Implicit learning: Indirect, not unconscious

BRUCE W. A. WHITTLESEA and MICHAEL D. DORKEN
Simon Fraser University, Burnaby, British Columbia, Canada

Dienes and Berry (1997) argue that there really are two kinds of knowledge, implicit and explicit, that can be characterized through a subjective threshold. We argue that the subjective threshold does not separate two forms of knowledge, but instead two uses of knowledge, for two different purposes. With Neal and Hesketh (1997), we suggest that an understanding of learning will be better served by investigating the processing that occurs under various circumstances than by attempting to isolate learning that occurs with awareness from that which occurs without it.

Psychologists love dichotomies. Faced with the complexity of behavior, we have traditionally responded by subdividing performance into modules governed by different principles: procedural versus declarative, semantic versus episodic, short- versus long-term remembering, and so on. In attempting to understand how organisms adapt to their environment by changing their behavioral tendencies, we have imposed multiple dichotomies, including controlled/automatic, incidental/intentional, and implicit/explicit. Like the descriptions of an elephant by the seven blind men, each of these has merit, but none quite captures the subject matter.

In their comprehensive review, Dienes and Berry (1997) defend the utility of the implicit/explicit distinction in the implicit learning literature. They argue that many of the studies they review demonstrate learning occurring above an objective threshold but below a subjective threshold, and that this learning has properties different from those of learning that occurs with awareness. In contrast, we argue that every act of learning carries with it a change in the potential to perform many other activities. People are unaware of these changed potentials, not because the learning occurs under a threshold, but because they are only incidentally related to what the person actually learns. We argue that implicit learning is just ordinary learning without becoming aware of the implications of that learning for performing unanticipated activities.

The Original Learning:
Implicit Learning or Acquiring a Potential?

In a standard "implicit learning" experiment, subjects study a set of exemplars of a grammar and are then given a surprise classification test, in which they show sensitivity to the underlying rules, without apparent awareness of the basis of their success. This observation has led many

authors to conclude that the subjects acquired some knowledge implicitly during the training and used that knowledge implicitly in test. We will question the implicitness of the use of knowledge in the next section; first, we will discuss the implicitness of the original acquisition.

There is an important ambiguity in the meaning of the words "learning" and "knowledge." One sense is the actual, direct computation of information about a stimulus, such as the computation that ROAD is a synonym of STREET. In that case, the person actually acquires a knowledge unit, a representation of the experience of computing the relationship. The person can know that he has that knowledge by recalling it. However, along with that representation, the person also acquires the potential to perform differently on future occasions. For example, he may experience savings on a later occasion in which he is again asked about the relationship of ROAD and STREET. This potential to perform is not directly computed and represented in memory, nor is it computable in detail: Any stimulus experience provides resources for an infinite number of possible future occurrences. The person cannot know the potential abilities that the new representation entails until that potential is realized in some future stimulus encounter.

On the first occasion that a person compares ROAD and STREET, that person already possesses information about each that makes it possible to perform the judgment. But that information was not directly about the relationship between ROAD and STREET, and until it occurs to the person to perform the comparison, she cannot know that she has the knowledge to do it. In learning separately about the concepts, the person acquired the potential to perform the comparison, but it remains only a potential until the person actually computes that relationship. Is this implicit knowledge? Did the person implicitly learn the relationship between ROAD and STREET, and also implicitly learn about all of the other future uses to which that information could be put, in learning earlier about the separate concepts? Or did she simply learn about those concepts in terms of whatever purpose she had at the time, and in so doing acquire the potential to perform other tasks? Or take a procedural example: a person learning to

play the piano is acquiring finger dexterity skills that may later facilitate learning to type or to be a mechanic. This person will probably not realize the implication of the early training at that time, and later may not connect that practice with ease in learning the new vocation. Clearly, there is implicit transfer of skill, the potential for which was created in the early training: but piano practice is not implicit learning to be a mechanic.

Similarly, subjects memorizing instances of an artificial grammar compute some properties of those items, encoding bigrams, substrings, or representations of whole items. This learning enables them to classify novel exemplars as legal or illegal, because legal items are necessarily more similar to encoded representations than are illegal items. That is, they directly encode information about items in one task that indirectly gives them the ability to deal with related stimuli in an unanticipated task. However, the learning of items in the memorization task is not direct learning of the ability to classify other items: It is simply learning of the items for the purpose given by that task, which incidentally grants the potential to perform other activities on related stimuli. The ability to classify later items is implicit in the original learning, in the sense of existing as a potential, but not in the sense of a direct computation of information relevant to that activity. In consequence, we argue that the original acquisition of knowledge in an "implicit learning" experiment is not qualitatively different from learning performed under any other circumstances.

The original motive for considering the possibility of implicit learning as a separate mode of learning was the idea that people memorizing instances of a grammar also learn general, abstract information about the deep structure of the domain without being aware of the latter (e.g., Reber & Allen, 1978). That is, it was suggested that they acquire the ability to perform the transfer task through direct computation and representation of information that is irrelevant to their overt task. If people did directly compute such extra information without intending to, there would be good reason to describe that learning as a special mode that was different from explicit learning. However, as discussed in the review by Neal and Hesketh (1997), there is no positive evidence of such extra computation. We conclude that the ability to perform a transfer task is simply the realization of a potential that was created in the act of encoding items for the purpose of memorization.

Metaknowledge: Awareness or Inference?

We now turn to the question of people's ability to identify the source of their ability to perform the transfer task. Dienes and Berry (1997) argued for a threshold account, in which knowledge that was inaccessible to consciousness could still control behavior. They offered two criteria for establishing that knowledge was beneath a subjective threshold: the guessing criterion, that subjects believe they are guessing while performance is above chance, and the zero-correlation criterion, that confidence is unrelated

to accuracy while performance is above chance. Both are common-sense ideas, corresponding to the intuitive notion of what it means to be conscious of one's ability or knowledge. However, these ideas turn out to be more complex than they initially appear.

We believe that the idea of "being aware or not aware of the knowledge" miscasts the way in which people come to a belief about what they know. We argue that what people say about their knowledge and the confidence that they express in their performance are the product of an inferential process. People possess representations of the events they have experienced, and those representations drive future performance. However, people do not have direct, conscious access to those representations. Instead, when a stimulus is encountered for some purpose, they can only observe their performance and attempt to judge the likely source of their behavior, given their understanding of the requirements of the task (cf. Nisbett & Ross, 1980). We argue that failure to develop the attitude that one has knowledge of how to perform the transfer task in an "implicit-learning" experiment is not the product of learning below a threshold of awareness. Instead, it occurs because the subjects do not interpret their behavior as evidence of such knowledge, under their theory about what is necessary to perform the task. That is, we argue, consciousness is an interpretive activity and not a passive knowing.

This characterization of learning as acquiring the potential to perform unanticipated future tasks and of metaknowledge as interpreting one's performance goes to the heart of the controversy over implicit learning. If the episodic-processing account of learning (reviewed by Neal & Hesketh, 1997) is correct, then people learn during the training session in both actual and potential senses. In processing training items to meet the demand to memorize, people compute some organization of each training item (whether at the bigram, substring, or item level). They thus know the individual items to some extent, in the sense of having representations of encoding those items. That knowledge grants the potential to discriminate between novel legal and illegal items in the classification test. They can also be aware of possessing this knowledge, in the sense that in a recognition test they can discriminate those items from others, and do so with a feeling of familiarity for the original items.[1] However, they do not become aware that it is that knowledge that permits them to succeed in the classification test. Being informed by the experimenter that test items differ in conformity to a set of abstract rules, the subjects assume that they could only succeed in discriminating the items if they possessed abstract knowledge. Being unable to recall any such knowledge (for the good reason that it was never computed), the subjects lack confidence in their classification judgments and claim to know nothing that could assist them in the test. In effect, the subjects lack meta-metaknowledge: Operating under the wrong theory about what knowledge is necessary in order to perform the surprise classification, they are unable to in-

terpret their ability to recall individual training items as metaknowledge about their ability to classify the stimuli.

This understanding of metaknowledge can be illustrated using Chan's (1992) data, which Dienes and Berry (1997) quote as a demonstration of performance beneath a subjective threshold. Chan's memorization and rule-searching subjects both apparently used knowledge of bigrams to classify test items as legal or illegal, but the memorization subjects' confidence was less related to their accuracy than was that of rule-searching subjects (zero-correlation criterion). Dienes and Berry concluded that memorization subjects lacked metaknowledge that rule-searching subjects possessed. Agreed—but was the knowledge they used to classify items inaccessible to the memorization subjects (below a threshold) or was it instead that memorization subjects failed to interpret their item knowledge as being relevant to the legality decision? Chan's further finding that confidence was found to be related to accuracy when memorization subjects were later asked to rate bigrams for grammaticality suggests to us that the subjects were able to recall and interpret their knowledge of bigrams as relevant when an appropriate context was provided at test.

A study by Whittlesea, Brooks, and Westcott (1994) offers another insight into the relationship between classification accuracy and confidence. They exposed subjects to exemplars of two categories, each based on a prototype. The subjects were trained twice on the same stimuli before being tested. In one training phase, they were required to copy the stimuli, thus encoding each item as a unique entity; in the other, they were asked to judge the frequency with which each feature of the item had occurred in previous items of that category, thereby learning about the typicality of the features.

There were two types of critical test stimuli, both of which were novel. One type (called typical) shared three features with its prototype and also three with the most similar training item; the other (called similar) shared only two features with its prototype but four features with the most similar training item. In making decisions about test stimuli, the subjects could rely on information about particular training items, in which case they would perform better on similar than on typical stimuli or, alternatively, rely on knowledge about the typicality of the component features of the item, resulting in typical items' being judged better than similar items.

At test, the subjects were asked to classify each item and then to rate their confidence that the classification was correct. In addition, they were asked to pronounce the item. We knew from previous studies that this task is usually performed using similarity information. The question was whether performing that task would influence the basis of performance of the other two tasks. Pronunciation was required after the other two, or before both, or between the other two.

When pronunciation came last, the subjects classified the stimuli above chance, and were more accurate on typical than on similar stimuli, indicating use of typicality information. Overall confidence was around .5 on a scale

between zero (no confidence) and unity (full confidence), and was also greater for typical than for similar stimuli. That is, (1) the subjects had some knowledge enabling them to classify test stimuli above chance, (2) the subjects claimed some confidence in having such knowledge, and (3) the patterns of confidence and classification were correlated. Thus, by both the guessing and zero-correlation criteria, the subjects had metaknowledge about their classification performance. When pronunciation came first, the pattern was reversed. Subjects were now more accurate on similar than on typical stimuli; confidence was around .7, but was now greater for similar than for typical stimuli. We interpreted this reversal to mean that subjects had shifted the basis of both judgments from typicality to similarity. However, in terms of the criteria of implicitness, the subjects again appeared to have metaknowledge.

In the third case, when pronouncing the item was inserted between the other tasks, accuracy was greater for typical than for similar stimuli, indicating use of typicality information. However, the subjects were more confident on similar than on typical items, indicating the use of similarity rather than typicality. Thus, by the zero-correlation logic, the subjects did not know what information they had based their classification decisions on: in fact, they used a completely different basis to perform the two decisions, even though the second was a decision about the accuracy of the first. In a grammar experiment, this is parallel to basing one's classification on abstracted knowledge about the general deep structure of the set, but basing one's confidence about the accuracy of that classification on the ease with which a similar training item comes to mind. However, the subjects certainly thought they knew how well they had classified: confidence in their accuracy was again high (about .7 on the zero-to-one scale).

Clearly, the subjects did not have explicit access to the knowledge they used to perform the classification, or they would have used that same knowledge to evaluate their confidence: they did not have metaknowledge about the basis of their performance. However, their confidence ratings were not random: they were based on whichever form of knowledge the subjects had used in the immediately preceding task. In that sense, the subjects were highly sensitive to the specific forms of knowledge they used to perform other judgments. Did they know what they knew?

We suggest that this is the wrong question. Just as people's "implicit, unconscious knowledge" is often not knowledge, in the sense of lying in a precomputed representation, but instead consists of the potential to perform toward some unanticipated stimulus—so people's "explicit, conscious knowledge" about their abilities is often not knowledge, in the sense of their having direct access to the actual basis of performance. Instead, it is an attribution about what one must know, an inference based on one's judgment about how one is performing now and what is required to perform the task (cf. Nisbett & Ross, 1980). We believe that subjects in this study based their confidence judgments on a feeling of goodness, a feeling

resulting from the relative fluency with which various tasks can be performed on different stimuli (e.g., Jacoby, Kelley, & Dywan, 1989; Whittlesea, 1993). More typical stimuli and stimuli that are more similar to training items are (for different reasons) easier to process than are less typical or less similar stimuli (Whittlesea, 1987). The feeling of confidence that these subjects experienced came not from privileged access to the basis on which they made their classification decisions, but from the ease they experienced in processing the stimuli.

Thus, we argue that the fundamental problem with the search for a way to distinguish implicit versus explicit knowledge is that knowledge has been characterized as the possession of some precomputed fact to which the subject has explicit or only implicit access. Instead, knowledge is the potential to perform some activity or judgment on an object, and metaknowledge is an inference about the source of that performance in prior experience. Performance is supported by representations of prior experience, but the ability to perform some activity toward a stimulus without developing an attitude of knowing the basis for that performance does not demonstrate the operation of a special implicit-learning mechanism.

Qualitative Differences: Two Modes?

Dienes and Berry (1997) pointed out that finding knowledge to be below a threshold of awareness is not of value unless knowledge that is above and below that threshold is qualitatively different. They argue that there really are qualitatively different modes of learning and knowledge, distinguished not only by metaknowledge, but also by breadth of transfer, probability of success in test, and robustness. We will comment only on the first two grounds; Neal and Hesketh (1997) have anticipated (however implicitly) our rebuttal to the third.

Transfer to related tasks. Dienes and Berry (1997) pointed out that transfer is often smaller when test stimuli are presented in novel surface features than when they are presented in the same surface features as training items. They concluded that implicit knowledge is to some extent perceptually bound and that this limitation on transfer reveals a form of learning that is qualitatively different from an explicit form.

Presumably, the contrast is with learning in the form of hypothesis testing, when the subject is told that training items possess a deep structure and is asked to discover it. We are not aware of studies on the breadth of transfer following such training, but we anticipate, with Dienes and Berry (1997), that it would differ from that seen after memorization: A subject who learned a useful rule would have equal success in transfer on stimuli presented either in the same or different features, because the rule is applicable regardless of the features in which items are presented.

Clearly, memorizing instances and educing rules are different activities, resulting in qualitatively different consequences for future performance. However, that is not a sufficient basis on which to argue that they reflect explicit and implicit modes of learning. Transfer across a change in surface features is limited following the memorization task, not because that task produces implicit knowledge, but because it encourages learning about the unique surface characteristics of items. Similarly, the broad transfer that accompanies rule eduction occurs not because the knowledge is explicit, but because it is abstract, independent of the features to which it is applied.

Whittlesea and Dorken (1993) asked some subjects to memorize grammatical items; others were shown the same stimuli, each with a marker under one of its letters, and asked to determine whether that letter was repeated elsewhere in the item. Neither group was informed about the existence of a grammar. At test, the subjects were informed of the grammar and asked to classify novel items. Both groups were tested on items presented in the same letter set as that of the training items and on the same items presented in a novel set of letters (the order of tests being counterbalanced). The memorization group showed above-chance classification on the same-letters test but less ability in the novel-letters condition, demonstrating the narrow transfer described by Dienes and Berry (1997). However, the other group performed at identical levels in the two tests. The analysis of repetition patterns that the second group was asked to perform provided them with abstract knowledge about the stimuli, information that was of equal applicability when the surface structure was changed. We thus argue that the narrow transfer that is often observed in implicit-learning experiments is not a distinguishing feature of a special implicit-learning mode. It is simply a result of the fact that, in order to distract subjects from the general rules of the set, researchers have used induction tasks that cause subjects to focus on surface properties of individual items.

Success in transfer. Dienes and Berry (1997) reported evidence that people who approach complex tasks in a relatively passive manner, simply memorizing the instances, perform as well as, if not better than, people who attempt to educe the deep structure of the task. They concluded that implicit learning leads to as much, if not more, success than explicit learning in dealing with complex domains.

Unquestionably, people learn different things about a domain of stimuli when they are told that the domain has a deep structure than when they are not. However, the superiority of memorizing instances over attempted rule eduction can be explained without resorting to an implicit learning mechanism. First, at any moment during training, a subject is faced with a particular instance but never encounters the deep structure as a unitary whole, unless mentally. That makes it easier to learn about individual members of the set than to learn about the abstract relations among features across items. To learn the separate stimuli, one has only to study each as it is presented. But to learn about the general abstract structure, one must hold in mind a large variety of potentially important structural units without confusing them.

Worse, the subjects have no chance of hitting on the right hypothesis: The only correct hypothesis is the complex set of rules used to generate instances. However, the

subjects might try one or two simple hypotheses that are correlated with the actual grammar. Initially, they are likely to select a few salient parts of early items, testing the frequency with which those parts occur in succeeding stimuli. This puts them at a disadvantage in two ways. First, they have not attended (and so not encoded) the remainder of each string, and so are missing a lot of the important structure. Second, if the parts they initially select are of low frequency across the set, they are likely to abandon those hypotheses as false (although any part of a training item represents an aspect of the deep structure), and begin testing some other parts. Because of the variability of stimuli, they are likely to conclude that no hypothesis has much merit. In the end, they have no useful rules, and also have not encoded much about the structure of particular items.

In contrast, although memorizing the items is not easy, it nevertheless causes the subjects to attend to all of the components that are relevant in defining the deep structure. Even poor encoding of whole items (e.g., the encoding of some bigrams) is thus likely to grant the subjects some basis for future well-formedness judgments. Thus, in dealing with complex systems, it may easily be more useful to encode some clear examples of the effect of the system than to attempt to educe its principles. If deliberate rule-induction succeeds in identifying the underlying principles, then it grants perfect control; but if it fails, it can leave the learner with no basis for future judgments. In contrast, learning about instances grants an imperfect but robust basis for future performance.

Moreover, the superiority of memorization over rule-eduction instructions is not an absolute fact; it depends on the way in which training items are shown. It occurs when items are presented one at a time, in random order (e.g., Reber, 1976). However, when the items are presented in an organized fashion, all at once, performance in the rule-learning condition can surpass that in the memorization condition (Reber, Kassin, Lewis, & Cantor, 1980). We conclude that any style of learning succeeds, to the extent that it is supported by the circumstances in which it occurs.

Fundamentally, better transfer following memorization than following hypothesis testing merely indicates that the former learning does not consist of hypothesis testing. It does not indicate that that learning consists of abstraction, or that it is passive, or implicit, or of a different mode of learning, controlled by different principles. Memorizing instances is an activity that is different from hypothesis testing, but it is not a mysterious process requiring a special implicit-learning mechanism.

Conclusion

Dienes and Berry (1997) have provided the most comprehensive summary to date of evidence for a distinction between implicit and explicit learning and knowledge. However, we cannot agree that this evidence makes a

real case for accepting the idea of a subjective threshold of metaknowledge, or for qualitative differences in performance that result directly from the implicit acquisition of knowledge. Instead, we agree with Neal and Hesketh's (1997) conclusion that abandoning the concept of awareness as the crucial dimension, and instead investigating the consequences of variation in processing resulting from differences in tasks, contexts, and prior history will prove most fruitful in our understanding of learning.

REFERENCES

CHAN, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer design systems.* Unpublished doctoral dissertation, University of Oxford.

DIENES, Z., & BERRY, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4,* 3-23.

JACOBY, L. L., KELLEY, C. M., & DYWAN, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

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

NISBETT, R. E., & ROSS, L. (1980). *Human inference: Strategies and shortcomings of human judgement.* Englewood Cliffs, NJ: Prentice-Hall.

REBER, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning & Memory, 2,* 88-94.

REBER, A. S., & ALLEN, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition, 6,* 193-221.

REBER, A. S., KASSIN, S. M., LEWIS, S., & CANTOR, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning & Memory, 6,* 492-502.

WHITTLESEA, B. W. A. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13,* 3-17.

WHITTLESEA, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19,* 1235-1253.

WHITTLESEA, B. W. A., BROOKS, L. R., & WESTCOTT, C. (1994). After the learning is over: Factors controlling the selective application of general and particular knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20,* 259-274.

WHITTLESEA, B. W. A., & DORKEN, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General, 122,* 227-248.

NOTE

1. That this awareness is not true access to representations of prior experience is demonstrated by the fact that people regularly mistake novel but grammatical items for items they have actually seen (cf. Whittlesea & Dorken, 1993). In a recognition test, in contrast to their claims of not knowing what enables them to classify novel items, people falsely claim to be able to know which items they have seen previously. Our claim is that metaknowledge is an inference based on the success of current processing, not direct access to the representations of past events that control current performance.