When knowledge is unconscious because of conscious knowledge and vice versa

Zoltan Dienes (dienes@biols.susx.ac.uk)
Experimental Psychology, Sussex University
Brighton BN1 9QG England

Josef Perner (Josef.Perner@sbg.ac.at)
Institut fuer Psychologie, Universitaet Salzburg
A-5020 Salzburg Austria

Abstract

This paper will offer a framework and a methodology for determining whether subjects have conscious or unconscious knowledge. The implicit-explicit distinction will be related to consciousness using the framework of Dienes & Perner (1999; 2001a,b,c) and the higher-order thought theory of Rosenthal (1986, 2000). Whether a mental state is conscious or not depends on whether certain inferences are unconscious or not, in a way we will specify; this is the interaction between implicit and explicit knowledge we will consider. The arguments will be illustrated with the artificial grammar learning paradigm from the implicit learning literature.

Introduction

In this paper we will argue that there is an intimate and generally unappreciated interaction between implicit and explicit knowledge that occurs all the time. Consideration of this interaction is important in determining whether a subject possesses conscious or unconscious states of knowledge. To make the argument, in the first section below we will overview the framework of Dienes and Perner (1999, 2001a,b,c) for understanding the implicit-explicit distinction in terms of the properties of representations. We will take an everyday use of the implicit-explicit distinction and apply it in a particular way to what it is to represent something. Given a representational theory of knowledge, this produces a hierarchy of ways in which knowledge can be implicit or explicit. We will then use the higher-order thought theory of Rosenthal (1986, 2000) to show full explicitness is almost the requirement for mental states of knowing to be conscious. There is one further stipulation above and beyond full-explicitness needed for consciousness and it is this that shows the importance of an interaction between implicit and explicit knowledge in producing conscious or unconscious states. We will discuss this relationship and illustrate how subjects' knowledge of an artificial grammar could be shown to be conscious or unconscious (in fact we will argue that the evidence shows that subjects can acquire fully unconscious knowledge).

Implicitly vs Explicitly Representing

Dienes and Perner's (1999, 2001a,b,c) framework could be structured as semi-independent modules: a notion of representation, a notion of the implicit-explicit distinction, the hierarchy of implicitness, and a theory of consciousness. To a degree, one can reject one of the modules and still accept the others to build an understanding of implicit knowledge and consciousness. We begin first with the notion of representation: In order to be clear how one might explicitly or implicitly represent something we need to be clear about what it is to represent something. In this we follow the functional theories of representation. For example, according to Millikan (1984, 1993), there must be a producer of the representation that has as its function that it brings about a mapping from the representation to a state of affairs. For example, in a bee dance, the bee can produce a dance such that the angle of the dance maps onto the location of the nectar. Further there will be consumers of the representation that perform various functions as they react to it. But they can only perform their functions under normal conditions if the representation does indeed map onto a certain state of affairs: This state of affairs is the content of the representation. This is what we mean by a representation. On this account, representations do not need to have further properties (e.g. compositional semantics) to be representations. The weights of a connectionist network are representations: They must map onto statistical regularities in the world in a certain way for the consumers of these weights to perform their functions, so the weights represent statistical regularities. What is it for a representation to represent something implicitly or explicitly, and what makes some representations conscious (those that have the contents that are the contents of our consciousness) while other representations are unconscious?

A bee dance represents the location of nectar. We say it represents location explicitly, because variations in the representational medium (angle of dance) map onto variations in location. However, it does not explicitly represent that it is about nectar: There is nothing in the medium that varies with whether it is nectar that it is about or something else. We say it represents the fact that it is about nectar only implicitly. Just so, in everyday terms when one answers the question "what is this?" to a succession of animals, and responds with the statement "cat" (or "dog" etc), the statement explicitly represents the property of being a cat, because variations in the representational medium (words) map precisely into variations in this content. The statement implies that it is this that is a cat, but it does not say so explicitly: There is not a part of the medium that varies directly with this rather than that being a cat.

Now consider what it is to have knowledge. In general knowledge consists of a proposition ("this word has the meaning butter") towards which you have an attitude of knowing. One can know without making all of these components of knowledge explicit. One can represent the proposition explicitly but not the fact that it is knowledge. It can in fact be knowledge (because it is taken as true and acted upon) without there being a representation that goes into one state for it being knowledge and another state if it is not knowledge. Minimally one could just make explicit the property without making explicit the individual that has this property (compare the bee dance). In subliminal perception we argue that it is indeed just a property of a presented word that is represented explicitly e.g. having the meaning "butter". There is no representation with the content "I see that the word in front of me has the meaning butter" so this cannot be the content of any experience of the subject; but the representation of merely "butter" can allow the subject to e.g. say "butter" as the first dairy product that comes to mind. At the next stage, the full proposition is made explicit ("The word in front of me has the meaning butter"). This stage involves the binding of features to individuals. At the next stage the factuality or otherwise of the proposition is made explicit ("it is a fact that the word in front of me has the meaning butter"). This is precisely the developmental milestone that occurs in a child's representational capacity at about 18 months (Perner, 1991), and is needed for appreciating hypotheticals, changing temporal states of affairs (and hence is necessary for explicit memory), and counter-factual reasoning. At the final stage the propositional attitude by which one holds the proposition is made explicit ("I see that the word in front of me has the meaning butter"). This is full self and attitude explicitness. We argue that this is necessary for knowledge to be conscious knowledge. This link from full explicitness to consciousness is made via the higher order thought theory of consciousness (Rosenthal, 1986, 2000; Carruthers, 1992, 2000).

The Higher-Order Thought Theory

Rosenthal (1986, 2000) develops an account of when a mental state is a conscious mental state. He argues that when one is in a conscious mental state one is conscious of that mental state. It would be inconceivable to claim that one is in a conscious state of, for example, seeing the word butter, while at the same time denying being conscious of seeing the word butter. So the question is, how does one become conscious of mental states? The relevant way, Rosenthal argues, is to think about them. We become conscious of our seeing the word butter when we think that we are seeing the word butter. That is, when we are consciously seeing the word butter, we have a thought like "I see that the word is butter". Because this thought (this mental state) is about another mental state (seeing), it is called a higher order thought. Note that this higher order thought is just our requirement for knowledge to be fully explicit: There is a natural relationship between explicitness and consciousness.

A Method For Determining Unconscious States

These considerations show that knowledge states being conscious or not is essentially a metacognitive issue (Dienes & Perner, 2001b). Roughly, simply knowing something and hence being able to respond discriminatively does not make the knowing a conscious mental state; for the latter, one must know that one knows. Metacognition has both a monitoring and a control aspect, and both of these aspects can be used to form methodologies for determining the conscious status of knowledge via an analysis of the relationship of different types of control and monitoring to the hierarchy of implicitness (Dienes & Perner, 2001b). Here we will focus exclusively on monitoring; the criterion for a state being conscious or unconscious is essentially that of the subjective threshold in the subliminal perception literature (Cheesman & Merikle, 1984).

Consider a subject in an artificial grammar learning experiment (Reber, 1967, 1989). The subject is exposed to strings of letters generated by a finite state grammar and asked to memorize them. After some minutes exposure, the subject is told actually there was a set of complex rules that determined the order of letters within the strings, and could they now classify a new set of strings as obeying the rules or not. Reber found that subjects could do so above chance but they found it difficult to say what the rules were. How could we

determine whether their knowledge is actually unconscious?

When subjects classify a test string they bring to bear their knowledge of the grammar to produce a new piece of knowledge: Whether this string is grammatical or not. We must distinguish these different knowledge contents: knowledge of the grammar, and knowledge of a particular string being grammatical (the grammaticality judgement).

When subjects make grammaticality judgements to the same strings several times they respond with different degrees of consistency to different strings (Reber, 1989; Dienes, Kurz, Bernhaupt, & Perner, 1997). For some strings the subject responds highly consistently, for others the subject may give a "grammatical" or "non-grammatical" response with 50% probability. Our interpretation of this fact is that subjects are in different knowledge states about the different test strings. Regardless of whether they have induced the same grammar as the experimenter or not (in fact, their grammar is correlated with the experimenter's grammar), the subjects themselves are treating themselves as being in different knowledge states about different strings. But have they conceptualized themselves as being in those different knowledge states? That is, have they formed attitude-explicit representations - higher-order thoughts - about those states? (Note that it was important to establish that there were different knowledge states before this question could be asked.)

When confidence ratings are taken after each classification decision, subjects can classify at above chance rates even when they claim they are literally guessing (for a review see Dienes & Berry, 1997). Further, under some conditions, there will be no within-subject relationship between confidence and accuracy: Subject do not know about the different knowledge states they are in fact in (Dienes & Berry, 1997; see Dienes & Perner, 2001c for this finding with a context-free grammar). Their knowledge is attitude-implicit, and hence unconscious.

The Crucial Interaction Between Conscious and Unconscious Knowledge

Consider now two objections one may have to assessing unconscious knowledge with confidence ratings. First, Allwood, Granhag, Johansson (in press) found the normal evidence for attitude-implicit knowledge in an artificial grammar learning experiment when the typical amount of learning and testing was used. In a second experiment that involved greater exposure to strings at learning and test subjects' confidence and accuracy was well-calibrated and so the knowledge seemed entirely atittude-explicit, despite the authors' feeling that the fundamental nature of subjects' knowledge had not changed. Allwood et al suggest that confidence ratings can come to be based on implicit knowledge as much as explicit knowledge, and this possibility undermines the usefulness of a dissociation between confidence and accuracy as a measure of implicit knowledge. Second, there is the standard objection to subjective measures of consciousness: Are they not dependent on the vicissitudes of subjects' idiosyncratic theories of consciousness, knowledge, etc (e.g. Shanks & St John, 1994)? This second objection is taken up in Dienes & Perner (2001a) and especially Twyman and Dienes (2001); we will develop a different, complementary response in this paper. To address these objections we need to return to a subtlety of the higher-order thought theory that we glossed over.

Rosenthal argues that to make a mental state conscious, the higher order thought must assert that one is in the state and it must not arise from any inferences of which we are conscious. That final stipulation is crucial. If I am driving along and swerve, and wonder why I swerved, I might think "I must have swerved because I saw that truck". That thought is a higher order thought about one being in a mental state of seeing. But it does not make the original seeing conscious; it does not make it conscious precisely because it arose from an inference of which one was conscious.

A mental state is conscious if we are non-inferentially conscious of it; the mental state would be unconscious if we were conscious of the mental state only by virtue of inferences of which we are conscious. This is just why blindsight patients' seeing is still unconscious even though they may consciously infer they must be seeing "because the experimenter tells me I am consistently correct." They have a higher-order thought that they are seeing, but our intuitions are that the seeing remains unconscious; and it remains unconscious precisely because the higher-order thought (the attitude explicit representation) arose from an inference of which the patient is conscious.

How do these considerations apply to determining whether a subject in an experiment has acquired conscious or unconscious knowledge? One has to be careful when the mental state is knowing, because knowing need not refer to an occurrent mental state at all. Knowing is often used in a dispositional sense: if you are asked if you know your times tables, that does not mean to say you are actively thinking about them now; the question is just whether you could do so accurately if asked. We will see how this can lead to a paradox.

Imagine a person is asked a general knowledge question and they believe they know the answer. The person is asked "Why do you conclude that you know

that?" When a person consciously knows something they might be able to provide conscious inferences by means of which they conclude that they know. They might justify their knowledge as knowledge because e.g. " I can describe the reasoning by which I drew the conclusion"; "I remember the event in which a trusted authority told me the knowledge"; or, more generally, "I can consciously link the knowing to some conscious perception."

These conscious inferences do not make the knowledge unconscious. (Further, they are not essential for the knowledge being conscious either: I might not know why I know something, I just insist that I know it.) This seems to go against the conclusion that mental states are only conscious if one knows about them by inferences of which one is unconscious.

The problem arises because knowing is not an occurrent mental state. An occurrent mental state associated with knowing is "thinking with conviction" or "thinking with a certain degree of conviction". Even if a person is aware of the inferences by which they know something, they just directly know that "I am thinking with conviction" if the thinking is a conscious state. In answer to the question "How do you know you are thinking with conviction?" one does not need to list the inferences that justify the knowledge as knowledge; they are not inferences leading to the conclusion that one is thinking with conviction. One just directly knows that one is thinking with conviction if the thinking-with-conviction is a conscious mental state. It is a conscious state because the inferences, if any, by which one ascertained that one was in the state were unconscious.

In answer to the question "Why do you conclude that you know that?" a person might provide conscious inferences by means of which they conclude they know something by observing their behaviour: "I respond consistently, quickly or effectively". For example, a person may select the correct capital of a country, let's presume, due to being in an unconscious state of thinking-with-conviction. This state makes the person respond consistently and quickly; the state of thinking-without-conviction (let's presume) makes the person respond inconsistently and slowly. The person does not know he is in a state of thinking with conviction at first; but he consciously infers from the speed with which the answer came to him that he must have been in an occurrent state of knowing. Because he is conscious of this inference, the state is unconscious. If the same inference had been drawn for the same reason but unconsciously, his thinking-with-conviction would be a conscious mental state. In this sense, explicit knowledge of ones mental state depends on that explicit knowledge being produced only implicitly; for example, only with inferences that are themselves implicitly represented. This is the crucial interaction between implicit and explicit knowledge we wish to dwell on.

Applying these notions to artificial grammar learning, consider a subject who sees a test string, applies knowledge-of-the-grammar, classifies the test string as grammatical or not (knowledge-of-the-test-string), and then gives a confidence rating.

Subjects' different degrees of consistency to different test strings show that subjects, when classifying different strings, are in states of thinking with different degrees of conviction. If subjects confidence ratings are unrelated to their consistency then their higher order thoughts (confidence ratings) are not sensitive to their actual mental states of thinking with more or less conviction ("knowledge states", for short). We have taken this to be evidence of the knowledge states being unconscious. In fact, however, in some cases they will be in a state of thinking-with-a-lot-of-conviction and give a high confidence rating; in these cases, they do have a higher-order thought (attitude-explicit knowledge) to the effect that they are in a state that they are in; so if the confidence rating came to them in a way that appeared unmediated, the state would be a conscious state. If confidence ratings appear unmediated to the subject, the lack of relationship between confidence and consistency implies some knowledge is unconscious, even though it allows some knowledge to be conscious. This is one refinement we must add to our previous interpretation of a lack of relationship between confidence and accuracy.

Now consider Allwood et al's results. Demonstrating a relationship between confidence and accuracy is not sufficient for demonstrating that the knowledge is conscious; one also needs to determine what the subject believes the confidence ratings are based on. Strictly speaking, if subjects base the ratings on inferences (e.g. perceived reaction times, perceived fluency) of which they are conscious, the knowledge states are still unconscious. Although we are in progress with an experiment that includes asking subjects to report on the bases of their confidence ratings, we regard this more as a means for us as psychologists to generate ideas, rather than as a test of the conscious status of their knowledge; the latter needs to be methodologically simpler. A lack of relationship between confidence and accuracy does imply that at least some of the knowledge states are unconscious, and so this remains a valuable criterion.

If subjects' ratings are based on explicit inferences, including the products of implicit knowledge (e.g. fluency), the knowledge states are unconscious; if confidence ratings come to be based on implicit

knowledge in a way that appears unmediated to the subject, and hence confidence is calibrated with accuracy, then the states of knowledge are conscious. The knowledge states referred to here are knowledge about the grammatical status of test strings. Even if these knowledge states were all conscious, it would leave open the possibility that knowledge of the grammar was unconscious.

Thus, Allwood et al's intuitions that subjects in their experiment two still had, in some sense, implicit knowledge could be due to (a) despite the subjects' well calibrated confidence ratings, this calibration was based on conscious inferences regarding the mental state the subjects must have been in (states of knowing the grammatical status of strings), and so those mental states were still in fact unconscious; or (b) the mental states of knowing the grammatical status of the strings were in fact conscious (the confidence ratings were not based on conscious inferences), but the mental states of knowing the grammatical rules were not conscious; subjects did not have non-inferential higher order thoughts about being in those latter states.

If subjects' become conscious of their mental states responsible for grammaticality judgements, methods of using confidence ratings for grammatcaility judgements can no longer be used to show knowledge of grammatcal rules is unconscious. Dienes and Perner (2001b) discuss how to measure implicit knowledge under these conditions.

Rosenthal (2000) discusses how higher order thoughts need not be produced by a 100% reliable means; however they are produced, so long as they appear unmediated, they produce conscious awareness of being in a certain mental state. The first order mental state about which one has a second order thought need not even exist for the subject to consciously experience being in a certain state. Subjects' higher-order thoughts partly constitute their theories about their mental states; such theories are not therefore a nuisance that get in our way as experimenters; they are part of the very thing to be investigated and explained.

Conclusion

We have argued for a second-order representational account of consciousness. Consciousness can never be produced just by, for example, a sustained pattern of activation in a connectionist network per se (e.g. O'Brien & Opie, 1999); the properties of the representation must in certain respects be like those of people to count as mental states and the system must be able to refer to those states in further representations (Perner & Dienes, 1999; see Dienes & Perner, 2001c for discussion). Such considerations lead directly to metacognitive measures of the consciousness of mental states. Conscious states cannot be measured just by discriminative responding, but only by evidence that the subject is conscious of the mental state.

To summarize the argument of this paper, we have taken two points made by Rosenthal regarding his higher order thought theory; namely that (a) higher order thoughts are occurrent states rather than dispositional states (like knowing might be); and (b) the higher order thoughts must not result from inferences of which the person is conscious. These two points turn out to have important implications for the measurement of implicit or explicit knowledge in (for example) the implicit learning literature. The contribution of this paper is to show in detail the relevance of Rosenthal's theory to psychologists interested in determining the conscious or unconscious status of mental states in (for example) implicit learning studies.

We are conscious of mental states when our explicit knowledge of them is based purely on implicit knowledge, when the knowledge of being in the state does not arise out of any inference of which we are conscious. Note there is a symmetry here with volition; we have voluntary control over an act only when the intention produces the act by mechanisms of which we are unconscious (Dienes & Perner, 2001b).

The importance of considering the conscious status of the inferences leading to a judgement has also been highlighted by Koriat (e.g. 1998). We hope we have elucidated further implications of the interaction between implicit and explicit inferences and implicit and explicit knowledge states.

Acknowledgments

Thanks to David Rosenthal for valuable discussion before this paper was written.

References

Allwood, C. M., Granhag, P. A., Johansson, H. (in press). Realism in confidence judgements of performance based on implicit learning. *European Journal of Cognitive Psychology*.

Carruthers, P. (1992). Consciousness and concepts. Proceedings of the Aristotelian Society, Supplementary Vol. LXVI, 42-59.

Carruthers, P. (2000). *Phenomenal consciousness naturally*. Cambridge: Cambridge University Press.

Cheesman J. & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, 36, 387-395.

Dienes, Z., & Berry, D. (1997). Implicit learning: below the subjective threshold. Psychonomic Bulletin and Review, 4, 3-23.

Dienes, Z., Kurz, A., Bernhaupt, R., & Perner, J. (1997). Application of implicit knowledge: deterministic or probabilistic? *Psychologica Belgica, 37*, 89-112.

Dienes, Z., & Perner, J. (1999) A theory of implicit and explicit knowledge. *Behavioural and Brain Sciences*, *22*, 735-755.

Dienes, Z., & Perner, J. (2001a) A theory of the implicit nature of implicit learning. In Cleeremans, A., & French, R. (Eds), *Implicit learning*. Psychology Press.

Dienes, Z., & Perner, J. (2001b). The metacognitive implications of the implicit-explicit distinction. In Chambres, P., Marescaux, P.-J., Izaute, M. (Eds), *Metacognition: Process, function, and use*. Kluwer.

Dienes, Z., & Perner, J. (2001c). Unifying consciousness with explicit knowledge. In Cleeremans, A. (Ed.) *The unity of consciousness: binding, integration, and dissociation.* Oxford University Press.

Koriat, A. (1998). Metamemory: the feeling of knowing and its vagaries. In M. Sabourin, F. Craik, & M. Roberts (Eds), *Advances in psychological science* (Vol. 2). Hove, UK: Psychology Press.

Millikan, R. G. (1984). Language, thought, and other biological categories. Cambridge, MA: MIT Press.

Millikan, R. G. (1993). *White queen psychology and other essays for Alice*. Cambridge, MA: Bradford Books/MIT-Press.

O'Brien, G., & Opie, J. (1999). A connectionist theory of phenomenal experience. *Behavioural and Brain Sciences, 22*, 127-196.

Perner, J. (1991). *Understanding the representational mind*.. Cambridge, MA: MIT Press. A Bradford Book.

Perner, J., and Dienes, Z. (1999) Higher order thinking. Behavioural and Brain Sciences, 22, 164-165.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour, 6,* 855-863.

Reber, A.S. (1989). Implicit learning and tactic knowledge. *Journal of Experimental Psychology: General, 118*, 219-235.

Rosenthal, D.M. (1986). Two concepts of consciousness. *Philosophical Studies*, **49**, 329-359.

Rosenthal, D.M. (2000).Consciousness, Content, and Metacognitive Judgments, *Consciousness and Cognition*, 9, 203-214.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioural and Brain Sciences*, 17, 367-448.

Twyman, M., & Dienes, Z. (2001). Metacognitive Measures of Implicit Knowledge. 2001 Convention of The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), York March 21st-24th.