

**Fishing with the wrong nets:  
How the implicit slips through  
the Representational Theory of Mind**

**Luis Jiménez**

Department of Psychology  
Universidad de Santiago  
15706 Santiago  
SPAIN

Electronic Mail: [jimenez@usc.es](mailto:jimenez@usc.es)

WWW: <http://web.usc.es/~psljim/englishpage.html>

**Axel Cleeremans**

Cognitive Science Research Unit  
Université Libre de Bruxelles CP 122  
Avenue F.-D. Roosevelt, 50  
1050 Bruxelles  
BELGIUM

Electronic Mail: [axcleer@ulb.ac.be](mailto:axcleer@ulb.ac.be)

WWW: <http://srsc.ulb.ac.be/axcWWW/axc.html>

## Abstract

Dienes and Perner's theory is not a satisfactory theory of implicit knowledge because by endorsing the representational theory of knowledge, the authors also inadvertently accept that only explicit knowledge can be causally efficacious, and hence that implicit knowledge is an inert category. This conflation between causal efficacy, knowledge, and explicitness is made clear through the authors' strategy, which consists of attributing any observable effect to the existence of representations that are as minimally explicit as needed to account for behavior. In contrast, we believe that causally efficacious and fully implicit knowledge exists, and is best embodied in frameworks that depart radically from classical assumptions.

The goal pursued by Dienes and Perner (henceforth, D&P) in this target article is an ambitious one, as they aim to build a theory of implicit and explicit knowledge that would enable cognitive scientists to distinguish among the diversity of senses in which one can consider knowledge to be “implicitly” held by a cognitive system. This kind of conceptual effort is commendable in and of itself given that common-sense and technical terms often get blended in maturing disciplines such as cognitive science. D&P’s proposal is therefore to be welcomed, even if, as we will show through our commentary, their effort tends to be conceptually misguided. To put it simply, we believe that D&P have dropped a conceptual net that is ill-suited to bring back the intended fishing.

To clarify the argument, consider D&P’s strategy — their “conceptual net”. This conceptual net, as described by the authors, consists of analyzing knowledge as a “propositional attitude according to the representational theory of mind” (p. 4), a theory that assumes that a representation “constitutes knowledge if it is put in a [...] knowledge box or [...] data base” (p.6). In this context, D&P propose a functional distinction between explicit and implicit knowledge, according to which the content of any knowledge (i.e., the content of the representations in the knowledge box) is explicit if it is a function of that representation to indicate precisely that content. In contrast, information is implicit if it is only conveyed as an indirect consequence, or supporting fact, of what is explicitly represented. The authors offer this functional criterion as a way of independently distinguishing between different types of knowledge that shape human cognition. However, as the discussion provided through the rest of the target article clearly illustrates, such a functional criterion is not independent from the criterion of knowledge efficacy, and hence results in a disturbing conflation between causal efficacy, knowledge, and explicitness.

This conflation seems unavoidable from the authors's conceptual approach, given that the only external way to ascertain the function of some knowledge is to ascertain its cognitive and behavioral effects. Thus, by relying on this criterion, the authors have no choice but to consider that all the knowledge that produces observable effects must necessarily count as explicit knowledge at the specific level that is minimally needed to account for the observed effects. The application of this logic is widely illustrated through the target article, and goes roughly like this: If you observe an effect that appears to depend on the presence of some knowledge, label this knowledge as explicit at that particular level, and let any other knowledge that appears to have no bearing on the observed behavior be implicit. If the efficacious knowledge includes properties of the stimulus, then call it “property” explicit; if it includes, for instance, the holder of that knowledge, then call it “self” explicit.

Clearly, the problem with this strategy is that by labeling as implicit only the knowledge that has no bearing on a specific situation, it becomes somewhat odd to even consider it as “knowledge”, no matter how tightly related to explicit knowledge this implicit knowledge might appear to be from an external point of view. Thus, to paraphrase a sentence that D&P apply to the criterion of accessibility (p.19), we might say that if implicit knowledge is not causally efficacious, then it would not qualify as knowledge and, in any case, there would be no evidence that there is any implicit knowledge at all. Implicit knowledge, then, in this framework, is the name for a non-entity — a fish that has slipped through the authors’ conceptual net.

How did the fish slip through the net? We surmise that the problem lies in the author's adoption of the "representational theory of mind". According to this framework, representations constitute knowledge if they appear in a "knowledge box". Importantly, whether a particular representation enters the knowledge box or not is defined by whether "the representation is used as a reflection of the state of the world and not, e.g., if it were in a goal box, as a typically non-existing but desirable state of the world"

(p. 6). Now, the only way for an outside observer to ascertain whether a particular representation that an agent possesses is in its knowledge box (i.e., is "used as a reflection of the state of the world") or not is to examine whether it influences the agent's performance on some task. But note that this reasoning is exactly identical with that entailed by the criterion used by D&P to ascertain whether knowledge is explicit at any given level. In other words, we seem to be caught in a maze of twisty little conceptual corridors that all point to the same conclusions: In D&P's framework, (1) a representation can only constitute the agent's knowledge if it is in the agent's knowledge box, (2) a representation can only influence performance if it is in the agent's knowledge box, and (3) any representation that is in the knowledge box is necessarily explicit in at least the specific way needed to account for observable behavior. It should be clear this conceptual net has shark-sized holes in it.

Where do the holes come from? The problem, we surmise, is with the knowledge box. The notion that one can account for the way in which our representational systems are organized by assuming that representations are tokened and put in databases of different kinds is simply inadequate to capture the dynamics of cognition. What is the alternative? We suggest to eliminate the "knowledge box" as a requirement for the definition of knowledge, and to assume that representations can simultaneously constitute knowledge and be causally efficacious without ever being tokened in any way. For instance, observing that "butter" has been perceived in a subliminal perception experiment because it exerts detectable effects on performance does not imply that the property of "butter" has been somehow unconsciously represented in the subject's knowledge box (as D&P strangely suggest on p. 39) or, worse, that it has been represented in some unconscious zombie-like twin of the knowledge box. It simply means that the relevant neural pathways were activated sufficiently to bias further processing in the relevant direction when the stem completion or lexical decision task is actually performed. The knowledge embedded in such pathways is knowledge that is simultaneously causally efficacious and fully implicit. It does not produce any kind of

conscious or unconscious “attitude” and hence cannot be accounted for based on the representational theory of mind. Clearly, such knowledge is better captured through dynamical approaches such as the connectionist framework (see Cleeremans, 1997; Cleeremans & Jiménez, submitted; O’Brien & Opie, in press, Mathis & Mozer, 1996) — a perspective with which D&P otherwise agree. A particularly important and difficult issue in this context is to chart the divide between processes and representations, but this is a matter for further debate.

## Acknowledgments

Axel Cleeremans is a Research Associate of the National Fund for Scientific Research (Belgium). This work was also supported by Grant XUGA21106B98 from the Xunta de Galicia to Luis Jiménez.

## References

- Cleeremans, A. (1997). Principles for implicit learning. In: *How implicit is implicit learning?*, ed. D.C. Berry. Oxford University Press.
- Cleeremans, A. & Jiménez, L. (submitted). Implicit cognition with the symbolic metaphor of mind: Theoretical and methodological issues.
- Mathis, W.D. & Mozer, M. C. (1996) Conscious and unconscious perception: A computational Theory. In: *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pp. 324-328, Erlbaum
- O'Brien, G. & Opie, J. (in press) A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences*