenge for humans is to implement deductive rules in order to redirect the mind from reasoning errors to logical thinking.

My commentary is specifically about section 7 of the target article, that is, about problem solving, decision making, and automaticity. Recent cognitive psychology and neuroimaging studies by my group have dealt with the mechanisms by means of which the human brain corrects initially unconscious logical reasoning errors (Houdé et al. 2000; 2001). They show that the activated brain networks are different, depending on whether (1) subjects

think they are responding correctly to a logic problem when in fact their reasoning is biased by an erroneous perceptual strategy (an automatic strategy, in accordance with Evans's (1989) model); or (2) they become aware of their error and correct it after being trained to inhibit the initial perceptual strategy. In the second stage (after training), regions in the left lateral prefrontal cortex devoted to executive functions, inner speech, and deductive logic are activated, along with a right ventromedial prefrontal area dedicated to self-feeling and relationships between emotions and reasoning (see Damasio's theory on consciousness; Damasio 1999).¹ None of these regions are implicated in the first stage (before training), where the only activation observed is in a posterior network strongly anchored in perception (ventral and dorsal pathways).

Interestingly, from the famous case of Phineas Gage in the nineteenth century (Damasio et al. 1994; Harlow 1848) to Damasio's patients today (Damasio 1994; 1999), neuropsychological findings clearly indicate that right ventromedial prefrontal damage is consistently associated with impairments of reasoning/decision making, emotion, and self-feeling. For the first time, our neuroimaging results demonstrate the direct involvement, in neurologically intact subjects, of a right ventromedial prefrontal area in the making of logical consciousness, that is, in what puts the mind on "the logical track," where it can implement the instruments of deduction. (Note that this brain area was not activated in a group of subjects who were unable to inhibit the initial perceptual strategy and therefore could not avoid the reasoning error; see Houdé et al. 2001.) Hence, the right ventromedial prefrontal cortex may be the emotional component of the brain's error correction device. More exactly, this area may correspond to the self-feeling device that detects the conditions under which logical reasoning errors are likely to occur (in connection with the anterior cingulate cortex; see Bush et al. 2000).

From the standpoint of evolutionist psychology (Bjorklund 1997; Tooby & Cosmides 2000), it is interesting to relate these neuroimaging results to the role classically ascribed to emotions in survival, namely, that in the face of danger (here, logical reasoning errors), fear leads animals and thus humans to flee, to avoid. In Darwinian terms, we can contend that evolution must have fashioned the brain to sense the emotions needed to inhibit nonadaptive behavior, even nonadaptive reasoning strategies (Houdé et al. 2000).

The findings of these studies (Houdé et al. 2000; 2001) allow me to challenge Perruchet and Vinter's concept of self-organizing consciousness (SOC) in regard to three interrelated points: (1) the existence – denied by the authors – of a hierarchy of levels of consciousness (and thus of partial unconsciousness) implemented in different brain networks (see Stage 1 and Stage 2 in our training paradigm below); (2) the important problem in adults, but also in children (Houdé 2000; Piaget 1984), of the consciousness or unconsciousness of reasoning errors (an issue the authors do not really address in the article); and (3) the question of the involvement of emotions and self-feelings in logical reasoning, if and when the subject becomes aware of the fact that there are several ways to solve a problem (and that one of them is based on a misleading strategy).

NOTE

1. *The Feeling of What Happens: Body and Emotions in the Making of Consciousness*. I am surprised to find no references to Damasio's work, and in particular, with regard to the issue of decision making, no mention of the 1997 article in *Science* that he co-wrote (Bechara et al. 1997).