

REPLIES

Implicit synthesis

ZOLTÁN DIENES

University of Sussex, Brighton, England

and

DIANNE BERRY

University of Reading, Reading, England

In this reply to Neal and Hesketh and to the commentators, we argue that implicit knowledge is partly abstract and can be usefully defined by the criteria of both metaknowledge and intentional control. We suggest that the pattern of dissociations supports a claim of separate implicit and explicit learning modes. According to our characterization, implicit learning leads to knowledge that is not automatically represented as knowledge by the learning process; instead, the presence of knowledge has to be inferred by the subject (partial explicitation) if metaknowledge is gained at all. During explicit learning, knowledge is automatically labeled as knowledge by the learning process, so that metaknowledge comes immediately and to the fullest extent. Finally, we suggest that implicit knowledge may to some degree apply regardless of intention.

The commentators (Mathews, 1997; Perruchet, Vinter, & Gallego, 1997; Reber, 1997; Stadler, 1997; Whittlesea & Dorken, 1997) and Neal and Hesketh (1997) all raise interesting points that sharpen the debate on implicit learning. First, we will deal with the claim that implicit knowledge is episodic and not abstract; second, with issues to do with the dissociations between implicit and explicit learning; and finally, with issues to do with defining implicit knowledge.

Abstractness of Implicit Knowledge

Neal and Hesketh (1997) claim that implicit learning produces knowledge that is episodic and not abstract. As noted by Reber (1997), most current computational models of human learning include an abstractive component. For example, within the concept formation literature, pure exemplar models have been superseded by those that include a delta rule component between exemplars and categories (Estes, 1994; Nosofsky, Kruschke, & McKinley, 1992); within the implicit learning literature, the simple recurrent network (SRN) has been successfully applied to the sequential reaction time task (Cleeremans, 1993) and to artificial grammar learning (Dienes, Altmann, & Gao, 1995). Cleeremans pointed out that the SRN forms representations that could be regarded as lying on some continuum of abstractness between that of exemplar models and formal rules. So on the issue of abstractness, we reject

a clear dichotomy. Further progress is most likely to be made by formulating precise models, which we believe will have abstract components, as well as sensitivity to the encoding conditions present at training.

Dissociations Between Implicit and Explicit Learning

We take a different slant from that of Neal and Hesketh (1997) on the various criticisms of the use of dissociations. Any set of data always underdetermines explanations. A set of data may not *prove* the existence of implicit and explicit learning, but the data still need to be explained. So the question is, do implicit and explicit learning models provide the most coherent and elegant account? Neal and Hesketh dismiss the data on the effects of secondary tasks on learning as not decisively indicating different implicit and explicit knowledge bases. But what is their alternative explanation? Whittlesea and Dorken (1997) also question the use of dissociations in inferring different implicit and explicit learning modes, but also do not, in our view, provide compelling alternative explanations. For example, Whittlesea and Dorken consider the finding that incidentally trained subjects may not transfer as well as intentionally trained subjects to embodiments of a grammar in different surface features. Whittlesea and Dorken suggest that this finding may arise simply because rules induced by the intentional subjects apply to abstract features, but incidental subjects would have by-and-large memorized surface features. However, this explanation does not account for transfer failing between domains where the mapping is highly transparent—for example, between colors and their equivalent color names (Dienes & Alt-

Correspondence should be addressed to Z. Dienes, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, U.K. (e-mail: dienes@epunix.biols.susx.ac.uk, or d.c.berry@reading.ac.uk).

mann, in press). Any memorized training exemplars in either domain are in principle equally useful in making classification decisions in both domains, so why should incidentally trained subjects fail to transfer completely? Note that if the retrieval of the training exemplars is an explicit process based on cued recall, aspects of context that do not affect the meaning of stimuli (as they do not in this experiment) should not affect retrieval (Baddeley, 1990, p. 287). Only implicit rather than explicit memory seems to be sensitive to perceptual cues when cues to the identity of the stimulus are present (Schacter, 1987). That is, the failure to transfer between colors and color names is not plausibly explained by a failure of explicit retrieval across different contexts. So a distinction between implicit and explicit processes is still needed in order to provide a coherent account of all the data.

Stadler (1997) also questions the usefulness of the flexibility criterion and points out that free recall can be context bound. That is, both implicit and explicit processes can be context bound. Although Stadler is making a different point from that of Whittlesea and Dorken (1997), our response to Whittlesea and Dorken also partly covers Stadler's question. Explicit retrieval is not affected by purely perceptual changes when a part or whole of the stimulus is present (i.e., cued recall or recognition); only implicit retrieval is affected by perceptual changes under these conditions. Stadler, Mathews (1997), and Reber (1997) all point out cases where implicit knowledge shows some flexibility. Reber lists a number of studies in the artificial grammar learning paradigm finding no significant difference between transfer and same domain performance. All these studies had confidence intervals that included the size of difference detected in the other studies that did show a drop in performance from same domain to transfer (of about 20%). That is, the "null result" studies are not informative about whether performance drops between transfer and same domain. Mathews may be right in suggesting that flexibility improves with practice. His is an intriguing suggestion, because the opposite is often assumed to be the case, but it would account for the natural language case. Perhaps what is required for increasing flexibility is training under different conditions with different exemplars. On the issue of robustness, it strikes us that Stadler is right to say that robustness will turn out to be too broad a concept to distinguish implicit from explicit learning. But we believe that it points in the right direction until detailed process models specify more precisely the ways in which implicit learning is relatively robust and the ways in which it is not.

Reber casts us as "loving dichotomies." Partly: we do argue that there are distinct implicit and explicit learning modes, but performance is typically a blend of contributions from both modes (Berry & Dienes, 1993). Like Stadler (1997) and Perruchet, Vinter, and Gallego (1997), Berry and Dienes argued that implicit learning is associative, like a connectionist network; explicit learning is more symbolic, so there is a qualitative computational difference between implicit and explicit learning modes.

Defining Implicit

Both Neal and Hesketh (1997) and Whittlesea and Dorken (1997) suggest that the concept of consciousness as a crucial dimension should be abandoned. Part of the problem in accepting consciousness as a useful construct may be that people expect to be presented with *the* definition or criterion of consciousness. But our everyday notion of consciousness may not be that of a single natural kind; our intuitions are complex, and they may only approximate the real divide or divides in nature. For example, Dienes, Altmann, Kwan, and Goode (1995) found that knowledge that people did not know they had could be applied intentionally, dissociating awareness from intention. Rather than argue over what the essence of consciousness *really* is, we can take our everyday intuitions to see what criteria they inspire, and then take those criteria as potentially interesting in their own right (see Stadler, 1997, for a similar argument): Do they separate qualitatively different types of learning modes or knowledge bases? We need not be concerned if these criteria pick out all or only those cases that everyday intuition decides is conscious; our folk theories are to provide inspiration, not be the ultimate arbiters of truth. Also, we need not be worried about the "hard" problem of qualia, which Neal and Hesketh dwell on; there is the "easy" problem of the functional role of consciousness (the sort that qualia-less zombies could have). The functional role of consciousness could be defined according to any criteria that are inspired by, but necessarily determined by, everyday intuitions. We believe that criteria in terms of both metaknowledge (awareness) and intention will prove to be useful (see Richardson-Klavehn, Gardiner, & Java, 1996, for a similar argument in the implicit memory literature).

Metaknowledge

In criticizing the use of metaknowledge as a useful criterion, Whittlesea and Dorken (1997) suggest that people "do not have direct, conscious access to those representations" that drive performance (p. 64). What representations do subjects have access to? As cognitive psychologists, we all have representational theories of mind; thus, the content of our experience is just the content of some representation. So we do have direct access to some representations. That is just the point: We have direct access to some (the explicit) and not others (the implicit). Furthermore, the process of forming some of these representations directly produces representations about how we formed those representations. A person looks at a list of words and sees that the first word is *antelope*. She must have formed the representation "the first word on the list is *antelope*." Furthermore, if that representation is conscious, she will have formed the higher order representation "I am seeing that the first word on the list is *antelope*" (Dienes & Perner, 1996; Rosenthal, 1986). It is this higher order representation that allows *explicit* recollection of the word *antelope*: she represents not just the word form *antelope* (this much is sufficient for performance on implicit memory tasks), but the process by

which she knows it was on the list (i.e., she saw it). Explicitly formulating and testing hypotheses, explicitly retrieving previous instances and drawing analogies are explicit by virtue of representing the fact that we are doing these things. On the other hand, implicit learning is likely to be something like learning in connectionist networks; the representations involved in learning are not themselves capable of being objects of further representations (Cleeremans, in press; Dienes & Perner, 1996). In that case, metaknowledge does not arise directly from the processes of learning, but, as suggested by Whittlesea and Dorken (and Dienes, Altmann, et al., 1995), has to be inferred (on the basis of, e.g., familiarity) if it is achieved at all.

Whereas Whittlesea and Dorken (1997) suggested that representations are unconscious, Perruchet et al. (1997) suggested that all mental representations are conscious, but that it is just the mechanisms that produce them that are unconscious—a claim we find just as inappropriate. For the suggestion to have substance, a definition of representation is needed. One way of defining representation is taken by Dretske (1988): *Y* (e.g., a pattern of neural activity) is a representation of *X* just in case it is the function of *Y* to covary with *X*. Given this definition, it is clear that not all representations are conscious (why should they be?). Is it just *mental* representations that are conscious? But then, what makes a representation mental? Following the higher order thought theory of consciousness (Armstrong, 1968; Carruthers, 1992; Rosenthal, 1986), Dienes and Perner (1996) have suggested that to be conscious of some state of affairs (e.g., that this chair is red), the mental state by which this state of affairs is beheld must be represented (i.e., that I *see* that the chair is red). That is, for *Y*, the representation of a fact (e.g., “the chair is red”), to be conscious, it must be that the representation *Y* is an object of a further representation (“I represent ‘*Y*’ by virtue of seeing it”). This representational way of looking at consciousness leads naturally to the subjective threshold criterion of consciousness; if the learning process (e.g., a connectionist one) does not automatically lead one to represent one’s knowledge as knowledge, the knowledge will be below a subjective threshold. However, we agree with Mathews (1997), Reber (1997), and Whittlesea and Dorken in the claim that suitable post hoc inferences may later enable the knowledge to be represented as knowledge and thus to be partly explicated; inferences are unlikely to fully explicate all the knowledge. The interesting comparison occurs between this learning system and one in which representations of the status of the putative knowledge are automatically formed during the process of learning (i.e., explicit learning).

Perruchet et al. (1997) point out (as Jacoby and his colleagues have repeatedly noted) that unconscious processes can give rise to conscious experiences, and they suggest that this undermines the approach of Jacoby and his colleagues. But one always has to ask “Conscious of what?”

Jacoby (e.g., 1991) can talk of unconscious influences of memory because subjects are not conscious of the experience *as* being a memory; similarly, subjects may have unconscious knowledge in that subjects are not conscious of the knowledge *as* knowledge (except, in both cases, by inference). (Note that the unconscious influence of a prior episode does not always give rise to a conscious experience of recollection *or* familiarity: Richardson-Klavehn, Gardiner, and Java [1994] showed that studying words could lead them to be included on an exclusion test without any accompanying feelings of familiarity; Jacoby, Allan, Collins, and Larwell [1988] showed that studying words affected the later perception of background noise.)

Perruchet et al. (1997) present their difficulties with the subjective threshold criterion. As we defined it in our paper, “people’s knowledge could be said to be below a subjective threshold if they lack metaknowledge about their knowledge” (Dienes & Berry, 1997, p. 5). Perruchet et al. claim that there is “internal inconsistency” in the application of this concept to subjects’ performance on grammaticality judgment tasks and on cued report tests, but they fail to specify where the inconsistency lies. If we ask subjects to rate their confidence in a grammaticality decision, we can determine whether subjects know that they know that particular strings are grammatical. If we ask subjects to rate their confidence in deciding which bigrams are allowed by the grammar, we can determine whether subjects know that they know which bigrams are allowed by the grammar. Clearly, the task we give subjects determines what knowledge we are testing, but as long as this is clear, there are no inconsistencies. (The distinguishing of different types of knowledge to which the subjective threshold could apply—i.e., rules of the grammar versus particular strings being grammatical—is discussed in detail in Dienes & Perner, 1996.)

Perruchet et al. (1997) dismiss the zero correlation criterion on the grounds that in our native language we are more likely to be correct in grammaticality judgments when we are confident rather than unconfident. One response to this comment has already been made in the target article (Dienes & Berry, 1997, p. 5). Suffice it to point out that finding a relationship between confidence and accuracy indicates the presence of *some* metaknowledge, and no one will be surprised to hear that we have *some* explicit knowledge about our native language. In a complementary way, a significant result with the guessing criterion indicates the presence of some implicit knowledge. It is quite possible for the zero correlation criterion to indicate the presence of some explicit knowledge and simultaneously for the guessing criterion to indicate the presence of some implicit knowledge.

Perruchet et al. (1997) also criticize the notion of a subjective threshold because they believe it implies that there is a hypothetical knowledge base that could be made explicit with more effort or time, a notion to which they object. In fact, the notion of a subjective threshold is

entirely consistent with the knowledge's always being implicit (as in the case of subliminal perception) or in principle being explicable, so this "criticism" is a red herring.

Intentional Control

Neal and Hesketh (1997) and Stadler (1997) recommend using intentional control (at testing and learning, respectively) as a criterion for distinguishing implicit from explicit knowledge. We agree that this may prove to be a useful criterion, in addition to one in terms of the subjective threshold. Perruchet et al. (1997) argue that pursuing a criterion in terms of intentional control is "objectless." They discount the Jacoby process dissociation procedure on the grounds that it may only separate memory in which one remembers the context from that in which one does not (for a similar point, see Dodson & Johnson, 1996). Dienes, Altmann, et al. (1995) earlier pointed out that this argument did not apply to the Stroop interference effects that flanking items can produce on a target item (subjects are faster if the flanking items and the target items are consistent in their old/new status rather than inconsistent) shown by Jacoby, Ste-Marie, and Toth (1993). Dienes (1996) extended this paradigm to the artificial grammar learning case: Subjects were faster to respond "grammatical" or "nongrammatical" to a target test item if the to-be-ignored flanking items were consistent rather than inconsistent. The same effect emerged if the flanking items were from a completely different domain, demonstrating the nonintentional transfer of knowledge across domains (between letters and colors). The demonstration of Stroop-like effects in memory lends credence to the idea that the assessed "automatic" influences in other of Jacoby's experiments really reflected, at least partly, the automatic application of knowledge. Note that the automatic application of knowledge in Dienes (1996) contrasts with the strategic control over which grammar to use shown by subjects in Dienes, Altmann, et al. (1995). The solution may be that people have strategic control over which body of knowledge to use, but once a body of knowledge is chosen, it applies automatically to everything in sight, even across domains.

Concluding Note

Mathews (1997) recommended looking at what one could agree with in the papers being considered. It is surprising that, amidst the claims of violent disagreement, the actual content, aside from terminological wrangles, of the papers by believers and nonbelievers and those who say they have transcended the issue remains exceedingly similar. Everyone, we think, is agreed that there is an important learning mechanism, pervasive in its effects, sensitive to the conditions of training and testing, but also capable of some flexibility, producing knowledge about which the subject does not directly know; producing knowledge by unconscious associative processes—knowledge that affects conscious experience and with time allowing inferences (rather than direct perception)

as to its status. This agreement is a measure of what we have actually achieved in the field.

REFERENCES

ARMSTRONG, D. M. (1968). *A materialist theory of mind*. London: Routledge.

BADDELEY, A. D. (1990). *Human memory: Theory and practice*. Hove, U.K.: Erlbaum.

BERRY, D. C., & DIENES, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove, U.K.: Erlbaum.

CARRUTHERS, P. (1992). Consciousness and concepts. *Proceedings of the Aristotelian Society*, **67**, 41-59.

CLEEREMANS, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.

CLEEREMANS, A. (in press). Principles for implicit learning. In D. C. Berry (Ed.), *How implicit is implicit learning?* Oxford: Oxford University Press.

DIENES, Z. (1996, April). *How unconscious is the transfer of knowledge of artificial grammars across domains?* Paper presented at the 1996 annual conference of the British Psychological Society, Brighton.

DIENES, Z., & ALTMANN, G. T. M. (in press). Transfer of implicit knowledge across domains: How implicit and how abstract? In D. C. Berry (Ed.), *How implicit is implicit learning?* Oxford: Oxford University Press.

DIENES, Z., ALTMANN, G. T. M., & GAO, S.-J. (1995). Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. In L. S. Smith & P. J. B. Hancock (Eds.), *Neural computation and psychology: Proceedings of the 3rd Neural Computation and Psychology Workshop (NCPW3), Stirling, Scotland, 31 August–2 September 1994* (pp. 19-33). London: Springer-Verlag.

DIENES, Z., ALTMANN, G. T. M., KWAN, L., & GOODE, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1322-1338.

DIENES, Z., & PERNER, J. (1996). Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit cognition* (pp. 227-256). Oxford: Oxford University Press.

DODSON, C. S., & JOHNSON, M. K. (1996). Some problems with the process dissociation approach to memory. *Journal of Experimental Psychology: General*, **125**, 181-194.

DRETSKE, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

ESTES, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

JACOBY, L. L., ALLAN, L. G., COLLINS, J. C., & LARWILL, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 240-247.

JACOBY, L. L., STE-MARIE, D., & TOTH, J. P. (1993) Redefining automaticity: Unconscious influences, awareness, and control. In A. Baddeley & L. Weiskrantz (Eds.), *Attention, selection, awareness, and control: A tribute to Donald Broadbent* (pp. 261-282). Oxford: Oxford University Press.

MATHEWS, R. C. (1997). Is research painting a biased picture of implicit learning? The dangers of methodological purity in scientific debate. *Psychonomic Bulletin & Review*, **4**, 38-42.

NEAL, A., & HESKETH, B. (1997). Episodic knowledge and implicit learning. *Psychonomic Bulletin & Review*, **4**, 24-37.

NOSOFSKY, R. M., KRUSCHKE, J. K., & MCKINLEY, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 211-233.

PERRUCHET, P., VINTER, A., & GALLEGO, J. (1997). Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin & Review*, **4**, 43-48.

- REBER, A. S. (1997). Implicit ruminations. *Psychonomic Bulletin & Review*, *4*, 49-55.
- RICHARDSON-KLAVEHN, A., GARDINER, J. M., & JAVA, R. I. (1994). Involuntary conscious memory and the method of opposition. *Memory*, *2*, 1-29.
- RICHARDSON-KLAVEHN, A., GARDINER, J. M., & JAVA, R. I. (1996). Memory: Task dissociations, process dissociations, and dissociations of consciousness. In G. Underwood (Ed.), *Implicit cognition* (pp. 85-158). Oxford: Oxford University Press.
- ROSENTHAL, D. (1986). Two concepts of consciousness. *Philosophical Studies*, *94*, 329-359.
- SCHACTER, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 501-518.
- STADLER, M. A. (1997). Distinguishing implicit and explicit learning. *Psychonomic Bulletin & Review*, *4*, 56-62.
- WHITTLESEA, B. W. A., & DORKEN, M. D. (1997). Implicit learning: Indirect, not unconscious. *Psychonomic Bulletin & Review*, *4*, 63-67.

(Manuscript received November 20, 1996;
accepted for publication November 25, 1996.)