Journal of Experimental Psychology: General
1993, Vol. 122, No. 3, 397–400

0096-3445/93/$3.00

Automatic Abstraction of Stimulus Structure From Episodes: Comment on Whittlesea and Dorken (1993)

Robert C. Mathews and Lewis G. Roussel

B. W. Whittlesea and M. D. Dorken's (1993) experiments were designed to examine whether implicit learning reflects automatic abstraction of the structure underlying stimuli or instead reflects task-dependent encoding of particular experiences. Their episodic approach presents 2 difficulties: One problem is their assumption that automatic abstraction implies abstraction of knowledge directly from the stimulus rather than from memory representations of experiences. A 2nd problem is that they dichotomized information (item information and global regularities) that is naturally and easily acquired simultaneously. Purely episodic systems become inefficient and waste resources in real-life situations involving high levels of experience. We argue that it is more fruitful to view memory as a conceptualizer that enables efficient interactions with the environment.

Implicit learning phenomena are important because they highlight a paradigm shift away from viewing memory as an information storage device toward viewing it as an active mechanism of knowledge acquisition. Historically, psychology adopted an information-processing perspective on memory in which memory representations of experiences are created (encoding) and then stored in a warehouse (long-term memory) until reactivated by appropriate retrieval cues and subsequently used for some purpose. We believe that this emphasis on exact retrieval of information previously input into memory is derived more from the type of tasks that psychologists chose to study (e.g., list recall experiments) than from the function of human memory.

The primary function of human cognition is to facilitate effective interactions with the environment. As such, a system that minimizes the amount of information that must be processed to perform a task is more adaptive than one that more perfectly catalogs past experiences. Abstraction of stimulus structure reduces the amount of information that must be processed to effectively interact with the environment (e.g., Collins & Quillian, 1969; Smith & Medin, 1981). Therefore, systems that abstract stimulus structure are more adaptive and are preferable to systems that catalog on purely rational grounds (cf. Aha & Goldstone, 1990; Anderson, 1991).

People who have attempted to build artificial intelligence systems that acquire new information are well aware of the problems of putting too much information in memory (e.g., Lenat, 1983). Redundant and unnecessary information (appropriately called "mud") quickly makes these systems inoperable. Also, inattention to and failure to learn about redundant cues has been well documented in the learning literature with animal experiments (e.g., Kamin, 1969). We doubt that human memory systems are any less efficient than those of other animals.

The Whittlesea and Dorken (1993) article attempted to explain implicit learning phenomena in terms of the traditional memory theory embracing the warehouse metaphor. Whittlesea and Dorken provided evidence that (a) tasks influence the encoding of experiences and thereby affect the ultimate knowledge derived from a set of experiences and (b) particular memories of experiences can often serve as an adequate guide to performance without implicit abstraction of stimulus structure. We have no quarrel with either of these ideas or the evidence supporting these notions. However, these findings are equally consistent with modern abstractionist views and are irrelevant to the more fundamental question of whether the warehouse model is an adequate model of implicit knowledge.

The main objective of Whittlesea and Dorken's (1993) experiments was to demonstrate that encoding tasks alter the type and quantity of knowledge acquired from experience. For example, in their Experiment 5 they showed that tasks that require maintenance rehearsal of strings (5a), detection of repetition patterns (5b), or memory of strings (5c) result in different patterns of transfer performance that are consistent with the operations needed to perform the respective encoding task. Because encoding tasks influence the knowledge acquired from experiencing the strings, Whittlesea and Dorken argued that information about the strings was not automatically abstracted. Instead, the knowledge acquired was determined by the encoding task.

Whittlesea and Dorken's (1993) argument is predicated on the assumption that abstraction theorists claim knowledge is abstracted from the stimulus rather than from mental representations of experienced stimuli. Whittlesea and Dorken claimed, "The earliest, and still most prevalent, interpretation of implicitly acquired sensitivity to deep structure is that memory is in some way directly sensitive to abstract, structural properties of the environment" (p. 227). They argued that their episodic-processing account differs from other views on this fundamental issue: "The episodic-processing account shares similarities with all of the accounts of deep-structural sensitivity described in earlier sections, but differs

Robert C. Mathews and Lewis G. Roussel, Department of Psychology, Louisiana State University.

Correspondence concerning this article should be addressed to Robert C. Mathews, Department of Psychology, Louisiana State University, Baton Rouge, Louisiana 70803-5501.

from them in a fundamental way. According to this account, memory preserves experiences, not stimulus structures" (p. 230).

The assumption concerning direct abstraction from the stimulus rather than memory representations, which is central to Whittlesea and Dorken's (1993) thesis, might fit early notions about artificial grammar learning (e.g., Reber, 1967, 1976) but is not applicable to modern abstractionist theories. Contemporary theories of implicit abstraction are based on connectionist models (Cleeremans & McClelland, 1991; Dienes, 1992), induction theory (Holland, Holyoak, Nisbett, & Thagard, 1986; Mathews, 1991), or competitive chunking (Servan-Schreiber & Anderson, 1990). All of these theories assume that knowledge is abstracted from memory representations of experiences. Thus, any variables that affect what is encoded about events (like the variables manipulated by Whittlesea & Dorken) will affect the outcome of learning (what knowledge is abstracted). Models based on these theories are also highly sensitive to, and preserve much detail of, the individual episodes to which they were exposed. Consequently, all of Whittlesea and Dorken's results are consistent with these modern abstractionist theories.

Computational models of implicit learning have demonstrated that exemplar models and abstraction models are much more similar than they first appear (cf Estes, 1986). Only three minimal assumptions are necessary to convert a purely episodic memory system to one that abstracts structure directly from experience. They are as follows:

1. Memories of experiences are incomplete. That is, memories are mental representations that include a subset of the possible features of an experience that could have been encoded. In many theories of implicit learning, fragmentary memories of exemplars serve as the building blocks for representing the structure of related stimuli (e.g., Dulany, Carlson, & Dewey, 1984; Mathews, 1990, 1991; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990).

2. Some experiences are better remembered than others without conscious decisions to adjust their memorability (strength).

3. Differences in memorability of episodes are related to either frequency of encountering similar episodes (Servan-Schreiber & Anderson, 1990) or usefulness of these memories for guiding behavior (Druhan & Mathews, 1989; Roussel, Mathews, & Druhan, 1990).

We doubt that Whittlesea and Dorken (1993) would strongly object to any of these assumptions. Yet they miss the point that episodic systems with these features are capable of simultaneously abstracting regularities and preserving episodic information when it is needed (Holland et al., 1986; McClelland & Rumelhart, 1985). They insisted on dichotomizing abstract information of stimulus structure and information about individual episodes. They have said their experiments are designed to determine "whether implicit sensitivity to deep structure reflects automatic abstraction of the structure underlying stimuli or whether it instead reflects task-dependent encoding of particular experiences of stimuli" (p. 227).

Holland et al.'s (1986) induction theory illustrates how systems can use instance-derived rules to simultaneously abstract regularities and preserve local variations. Like Whittlesea and Dorken's (1993) episodic approach, partial memories of episodes can be used as "rules" to guide behavior in future situations (see Mathews, 1991, for applications of induction theory to artificial grammar learning). However, the induction theory (Holland et al., 1986) assumes that when memories (instance-based rules) are applied successfully in new situations, they are automatically strengthened. Thus, as experience accumulates, good rules (memories) that capture relevant aspects of stimulus structure are strengthened and eventually dominate behavior. Even if features of stimuli are randomly sampled to be included in particular memories (rules), the process of strengthening rules that successfully predict future outcomes eventually abstracts stimulus structure (Roussel et al., 1990). This automatic tuning or strengthening of fragmentary traces, which are good predictors of future outcomes (see Mathews, 1991), is the passive abstraction procedure that Whittlesea and Dorken denied. It "weights or selects stimulus properties in terms of their relevance to the set as a whole" (Whittlesea & Dorken, 1993, p. 227). Rather than being a degenerate case of inefficient encoding (Whittlesea & Dorken, 1993, pp. 228–229), such automatic detection and abstraction of stimulus structure may be an important key to the power of human thinking.

Another important difference between the two theories is that induction theory does not continue adding new rules (memories) to the system after each new experience. New rules are added to the system only when needed to improve accuracy, that is, when predictions from current rules fail. Adding new rules only when the system fails avoids the accumulation of mud that would otherwise make the system much less efficient (see Holland et al., 1986).

Encoding sets limits on what can be learned from past experiences. People will never discover regularities involving features that were not encoded. However, contrary to the view of Whittlesea and Dorken (1993), encoding tasks do not exercise complete control over learning about regularities across features that were encoded in memories of past experiences. In other words, people's minds are capable of noticing regularities that they were not looking for. Such automatic use of regularities in stimuli, often without conscious awareness that one is doing so, is the hallmark of implicit learning (e.g., Berry & Broadbent, 1984; Lewicki, 1986; Nissen & Bullemer, 1987; Polanyi, 1969; Reber, 1989). Perkins (1981) considered the ability to discover regularities that one was not looking for to be one of the three key features that enable creativity.

Contrary to the pure episodic view expressed by Whittlesea and Dorken (1993), there is considerable evidence suggesting that memory traces dynamically interact rather than being stored as independent episodes in memory. For example, the thousands of experiments demonstrating proactive or retroactive interference show that traces interact despite task demands to keep episodes distinct. In proactive interference experiments, subjects lose item information (they forget particular list items), but they gain collective information about list structure (e.g., they recognize that all of the lists contained fruits). Other examples of trace interactions that have been well documented in the literature include automatic integration of memorized sentences following a common theme (e.g., Bransford & Franks, 1972) and unintentional infusion of information provided after an event with memory of the original event (Loftus & Palmer, 1974). Interactions of traces in memory are adaptive because the primary function of memory is to make predictions about actions in future situations, not to serve as a diary of life experiences. Whether such interactions occur in storage or at retrieval, they serve the function of implicitly abstracting stimulus structure (Medin & Florian, 1992).

One final problem with purely episodic theory concerns automatic selectivity or increased attention to relevant features of stimuli. In situations where subjects are induced to encode and remember all the letters in artificial grammar strings, subjects appear to focus on certain typical features in recognizing strings (Mathews et al., 1989). Thus, selectivity to cues that are typical of sets of experienced exemplars appears to occur automatically even when subjects are attempting to remember the items as single unique instances. We believe that the learning process at work here is identical to what happens in proactive interference experiments. More frequent or common sets of features are better remembered and stand out as characteristics of valid strings.

Ironically, the ultimate objective of Whittlesea and Dorken (1993) and earlier work on exemplar models (Brooks, 1978; Brooks & Vokey, 1991; Vokey & Brooks, 1992) is the same as ours. We all agree that the distinction between thinking and memory is an artificial one. People think (recognize, categorize, choose, etc.) with their memories. Thus, research on memory, learning, and thinking needs to be integrated. All of the work on episodic memory is highly relevant to categorization and implicit learning. We argue that (implicit) learning processes operate outside awareness, altering the weights of memories with the end result that people learn to pay most attention to relevant aspects of stimulus structure. Abstraction of stimulus structure and memory of episodes are part of the same process of adaptive knowledge acquisition.

References

Aha, D. W., and Goldstone, R. L. (1990). Learning attribute relevance in context in instance-based learning algorithms. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (pp. 141–148). Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Berry, D. C., and Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology, 36A,* 209–231.

Bransford, J. D., and Franks, J. J. (1972). The abstraction of linguistic ideas: A review. *Cognitive Psychology, 2,* 331–350.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Brooks, L. R., and Vokey, J. R. (1991). Abstract analogies and abstracted grammars: A comment on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General, 120,* 316–320.

Cleeremans, A., and McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120,* 235–253.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240–247.

Dienes, Z. (1992). Connectionist and memory array models of artificial grammar learning. *Cognitive Science, 16,* 41–79.

Druhan, B., and Mathews, R. C. (1989). THIYOS: A classifier system model of implicit knowledge of artificial grammars. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 66–73). Hillsdale, NJ: Erlbaum.

Dulany, D. E., Carlson, R. A., and Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General, 113,* 541–555.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18,* 500–549.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Lenat, D. B. (1983). The role of heuristics in learning by discovery: Three case studies. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 243–306). Palo Alto, CA: Tioga.

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 135–146.

Loftus, E. F., and Palmer, J. C. (1974). Reconstruction of automobile destruction. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.

Mathews, R. C. (1990). Abstractness of implicit grammar knowledge: Comments on Perruchet and Pacteau's analysis of synthetic grammar learning. *Journal of Experimental Psychology: General, 119,* 412–416.

Mathews, R. C. (1991). The forgetting algorithm: How fragmentary knowledge of exemplars can abstract knowledge. *Journal of Experimental Psychology: General, 120,* 117–119.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1083–1100.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159–188.

Medin, D. L., & Florian, J. E. (1992). Abstraction and selective coding in exemplar-based models of categorization. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes, Vol. 2* (pp. 207–234). Hillsdale, NJ: Erlbaum.

Nissen, M. J., and Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1–32.

Perkins, D. N. (1981). *The mind's best work.* Cambridge, MA: Harvard University Press.

Perruchet, P., and Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119,* 264–275.

Polanyi, M. (1969). *Knowing and being.* Chicago: University of Chicago Press.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 5,* 855–863.
Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional sets. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 88–94.
Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235.
Roussel, L. G., Mathews, R. C., & Druhan, B. B. (1990). Rule induction and interference in the absence of feedback: A classifier system model. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (pp. 733–740). Hillsdale, NJ: Erlbaum.
Servan-Schreiber, E., and Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 592–608.
Smith, E. E., and Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.
Vokey, J. R., and Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 328–344.
Whittlesea, B. W., and 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.

Received January 6, 1993
Revision received February 19, 1993
Accepted February 25, 1993 ■