The Interaction of the Explicit and the Implicit in Skill Learning: A Dual-Process Approach

Ron Sun
Rensselaer Polytechnic Institute

Paul Slusarz
University of Missouri—Columbia

Chris Terry
University of Alabama

This article explicates the interaction between implicit and explicit processes in skill learning, in contrast to the tendency of researchers to study each type in isolation. It highlights various effects of the interaction on learning (including synergy effects). The authors argue for an integrated model of skill learning that takes into account both implicit and explicit processes. Moreover, they argue for a bottom-up approach (first learning implicit knowledge and then explicit knowledge) in the integrated model. A variety of qualitative data can be accounted for by the approach. A computational model, CLARION, is then used to simulate a range of quantitative data. The results demonstrate the plausibility of the model, which provides a new perspective on skill learning.

The role of implicit learning in skill acquisition and the distinction between implicit and explicit learning have been widely recognized in recent years (see, e.g., Cleeremans, Destrebecqz, & Boyer, 1998; Proctor & Dutta, 1995; Reber, 1989; Seger, 1994; Stadler & Frensch, 1998). However, although implicit learning has been actively investigated, complex and multifaceted interaction between the implicit and the explicit and the importance of this interaction have not been universally recognized (though with a few notable exceptions even early on, e.g., Mathews et al., 1989).[1] Similar oversight is also evident in computational simulation models of implicit learning (with a few exceptions such as Cleeremans, 1993, and Sun, Merrill, & Peterson, 2001).

Likewise, in the development of cognitive architectures (e.g., Anderson, 1983, 1993; Meyer & Kieras, 1997; Newell, 1990), the distinction between procedural and declarative knowledge has been adopted by many (Anderson, 1983, 1993). The distinction maps roughly onto that between explicit and implicit knowledge, because procedural knowledge is generally inaccessible whereas declarative knowledge is generally accessible and thus explicit. However, the focus has been mostly on top-down models (i.e., learning first explicit knowledge and then implicit knowledge); the

bottom-up direction (i.e., learning first implicit knowledge and then explicit knowledge or learning both in parallel) has been largely ignored, paralleling the related neglect of the interaction of explicit and implicit processes in the skill acquisition literature.

Despite such problems, it has been gaining recognition that it is difficult to find a situation in which only one type of learning is engaged (Mishkin, Malamut, & Bachevalier, 1984; Reber, 1989; Seger, 1994; Sun et al., 2001; Willingham, 1998; but see Lewicki, Czyzewska, & Hoffman, 1987). Our review of existing data (see the Existence of Interaction section) has indicated that although one can manipulate conditions to emphasize one or the other type, in most situations, both types of learning are involved, with varying amounts of contributions from each.

Many issues arise that we need to examine to better understand the interaction between implicit and explicit processes:

> How can we capture implicit and explicit processes in computational terms?

> How do the two types of knowledge develop alongside each other and influence each other's development (e.g., top down versus bottom up)?

> How can bottom-up learning be realized computationally?

> How do the two types of knowledge interact during skilled performance, and what is the impact of that interaction on performance?[2]

In the following section, Existence of Interaction, we present evidence that points to a complex, multifaceted interaction be-

[1] By the explicit, we mean processes involving some form of generalized (or generalizable) knowledge that is consciously accessible.

[2] For example, the synergy of the two may result, as described in Sun et al. (2001).

tween implicit and explicit processes. On the basis of this information, in the section A Model, we present a theoretical model of skill learning incorporating both types of processes and emphasizing bottom-up learning. In the section Some Details of the Model, we present details of computational implementation of the model. Then, in the Analysis of Interaction section, we show how the model can account for or predict phenomena in skill learning qualitatively. We perform a detailed match of human and model data in the Simulation of Human Skill Learning Data section. Finally, in the General Discussion, we discuss a number of issues and examine existing models of skill learning. The discussion points to the uniqueness of the model and highlights the fact that it synthesizes a variety of data and provides a coherent, plausible interpretation based on the implicit–explicit interaction.

Existence of Interaction

Mathews et al. (1989) proposed that "subjects draw on two different knowledge sources to guide their behavior in complex cognitive tasks; one source is based on their explicit conceptual representation; the second, independent source of information is derived from memory-based processing, which automatically abstracts patterns of family resemblance through individual experiences" (p. 1098). Likewise, Sun (1994) pointed out that "cognitive processes are carried out in two distinct levels with qualitatively different mechanisms" (p. 44), although "the two sets of knowledge may overlap substantially" (p. 44). Reber (1989) pointed out that nearly all complex skills in the real world (as opposed to laboratory settings) involve a mixture of explicit and implicit processes interacting in complex ways (Mishkin et al., 1984; Willingham, 1998). The same point can also be found in Sun et al. (2001). Even commonly used implicit learning tasks likely involve a combination of explicit and implicit learning processes (Sun et al., 2001; see also the simulations in the Simulation of Human Skill Learning Data section).

Interaction of implicit and explicit knowledge has been demonstrated in the implicit learning literature. A brief review of three common tasks of implicit learning is as follows. The *serial reaction time* (SRT) tasks (Nissen & Bullemer, 1987) tapped subjects' ability to learn a repeating sequence. On each trial, one of the four lights on a screen was illuminated, and subjects were to press the button corresponding to the illuminated light. The lights were shown in a repeating 10-trial sequence. It was found that there was a significant reduction in response time to repeating sequences, attributed to the learning of the sequence. However, subjects were sometimes unaware that a repeating sequence was involved. Amnesic patients with Korsakoff's syndrome were found to undergo similar learning. Likewise, *dynamic control* (DC) tasks (Berry & Broadbent, 1988) examined subjects' ability to learn a relation between the input and output variables of a controllable system through interacting with the system dynamically. Subjects were required to control an output variable by manipulating an input variable. Although they often did not recognize the relation between input and output explicitly, subjects reached a high level of performance. Similarly, in *artificial grammar learning* tasks (Reber, 1989), subjects were presented strings of letters generated in accordance with a finite state grammar. After memorization, subjects showed an ability to distinguish between new strings that conformed to the grammar and those that did not. Although sub-

jects might not be explicitly aware of the underlying grammar (barring some fragmentary knowledge), they performed significantly beyond the chance level. In all, these tasks share the characteristic of heavily involving implicit processes,[3] which is also shared by some other tasks (such as some concept learning, automatization, and conditioning tasks).

Various demonstrations of interaction exist in which these tasks were used. For instance, Stanley, Mathews, Buss, and Kotler-Cope (1989) and Berry (1983) found that under some circumstances concurrent verbalization could help to improve subjects' performance in DC tasks. Reber and Allen (1978) similarly showed in artificial grammar learning that verbalization could help. Ahlum-Heath and DiVesta (1986) also found that verbalization led to better performance in learning the Tower of Hanoi task. In the same vein, although no verbalization was used, Willingham, Nissen, and Bullemer (1989) showed that those subjects who demonstrated more awareness of the regularities in the stimuli (i.e., those who had more explicit knowledge) performed better in an SRT task, which likewise seemed to show the helpful effects of explicit processes.

However, as Reber (1976, 1989) pointed out, verbalization and the resulting explicit knowledge might also hamper (implicit) learning under some circumstances. This may happen when too much verbalization induces an overly explicit learning mode in subjects performing a task that is not suitable for learning in an explicit way (e.g., when learning a complex artificial grammar). Similarly, in a minefield navigation task, Sun, Merrill, and Peterson (1998, 2001) reported that too much verbalization was detrimental to performance. Roussel (1999) showed that explicit reflection sometimes hurt performance. Dulaney, Carlson, and Dewey (1984) showed that correct and potentially useful explicit knowledge, when given at an inappropriate time, could hamper learning.

As variously demonstrated by Berry and Broadbent (1984, 1988), Stanley et al. (1989), and Reber, Kassin, Lewis, and Cantor (1980), verbal instructions (given prior to learning) could facilitate or hamper task performance too. One type of instruction was to encourage explicit search for regularities in stimuli that might aid in task performance. For example, Reber et al. (1980) found that depending on ways stimuli were presented, explicit search might help or hamper performance. Owen and Sweller (1985) and Schooler, Ohlsson, and Brooks (1993) found that explicit search hindered learning. Another type of instruction was explicit how-to instructions that specifically informed subjects how the tasks should be performed (including providing detailed information concerning regularities in stimuli). Stanley et al. (1989) and Berry and Broadbent (1988) found that this type of instruction helped to improve performance significantly. See also Boyd and Weistein (2001).

There is evidence that implicit and explicit knowledge may develop independently under some circumstances. Willingham et

---

[3] Some have disputed the existence of implicit processes, on the basis of the imperfection and incompleteness of tests for explicit knowledge (e.g., Shanks & St. John, 1994). We do not wish to engage in a methodological debate here. We refer readers to Sun et al. (2001) for relevant arguments in terms of the overwhelming evidence for the distinction between implicit and explicit processes. See also the *Potential Controversies* section.

al. (1989) reported data that were consistent with the parallel development of implicit and explicit knowledge. By using two different measures for assessing the two types of knowledge respectively, they compared the time course of implicit and explicit learning. It was shown that implicit knowledge might be acquired in the absence of explicit knowledge and vice versa. The data ruled out the possibility that one type of knowledge was always preceded by the other type. Rabinowitz and Goldberg (1995) similarly demonstrated parallel development of procedural and declarative knowledge in an alphabetic arithmetic task.

However, a subject's performance typically improves faster than explicit knowledge that is verbalized by the subject. For instance, in DC tasks, although the performance of subjects quickly rose to a high level, their verbal knowledge improved far slower: The subjects could not provide usable verbal knowledge until near the end of their training (Stanley et al., 1989). Bowers, Regehr, Balthazard, and Parker (1990) also showed delayed acquisition of explicit knowledge. When subjects were given patterns to complete, they first showed implicit recognition of proper completion. Their implicit recognition improved over time until eventually an explicit recognition was achieved. This phenomenon was also demonstrated by Reber and Lewis (1977) in artificial grammar learning. Sun et al. (1998, 2001) focused on this type of learning in a minefield navigation task. In all of these cases, because explicit knowledge lags behind but improves with implicit knowledge, explicit knowledge is in a way "extracted" from implicit knowledge of these tasks (as suggested by Seger, 1994; Stanley et al., 1989; Sun, 1997). That is, learning of explicit knowledge is through the (delayed) explication of implicit knowledge.

Given the voluminous evidence of complex interaction between implicit and explicit processes, the key questions now are: (a) how do we account for the interaction and (b) how do we explain the demonstrated effects of interaction?

A Model

Below, we discuss a model that incorporates both implicit and explicit processes (Sun, 1997; Sun et al., 2001).

Representation

Let us first consider the representations of a possible model incorporating the distinction between implicit and explicit processes. There is evidence that human action decision making in skilled performance is a largely implicit process (e.g., Ben-Zur, 1998; Reber, 1989; Seger, 1994). We note that the inaccessible nature of implicit knowledge is suitably captured by subsymbolic distributed representations provided by a backpropagation network (Rumelhart, McClelland, & the PDP Research Group, 1986). This is because representational units in a distributed representation (e.g., in the hidden layer of a backpropagation network) are capable of accomplishing tasks but are subsymbolic and generally not individually meaningful (see Rumelhart et al., 1986; Sun, 1994). This characteristic of distributed representation accords well with the inaccessibility of implicit knowledge. (However, it is generally not the case that distributed representations are not accessible at all, but they are definitely less accessible and not as direct and immediate as localist–symbolic representations. Distributed representations may be accessed through indirect, transformational processes.) In contrast, explicit knowledge may be best captured in

computational modeling by a symbolic or localist representation (Clark & Karmiloff-Smith, 1993), in which each unit is more easily interpretable and has a clearer conceptual meaning. This characteristic captures the property of explicit knowledge being accessible and manipulable (Smolensky, 1988; Sun, 1994). This radical difference in the representations of the two types of knowledge leads to a two-level model, CLARION (which stands for Connectionist Learning with Adaptive Rule Induction ONline; initially proposed in Sun, 1997, and Sun, 1999), whereby each "level" using one kind of representation captures one corresponding type of process (either implicit or explicit). Sun (1994, 1995, 1999), Dreyfus and Dreyfus (1987), and Smolensky (1988) presented relevant theoretical arguments for such two-level models.

At each level of the model, there may be multiple modules (both *action-centered* modules and *non-action-centered* modules; Moscovitch & Umilta, 1991; Schacter, 1990). At the bottom level, action-centered knowledge is highly modular: A number of backpropagation networks coexist, each adapted to a specific modality, task, or input stimulus type. This is consistent with the well-known modularity claim (Cosmides & Tooby, 1994; Fodor, 1983; Karmiloff-Smith, 1986) and is also similar to Shallice's (1972) idea of a multitude of "action systems" competing with each other.

The non-action-centered modules (at both levels) represent more static, declarative, and generic types of knowledge. The knowledge there includes what is commonly referred to as *semantic* memory (i.e., general knowledge about the world in a conceptual, symbolic form; Quillian, 1968; Tulving, 1972). We do not get into these modules in this work.

The reason for having both action-centered and non-action-centered modules (at each level) is because, as it should be obvious, action-centered knowledge (roughly, procedural knowledge) is not necessarily inaccessible (directly) and non-action-centered knowledge (roughly, declarative knowledge) is not necessarily accessible (directly). (Although it has been argued by some that all procedural knowledge is inaccessible and all declarative knowledge is accessible, such a clean mapping of the two dichotomies is untenable in our view.)

Learning

Only the learning of action-centered knowledge (which is most relevant to skill learning) is dealt with here. The learning of implicit action-centered knowledge at the bottom level can be done in a variety of ways consistent with the nature of distributed representations. In the learning settings in which correct input–output mappings are available, straight backpropagation (a supervised learning algorithm) can be used for each network (Rumelhart et al., 1986). Such supervised learning procedures require the a priori determination of a uniquely correct output for each input. In the learning settings in which there is no input–output mapping externally provided, reinforcement learning can be used (Watkins, 1989), especially Q-learning (Watkins, 1989) implemented with backpropagation networks (see the Same Details of the Model section). Such learning methods are cognitively justified: Shanks (1993) showed that a simple type of skill learning was best captured by associative models (i.e., neural networks), compared with a variety of rule-based models. Cleeremans (1997) argued that implicit learning could not be captured by symbolic models but could be captured by neural networks.

The action-centered explicit knowledge at the top level can also be learned in a variety of ways in accordance with the localist–symbolic representation used. Because of the representational characteristics, one-shot learning based on hypothesis testing (Bruner, Goodnow, & Austin, 1956; Busemeyer & Myung, 1992; Mitchell, 1998; Nosofsky, Palmeri, & McKinley, 1994; Sun et al., 1998, 2001) is needed. With such learning, individuals explore the world and dynamically acquire representations and modify them as needed, reflecting the dynamic, ongoing nature of skill learning (Heidegger, 1927/1962; Sun et al., 1998, 2001; Vygotsky, 1962). In so doing, the implicit knowledge already acquired in the bottom level can be used in learning explicit knowledge (through bottom-up learning; Sun et al., 1998, 2001). The basic idea of bottom-up learning is as follows: If an action chosen (by the bottom level) is successful (i.e., it satisfies a certain criterion), then a rule is extracted and set up at the top level. Then, in subsequent interactions with the world, the rule is refined by considering the outcome of applying the rule: If the outcome is successful, the conditions of the rule may be generalized to make it more universal; if the outcome is not successful, then the conditions of the rule should be made more specific and exclusive of the current case. This is an online version of hypothesis testing processes studied (in different contexts) by, for example, Bruner et al. (1956) and Nosofsky et al. (1994). Other types of learning are also possible, such as hypothesis testing without the help of the bottom level (for capturing independent rule learning, as discussed in the previous section).

Thus, the bottom level develops implicit skills using the *Q-learning-backpropagation* algorithm (QBP, for short; to be explained later), whereas the top level extracts explicit rules using the *rule-extraction-refinement* algorithm (RER, for short; to be explained later) and possibly others. The learning difference of the two levels is somewhat analogous to that between the corticostriatal "habit" system and the corticolimbic "memory" system as proposed by Mishkin et al. (1984). See the next section for implementational details.

Here is a list of the basic theoretical hypotheses of the model (see Figure 1):

Representational difference: The two levels use two different types of representations and thus have different degrees of accessibility.

Learning difference: Different learning methods are used for the two levels.

Bottom-up learning: When there is no sufficient a priori knowledge available, learning is bottom up.

Some Details of the Model

As proposed in Sun (1997), a high-level description of the operation of CLARION (see Figure 1) is as follows:

1. Observe the current state x.

2. Compute in the bottom level the "value" of each of the possible actions (a_is) associated with the state x: $Q(x, a_1)$, $Q(x, a_2)$, . . . , $Q(x, a_n)$.

3. Find out all the possible actions ($b_1, b_2, . . . , b_m$) at the top level, based on the state x (which goes up from the bottom level) and the existing action rules in place at the top level.

4. Choose an appropriate action a stochastically, based on combining the values of a_is (at the bottom level) and b_js (which are sent down from the top level).

5. Perform the action a, and observe the next state y and (possibly) the reinforcement r.

6. Update the bottom level in accordance with the Q-learning-backpropagation algorithm (or QBP; to be detailed later), based on the feedback information.

7. Update the top level using the rule-extraction-refinement algorithm (or RER, for constructing, refining, and deleting rules; to be detailed later).
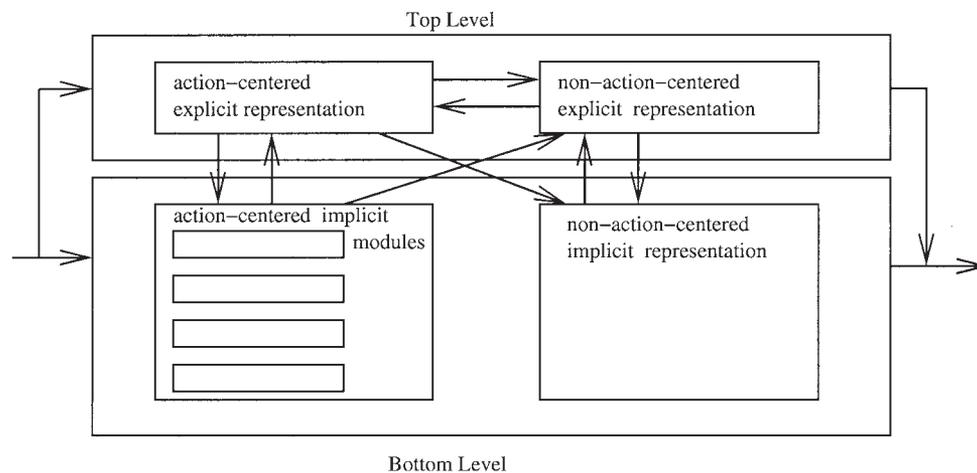

Figure 1. The CLARION architecture.

8. Go back to Step 1.

The following implementational details may be skipped on first reading. The reader may choose to go directly to the example in the *An Example* section.

The Bottom Level

In the bottom level, a Q-value is an evaluation of the "quality" of an action in a given state: $Q(x, a)$ indicates how desirable action a is in state x. We can choose an action based on Q-values. At each step, given the input x, we first compute the Q-values for all the possible actions (i.e., $Q(x, a)$ for all as). We then use the Q-values to decide probabilistically on an action to be performed, which is done by a Boltzmann distribution of Q-values:

$$p(a|x) = \frac{e^{Q(x,\, a)/\alpha}}{\sum_i e^{Q(x, a_i)/\alpha}}. \qquad (1)$$

Here α controls the degree of randomness (temperature) of the decision-making process.[4]

A four-layered connectionist network is used for implementing Q-learning (see the bottom half of Figure 2), in which the first three layers form a (either recurrent or feedforward) backpropagation network for computing Q-values and the fourth layer (with only one node) performs the aforementioned Boltzmann decision making. The network is internally subsymbolic and uses implicit representation in accordance with our previous considerations of representational forms. The output of the third layer (i.e., the output layer of the backpropagation network) indicates the Q-value of each action (represented by an individual node), and the node in the fourth layer determines probabilistically the action to be performed based on the Boltzmann distribution.[5]

To acquire the Q-values, supervised and/or reinforcement learning methods may be applied (as mentioned earlier). Particularly

applicable is the Q-learning algorithm (Watkins, 1989; a reinforcement learning algorithm). In the algorithm, $Q(x, a)$ estimates the maximum (discounted) total reinforcement that can be received from the current state x on

$$\max\left(\sum_{i=0}^{\infty} \gamma^i r_i\right), \qquad (2)$$

where γ is a discount factor that favors reinforcement received sooner relative to that received later and r_i is the reinforcement received at step i (which may be none). The updating of $Q(x, a)$ is based on

$$\Delta Q(x, a) = \alpha(r + \gamma e(y) - Q(x, a)), \qquad (3)$$

where γ is a discount factor, y is the new state resulting from action a in state x, and $e(y) = \max_b Q(y, b)$. Thus, the updating is based on the temporal difference in evaluating the current state and the action chosen: In the above formula, $Q(x, a)$ estimates, before action a is performed, the maximum (discounted) total reinforcement to be received if action a is performed, and $r + \gamma e(y)$ estimates the maximum (discounted) total reinforcement to be received, after action a is performed; so, their difference (the temporal difference in evaluating an action) enables the learning of Q-values that approximate the maximum (discounted) total reinforcement. Using Q-learning allows sequential behavior to emerge.

Applying Q-learning, the training of the backpropagation network is based on minimizing the following error at each step:

$$\text{err}_i = \begin{cases} r + \gamma e(y) - Q(x, a) & \text{if } a_i = a \\ 0 & \text{otherwise} \end{cases}, \qquad (4)$$

where i is the index for an output node representing the action a_i. On the basis of the above error measures, the backpropagation algorithm is applied to adjust internal weights (which are randomly initialized before training). When a correct input–output mapping is available for a step, the above error measure can be simplified to the desired output minus the actual output: $r - Q(x, a)$. This learning process (using Q-learning plus backpropagation) enables the development of implicit skills potentially solely based on exploring the world on a continuous and ongoing basis (without a priori knowledge).

The Top Level

In the top level (see Figure 2), explicit knowledge is captured in a propositional rule form. To facilitate correspondence with the bottom level (and to encourage uniformity and integration; Clark & Karmiloff-Smith, 1993), a localist connectionist network is used for implementing these rules (e.g., Sun, 1992), in accordance with our previous considerations. Basically, we translate the structure of
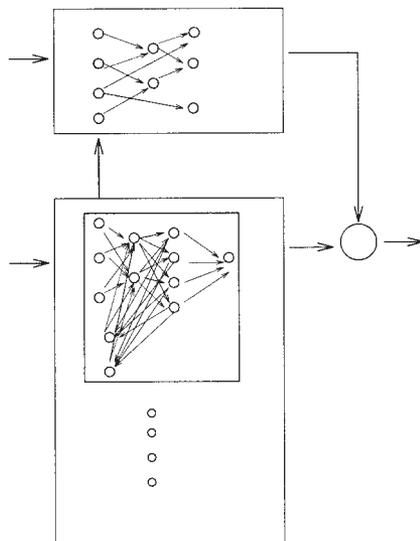

Figure 2. The implementation of CLARION (the action-oriented modules). The top level contains localist encoding of propositional rules. The bottom level contains backpropagation networks. The information flows are indicated with arrows.

---

[4] This method is also known as Luce's choice axiom (Watkins, 1989). It is cognitively well justified and found to match psychological data in a variety of domains.

[5] The calculation of Q-values for the current input with respect to all the possible actions is done in a connectionist fashion through parallel spreading activation and is thus highly efficient. Such spreading of activation is assumed to be implicit as, for example, in Hunt and Lansman (1986), Cleeremans and McClelland (1991), and G. Bower (1996).

a set of rules into that of a network. Assume that an input state x is made up of a number of dimensions (e.g., $x_1, x_2, \ldots, x_n$). Each dimension can have a number of possible values (e.g., $v_1, v_2, \ldots, v_m$).[6] Rules are in the following form: *current-state-specification* → *action,* where the left-hand side is a conjunction of individual elements, each of which refers to a dimension x_i of the input state x, specifying a value or a value range (i.e., $x_i \in [v_i, v_i]$ or $x_i \in [v_{i1}, v_{i2}]$), and the right-hand side is an action recommendation a.[7] Each element in the left-hand side is represented by an individual node. For each rule, a set of links are established, each of which connects a node representing an element in the left-hand side of a rule to the node representing the conclusion in the right-hand side of the rule. Further implementational details of rule representations are irrelevant to the discussion that follows and thus omitted (Sun & Peterson, 1998a, contains the full details).

Among other algorithms that we developed, we devised the rule-extraction-refinement algorithm (RER) for learning rules using information in the bottom level (to capture the bottom-up learning process):

1. Update the rule statistics (to be explained later).
2. Check the current criterion for rule extraction, generalization, and specialization:
   2.1. If the result is successful according to the current criterion and there is no rule matching the state and the action taken, then perform *extraction* of a new rule: state → action. Add the extracted rule to the rule network.
   2.2. If the result is unsuccessful according to the current criterion, revise all the matching rules through *specialization:*
       2.2.1. Remove the matching rules from the rule network.
       2.2.2. Add the revised (specialized) versions of the rules to the rule network.
   2.3. If the result is successful according to the current criterion, then generalize the matching rules through *generalization:*
       2.3.1. Remove the matching rules from the rule network.
       2.3.2. Add the generalized rules to the rule network.[8]

Let us discuss the details of the operations used in the above algorithm and the criteria measuring whether a result is successful (which are used in deciding whether to apply some of these operators). The criteria were mostly determined by an *information gain* (IG) measure that compares the quality of two candidate rules.

To calculate the IG measure, we do the following. At each step, we examine the following information: (x, y, r, a), where x is the state before action a is performed, y is the new state after an action a is performed, and r is the reinforcement received after action a. On the basis of this information, we update (in Step 1 of the above algorithm) the positive and negative match counts for each rule condition and each of its minor variations (i.e., the rule condition $\pm$ 1 possible value in one of the input dimensions), both of which are denoted as C, with regard to the action a performed: That is, $\text{PM}_a(C)$ (i.e., positive match) equals the number of times that an input matches the condition C, action a is performed, and the result is positive; $\text{NM}_a(C)$ (i.e., negative match) equals the number of times that an input matches the condition C, action a is performed, and the result is negative. Positivity–negativity may be determined on the basis of domain-specific information such as feedback from environments or information from the bottom level (see, e.g., Sun & Peterson, 1998a, 1998b). Each statistic is updated with the following formula: stat := stat + 1 (where *stat* stands for

PM or NM); at the end of each episode, it is discounted by stat := stat × 0.90.[9] On the basis of these statistics, we calculate the IG measure; that is,

$$IG(A, B) = \log_2 \frac{\text{PM}_a(A) + 1}{\text{PM}_a(A) + \text{NM}_a(A) + 2}$$
$$- \log_2 \frac{\text{PM}_a(B) + 1}{\text{PM}_a(B) + \text{NM}_a(B) + 2}, \quad (5)$$

where A and B are two different conditions that lead to the same action a. The measure compares essentially the percentage of positive matches under different conditions A and B (with the Laplace estimator; Lavrac & Dzeroski, 1994). If A can improve the percentage to a certain degree over B, then A is considered better than B. In the algorithm, if a rule is better compared with the corresponding *match-all* rule (i.e., the rule with the same action but with the condition that matches all possible input states), then the rule is considered successful.[10]

We decide on whether to construct a rule on the basis of a simple criterion that is determined by the current step (x, y, r, a):

> *Extraction:* If the current step is positive and if there is no rule that covers this step in the top level, set up a rule $C \rightarrow a$, where C specifies the values of all the input dimensions exactly as in x.[11]

However, the criterion for applying the generalization and specialization operators is based on the aforementioned IG measure. Generalization amounts to adding an additional value to one input dimension in the condition of a rule, so that the rule will have more opportunities of matching input, and specialization amounts to removing one value from one input dimension in the condition of a rule, so that it will have less opportunities of matching input. Here are the detailed descriptions of these two operators:

> *Generalization:* If IG(C, *all*) > threshold1 and $\max_{C'}$ IG(C', C) ≥ 0, where C is the current condition of a matching rule, *all* refers to the corresponding match-all rule (with regard to the same action specified by the rule), and C' is a modified condition such that $C' = C$ plus one value (i.e., C' has one

---

[6] Each dimension is either ordinal (discrete or continuous) or nominal. In the following discussion, we focus on ordinal values; nominal values can be handled similarly. A binary dimension is a special case of a discrete ordinal dimension.

[7] Alternatively, rules can be in the forms of *current-state-specification* → *action new-state* or *current-state-specification action* → *new-state.*

[8] We also merge rules whenever possible: If rule extraction, generalization, or specialization is performed at the current step, check to see if the conditions of any two rules are close enough and thus if the two rules may be combined: If one rule is covered completely by another, put it on the children list of the other. If one rule is covered by another except for one dimension, produce a new rule that covers both.

[9] The results are time-weighted statistics, which are useful in nonstationary situations.

[10] This is a commonly used method, especially in inductive logic programming, and well justified on the empirical ground. See, for example, Lavrac and Dzeroski (1994).

[11] Potentially, we might use attention to focus on fewer input dimensions, although such attention is not part of the current model.

more value in one of the input dimensions; i.e., if the current rule is successful and a generalized condition is potentially better), then set $C'' = \text{argmax}_{C'} \text{IG}(C', C)$ as the new (generalized) condition of the rule. Reset all the rule statistics. Any rule covered by the generalized rule will be placed on its children list.[12]

Specialization: If $\text{IG}(C, all) < threshold2$ and $\max_{C'} \text{IG}(C', C) > 0$, where C is the current condition of a matching rule, *all* refers to the corresponding match-all rule (with regard to the same action specified by the rule), and C' is a modified condition such that $C' = C$ minus one value (i.e., C' has one less value in one of the input dimensions; i.e., if the current rule is unsuccessful, but a specialized condition is better), then set $C'' = \text{argmax}_{C'} \text{IG}(C', C)$ as the new (specialized) condition of the rule.[13] Reset all the rule statistics. Restore those rules on the children list of the original rule that are not covered by the specialized rule and the other existing rules. If specializing the condition makes it impossible for a rule to match any input state, delete the rule.

In addition, the density parameter determines the minimum frequency of repetition necessary to keep a rule. For example, if this parameter is set at $1/n$, then at least one encounter of an input that matches a rule is necessary every n trials to maintain the rule. Otherwise, the rule will not be maintained (and will thus be deleted).

A variation of the above algorithm is independent rule learning (IRL) without using the bottom level for the initial extraction. In IRL, either completely randomly or in a particular domain-specific order (see examples later in the Simulation of Human Skill Learning Data section), rules of various forms are independently generated (hypothesized and wired up) at the top level. Then, these rules are tested through experience using the IG measure. When appropriate, generalization and specialization can also be performed on the basis of IG.

Given the explicit knowledge in the form of rules, a variety of explicit operations can be performed at the top level. These operations include backward chaining reasoning, forward chaining reasoning, and counterfactual reasoning (these operations will not be needed for the experiments reported in this article; see Sun et al., 2001, for details).

Combining the Two Levels

In the overall algorithm, Step 4 is for making the final decision on which action to take at each step by incorporating outcomes from both levels. Specifically, we combine the corresponding values for an action from the two levels by a weighted sum; that is, if the top level indicates that action a has an activation value v_a (which should be 0 or 1 as rules are binary) and the bottom level indicates that a has an activation value q_a (the Q-value), then the final outcome is $w_1 \times v_a + w_2 \times q_a$. Stochastic decision making with Boltzmann distribution (based on the weighted sums) is then performed to select an action out of all the possible actions. (w_1 and w_2 may be preset or automatically determined through probability matching.)[14] Stochastic decision making allows different operational modes: for example, relying only on the bottom level, relying only on the top level, or combining the outcomes from both levels and weighing them differently. Figure 2 shows the implementation of the two levels of the model.

An Example

Let us examine the simulation of a hypothetical SRT task. In an SRT task, a repeating sequence of X marks, each in one of the four possible positions, is presented to subjects (Curran & Keele, 1993). The subjects are instructed to press the key corresponding to the position of each X mark. Subjects learn to predict new positions on the basis of preceding positions through learning the sequential relations embedded in the sequence and thereby speed up their responses.

In the model, learning experience at the bottom level promotes the formation of implicit knowledge. Learning is done through iterative weight updating (the procedure of backpropagation was specified before). The resulting weights specify a function relating preceding positions (input) and the current position (output). For example, the iteratively updated weights may end up specifying the sequence 1 2 3 2 4 2. The sequence is specified in an implicit way (i.e., embedded in two layers of weights used for the sigmoidal computation within a backpropagation network).

The acquired sequential knowledge at the bottom level can lead to the extraction of explicit knowledge at the top level. The initial extraction step creates a rule that corresponds to the current input and the output (as determined by the bottom level). The generalization step adds more possible values (that may match new input) to the condition of the rule so that it may have more chances of matching new input. The specialization step, conversely, adds constraints to the rule (by removing possible matching values in the condition of the rule) to make it more specific and less likely to match new input. The applicability of these steps is determined on the basis of the IG measure described before. For example, the following rule may be initially extracted: $1\ 2\ 3\ 2 \rightarrow 4$. Generalization steps may lead to a simplified rule: $*\ 2\ 3\ 2 \rightarrow 4$ (provided that the IG measure calculated allows generalization, where * stands for "don't care"). Further generalization may lead to $*\ 2\ *\ 2 \rightarrow 4$, and so on. Continued revision (generalization and specialization) is likely to happen, as determined by the IG measure (which is in turn determined by the performance of the rule).

Analysis of Interaction

Below, we discuss a number of issues and phenomena observed or hypothesized regarding human skill learning relevant to the implicit–explicit interaction. The validity of our approach, as embodied in CLARION, lies in accounting for, or predicting, many such issues and phenomena.

Dissociation

Can implicit and explicit learning really be separated as two different processes empirically? Human skill learning data indicate that the answer to this question is yes (Cleeremans et al., 1998;

---

[12] The children list of a rule is created to keep aside and make inactive those rules that are more specific (thus fully covered) by the current rule. It is useful because if later on the rule is deleted or specialized, some or all of those rules on its children list may be reactivated if they are no longer covered.

[13] Clearly, we should have $threshold2 \leq threshold1$ to avoid oscillation.

[14] As shown by Willingham et al. (1989), explicit knowledge can influence skilled performance in humans.

Stadler & Frensch, 1998). Berry and Broadbent (1988) demonstrated this point using two DC tasks that differed in the degree to which the pattern of correct responding was salient to subjects.[15] Subjects in the two different conditions learned the tasks in different ways: Subjects in the nonsalient condition learned the task implicitly, whereas subjects in the salient condition learned the task more explicitly (as demonstrated by the difference in questionnaire scores). Lee (1995) showed a similar difference, but in relation to task complexity instead of saliency. A. Cohen, Ivry, and Keele (1990) described a similar situation in SRT tasks. When complex hierarchical relations were needed to predict next positions, subjects tended to use implicit learning, whereas explicit learning was more evident when simpler relations were involved. In artificial grammar learning tasks, there were also dissociations between learning simple (pairwise) relations and learning complex hierarchical relations (Mathews et al., 1989; Reber, 1989). Lee (1995) further demonstrated a reverse dissociation between implicit and explicit learning across two conditions: one with explicit instructions and the other without. With instructions, subjects did better on explicit tests than on implicit tests; without instructions, subjects did better on implicit tests than on explicit tests. There were also other corroborative findings to such dissociation results. For example, in learning physics, subjects could show different knowledge in different contexts: They showed incorrect knowledge while making explicit judgments; they nevertheless showed correct knowledge while actually performing a task (Seger, 1994).

Another line of evidence resulted from contrastive manipulations of implicit and explicit processes. First, through explicit search instructions, a more (or completely) explicit mode of learning may be used by subjects. The effect of such a mode change varies, depending on task difficulty (i.e., salience of stimulus materials). In the case of salient relations and/or regularities in stimuli, explicit search may be more successful and thus may improve performance (A. Cohen et al., 1990; Reber et al., 1980). In the case of nonsalient relations, explicit search may fail and thus lead to worsened performance (Berry & Broadbent, 1988; Owen & Sweller, 1985; see more discussion later). This contrast accentuates the differences between the two types of processes: Some tasks are more amenable to explicit processes than others. Second, through verbalization, a more explicit learning mode may likewise be used, as verbalization focuses subjects' attention explicitly on the relations and regularities in stimulus materials and thereby brings them out explicitly. Verbalization usually leads to comparable or improved performance (Gagne & Smith, 1962; Reber & Allen, 1978; Stanley et al., 1989; Sun et al., 1998) but may sometimes lead to worsened performance (Schooler et al., 1993; Sun et al., 2001). Third, through the use of dual-task conditions, a more implicit mode of learning may be attained. This is because, as has been argued in the literature, such manipulation affects explicit processes more than implicit processes (Cleeremans, 1993; Dienes & Berry, 1997; Hayes & Broadbent, 1988; Sun et al., 2001; Szymanski & MacLeod, 1996). Dual-task conditions often lead to worsened performance, which may mainly reflect the reduction of contributions from explicit processes (for alternative interpretations, see Frensch, Wenke, and Ruenger, 1999; Nissen & Bullemer, 1987; Stadler, 1995; more details later). Moreover, under dual-task conditions, the performance difference between subjects with explicit knowledge and those without may disappear (Curran & Keele, 1993). Contrasting these manipulations, we see the role played by explicit processes: Enhancing explicit processes often

helps performance, and weakening explicit processes hinders performance. Similar changes may be attained by using other methods (e.g., a wider range of task settings in training instead of a narrow focus on a single task setting), some of which may also serve to illustrate the separation of the two types of processes. There are also methods that involve process dissociation procedures (Destrebecqz & Cleeremans, 2001).

Posner, DiGirolamo, and Fernandez-Duque (1997) reviewed evidence from brain imaging research that indicated the possibility of different brain regions being associated with implicit and explicit learning and memory (see also Aizenstein et al., 2000; Poldrack et al., 2001). Such evidence relates biological findings to psychological constructs and lends additional support to the hypothesized separation of implicit and explicit processes. Keele, Ivry, Mayr, Hazeltine, and Heuer (2003) reviewed evidence from brain imaging studies that further delineated the nature of two separate learning processes. Mishkin et al. (1984) studied the contrast between (explicit) one-shot learning and (implicit) repeated habit formation training using brain lesioned primates, which showed physiological separation of these two types of processes. The studies involving brain damaged amnesic patients (such as by Nissen & Bullemer, 1987) indicated also that such patients were able to learn as well as normal subjects in task settings where implicit learning was dominant but not in settings where explicit learning was required, which also lent support for the physiological separation of the two types of processes. (However, the results of Boyd & Weistein, 2001, and Muente, Panning, Piepenbrock, & Muente, 2001, seemed different.)

This above interpretation of human data accords well with the CLARION model. With the model, one can select one type of learning or the other by engaging or disengaging the top level (and its learning mechanisms) or the bottom level (and its learning mechanisms). The question now is how a subject "decides" this on the fly and how the model accomplishes this "decision."

Division of Labor

A general pattern discernible from human data (especially those from the implicit learning literature) is that if to-be-learned relations are simple and the number of input dimensions is small (in other words, if the relations are salient to subjects), explicit learning usually prevails; if more complex relations and a larger number of input dimensions are involved (Halford, Wilson, & Philips, 1998), implicit learning becomes more prominent (Kemler-Nelson, 1984; Mathews et al., 1989; Reber, 1989; Seger, 1994; Sun et al., 2001). This pattern has been demonstrated in artificial grammar learning tasks, SRT tasks, and DC tasks, as reviewed earlier. Seger (1994) further pointed out that implicit learning was biased toward structures with a system of statistical relations (as indicated by the results of Kersten & Billman, 1992, and Stadler, 1992). The upshot of this is that the implicit learning mechanism appears more structurally sophisticated and able to handle more

---

[15] In the salient version, the computer responded in accordance with the subjects' immediately preceding response. In the nonsalient version, the computer responded in accordance with the response prior to the immediately preceding response.

complex situations (Kemler-Nelson, 1984; Lewicki, Hill, & Czyzewska, 1992).

This observation can be predicted by CLARION. Although explicitness of knowledge at the top level allows a variety of explicit processing (as described in the A Model section), it does not lend itself easily to the learning of complex structures because of crisp representation and selective hypothesis-testing learning process (see Hayes & Broadbent, 1988, regarding selectivity). However, in the bottom level, the distributed representation in the backpropagation network (that incorporates gradedness and temporal information through Q-learning) handles complex relations (including complex sequential relations) and high dimensionality better. A specific instance of this complexity difference effect is as follows. In correspondence with the psychological findings that implicit learning of sequences is biased toward sequences with a high degree of statistical structure (Stadler, 1992), as has been shown by Elman (1990) and Cleeremans and McClelland (1991), backpropagation networks (either feedforward or recurrent) perform well in capturing complex sequences (e.g., in SRT tasks and DC tasks). However, also in correspondence with available human data, the rule learning mechanism at the top level of CLARION has trouble handling complex stochastic sequences, because of the limitations of hypothesis testing and crisp representation (as we demonstrate below in the Simulation of Human Skill Learning Data section). Therefore, in such circumstances, although both levels are present, the bottom level prevails. This correspondence supports the two-level framework of CLARION. (Other instances of this difference abound, such as in terms of number of sequences, number of input dimensions, or distance of dependencies.)

This explanation implies that it is not necessary to deliberately and a priori decide when to use implicit or explicit learning. When complex relations and high dimensionalities are involved and the top level fails to learn (or is slow to learn), then we can expect a reliance on implicit learning at the bottom level. When the stimulus materials involved are simple, the top level may handle them better and therefore be more readily utilized. This accords well with the fact that in most situations, both types of learning are involved, with varying amounts of contributions from each (Seger, 1994; Sun et al., 1998, 2001).

There have also been other views concerning division of labor, for example, in terms of procedural versus declarative processes (Anderson, 1983, 1993; Anderson & Lebiere, 1998), in terms of nonselective versus selective processing (Hayes & Broadbent, 1988), in terms of algorithms versus instances (Stanley et al., 1989), or in terms of unidimensional versus multidimensional systems (Keele et al., 2003). However, these alternative views are not as generically applicable as the implicit–explicit distinction, although each of them may be supported in specific contexts (see the General Discussion).

Bottom-Up Learning

As reviewed earlier, subjects' ability to verbalize is often independent of their performance on implicit learning (Berry & Broadbent, 1988). Furthermore, performance typically improves earlier than explicit knowledge that can be verbalized by subjects (Stanley et al., 1989). For instance, in DC tasks, although the performance of subjects quickly rose to a high level, their verbal knowledge improved far slower: Subjects could not provide usable verbal knowledge until near the end of their training (e.g., as shown by

Stanley et al., 1989). This phenomenon has also been demonstrated by Reber and Lewis (1977) in artificial grammar learning. A more recent study of bottom-up learning was performed by Sun et al. (2001), who used a more complex minefield navigation task. In all of these tasks, it appears easier to acquire implicit skills than explicit knowledge (hence the delay in the development of explicit knowledge). In addition, the delay indicates that explicit learning may be triggered by implicit learning, and the process may be describe as delayed *explication* of implicit knowledge. Explicit knowledge is in a way "extracted" from implicit skills. Explicit learning can thus piggyback on implicit learning. (However, in some tasks, notably in artificial grammar learning, explicit and implicit knowledge appear to be more closely associated; see, e.g., Johnstone & Shanks, 2001.)

In the context of discovery tasks, Bowers et al. (1990) also showed evidence of explication of implicit knowledge. When subjects were given patterns to complete, they showed implicit recognition of what a proper completion might be, even though they did not have explicit recognition of a correct completion. The implicit recognition improved over time and eventually an explicit recognition was achieved. Siegler and Stern (1998) also showed in an arithmetic problem that children's strategy shifts often occurred several trials earlier than their explicit recognition of their strategy changes. Stanley et al. (1989) and Seger (1994) suggested that because explicit knowledge lagged behind but improved along with implicit knowledge, explicit knowledge could be viewed as obtained from implicit knowledge. Cleeremans and McClelland (1991) also pointed out this possibility in analyzing their data.

Several developmental theorists have considered a similar delayed explication process in child development. Karmiloff-Smith (1986) suggested that developmental changes involved "representational redescription": In children, first low-level implicit representations of stimuli were formed; then, when more knowledge was accumulated and stable behavior patterns developed, through a redescription process, more abstract representations were formed that transformed low-level representations and made them more explicit. This redescription process repeated itself a number of times, and a verbal form of representation emerged. Karmiloff-Smith (1986) proposed four representational forms: implicit, primary explicitation, secondary explicitation, and tertiary explicitation. A three-phase process was hypothesized, in which the first phase used only implicit representations and focused on input–output mappings, the second phase focused instead on gaining control over internal representations and resulted in gains in accessibility of representations, and the third phase was characterized by a balance between external information and internal knowledge (thus, a U curve might result). Mandler (1992) proposed a different kind of redescription: From perceptual stimuli, relatively abstract "image-schemas" were extracted that coded several basic types of movements. Then, on top of such image-schemas, concepts were formed using information therein. On the basis of data on perceptual analysis and categorization in infants, she suggested that an infant gradually formed "theories" of how his or her sensorimotor procedures worked and thereby gradually made such processes explicit and accessible. Although the mechanism of explicit representations had always been there, it was only with increasingly detailed perceptual analysis that such representations became detailed enough to allow conscious access. In a similar vein, Keil (1989) viewed conceptual representations as composed of an as-

sociative component (with frequency and correlational information; Hasher & Zacks, 1979) and a theory component (with explicit knowledge; Murphy & Medin, 1985). Developmentally, there was a shift from associative to theory-based representations. In the data concerning learning concepts of both natural and nominal kinds under a variety of conditions, simple similarity-based (or prototype) representations seemed to dominate at first, but gradually explicit and focused theories developed and became prominent. Keil (1989) pointed out that it was unlikely that theories developed independently, but rather they developed somehow from associative information already available. These theories and data testify to the ubiquity of the implicit-to-explicit transition.

CLARION captures this kind of bottom-up process. In the model, the bottom level develops implicit skills, on its own, using QBP, whereas the top level extracts explicit rules using RER (see the Some Details of the Model section). Thus, delayed bottom-up learning naturally falls out of the model.

As reviewed earlier, there is also evidence that explicit knowledge may develop independently (with little or no correlation with implicit skills), under some circumstances. Willingham et al. (1989) reported data that were consistent with the parallel development of implicit skills and explicit knowledge. Correspondingly, in CLARION, explicit hypothesis testing can be used for learning rules in the top level, independent of the bottom level (as explained in the A Model section; see also the Simulation of Human Skill Learning Data section), when to-be-learned materials are not too complex.[16]

Differences in Representation of Resulting Knowledge

Although there have been suggestions that implicit and explicit learning and performance may be based on the same knowledge base, most data support the position that the two have separate, though contentwise overlapping, knowledge bases (see Aizenstein et al., 2000; Faulkner & Foster, 2002; Poldrack et al., 2001; Sun, 1994, 1997). Such a separation was hypothesized in our model (see the A Model section).

Reber (1989) argued that implicit learning functioned by the induction of an underlying tacit representation that mirrored the structure intrinsic to the interaction of subjects with the world. Seger (1994) argued that in implicit learning, subjects developed abstract but instantiated representations. Being instantiated meant that representations were often tied to sensory–motor modalities of stimuli; being abstract meant that knowledge represented could be transferred or generalized to novel stimuli. However, the representation of explicit knowledge is almost universally accepted to be symbolic (Dulaney et al., 1984; Smolensky, 1988; Stanley et al., 1989; Sun, 1994, 1999; Sun et al., 2001).

CLARION predicts these differing characteristics of the two types of representations. At the bottom level, an instantiated representation is used for the input and the output layer of a back-propagation network (G. Bower, 1996). This type of network is also abstract in Seger's (1994) sense because distributed representations in the network provide a generalization ability (Rumelhart et al., 1986; Sun et al., 1998, 2001). The network is also tacit because of the lack of direct interpretability of distributed representations involved. However, at the top level of CLARION, rules

that are more abstract, less tacit, and less tied to specific sensory modalities are learned (because of generalization in rule learning; see the Some Details of the Model section) and represented symbolically, which corresponds to subjects' explicit knowledge.

Differences in Accessibility of Resulting Knowledge

There have been disagreements concerning what experimentally constitutes accessibility. It is also difficult to distinguish between explicit knowledge that is consciously used when a task is being performed and explicit knowledge that is retroactively attributed to task performance (e.g., when verbal reports are given; Nisbett & Wilson, 1977). Despite such difficulties, it is generally agreed that at least some part of skilled performance is not conscious under normal circumstances (by whatever experimental measures and whatever operational definitions). Reber (1989) pointed out that "although it is misleading to argue that implicitly acquired knowledge is completely unconscious, it is not misleading to argue that implicitly acquired epistemic contents of mind are always richer and more sophisticated than what can be explicated" (p. 229).

Consistent with the idea of bottom-up learning, Reber (1989) further pointed out, "Knowledge acquired from implicit learning procedures is knowledge that, in some raw fashion, is always ahead of the capability of its possessor to explicate it" (p. 229). For example, Mathews et al. (1989) asked their subjects in a dynamic control task to periodically explicate the rules that they used, and the information was then given to yoked subjects who were then tested on the same task (see also Roussel, 1999; Stanley et al., 1989). Over time, the original subjects improved their performance as well as their explicit knowledge (as evidenced by the improvement of performance of the yoked subjects). However, yoked subjects never caught up with original subjects, thus suggesting both the hypothesis of delayed explication (as discussed before) and that of relative inexplicability of implicit skills.

Such results are explained mechanistically by CLARION, in which delayed explication is the result of bottom-up learning and relative inexplicability of implicit skills is the result of distributed representations in backpropagation networks, which are always richer and more complex than crisp explicit representations used in the top level.

Differences in Flexibility, Generalizability, and Robustness

It has been argued that implicit learning often results in knowledge that is less flexible in some respects. As mentioned above, implicit learning often results in knowledge that is more tied to specifics of learning environments (Dienes & Berry, 1997; Seger, 1994), less reflective (e.g., lacking metaknowledge; Chan, 1992; Dienes & Perner, 1999), and often less adaptable to changing

---

[16] Such hypothesis testing, as mentioned before, is similar to models proposed by Bruner et al. (1956), Haygood and Bourne (1965), and more recently Nosofsky et al. (1994). In artificial intelligence research, many symbolic models for learning rules using hypothesis testing were proposed, such as in Michalski (1983), which may also be adapted for the top level of CLARION.

situations (Hayes & Broadbent, 1988). (There are different views on this issue; see, e.g., Willingham et al., 1989, and Willingham, 1998.) On the basis of psycholinguistic data, Karmiloff-Smith (1986) observed that with the growth of explicit representations, more flexibility was shown by subject children.

Consistent with the above view, CLARION predicts that explicit knowledge often entails a higher degree of flexibility in certain respects: With symbolic–localist representations at the top level of CLARION, a variety of explicit manipulations can be performed that are not available to the bottom level. For example, backward and forward chaining reasoning, counterfactual reasoning, explicit hypothesis testing, and so on can be used individually or in combination. These capacities lead to heightened flexibility of the top level in these respects. Thus, CLARION explains the claimed difference in flexibility between the two types of processes in those respects.

As observed in many experiments, following implicit learning, subjects were often able to handle novel stimuli or, in other words, to generalize. In artificial grammar learning, Reber (1967, 1976) found good transfer to strings involving different letters but based on the same grammar. Berry and Broadbent (1988) showed that subjects trained on one dynamic control task could transfer to another that had a similar cover story and an identical underlying relation.[17] As shown by Vokey and Brooks (1992) and others, both similarity between a novel stimulus and learned stimuli and more abstract commonalities between them (such as grammars and common category assignments) were used in generalization. (There were of course also demonstrations of failure of generalization after implicit learning.)

The bottom level of CLARION, which contains backpropagation networks, has the ability to capture generalization exhibited in skill learning. Generalization has been amply demonstrated in backpropagation networks in various contexts: Elman (1990) reported good generalization by recurrent backpropagation networks in grammar learning; Pollack (1991) found generalization of such networks to arbitrarily long sequences; Cleeremans and McClelland (1991) and Dienes (1992) modeled data of artificial grammar learning with such networks. As in human learning, generalization in networks is based in part on similarity of old and new sequences but also in part on certain structures exhibited by the sequences. Explicit processes in the top level of CLARION can also generalize, albeit following a different style via explicit hypothesis testing. This alternative style of generalization has been investigated to some extent by hypothesis-testing psychology, for example, in studies by Bruner et al. (1956), Haygood and Bourne (1965), and Dominowski (1972).

It has also been argued that implicit processes are often more robust than explicit processes (e.g., Reber, 1989) in the face of internal disorder and malfunctioning. For example, Hasher and Zacks (1979) found that encoding of frequency information (an implicit process) was correctly performed by clinically depressed patients, even though they could not perform explicit tasks. Warrington and Weiskrantz (1982) found that amnesic patients were more successful in performing implicit rather than explicit memory tasks. Implicit processes are also often more robust in the face of dual-task distractions, as shown by Curran and Keele (1993), although they can be affected as well. (In general, this robustness is limited; there have been demonstrations of impaired implicit processes in the literature.)

This view of differing degrees of robustness can be predicted within the dual-representation framework of CLARION: Whereas the top level uses crisp symbolic–localist representations and is thus more vulnerable to malfunctioning, the bottom level uses distributed representations that are more resilient, as demonstrated in connectionist modeling (e.g., Plaut & Shallice, 1994; Rumelhart et al., 1986). However, this difference may not be absolute, and the bottom level may be impaired under many circumstances as well (Plaut & Shallice, 1994).

Knowledge Interaction

Interactions of various sorts between implicit and explicit processes exist. With regard to the use of explicit knowledge to affect implicit processes (top-down information), the existing literature suggests that explicit knowledge may help subjects to learn when it directs subjects to focus on relevant features or when it heightens subjects' sensitivity to relevant information (see, e.g., Reber et al., 1980, and Howard & Ballas, 1980). Explicit knowledge may also help subjects to deal with high-order relations (Premack, 1988). However, as Reber (1976, 1989) pointed out, explicit knowledge may also hamper implicit learning, especially (a) when explicit prior instructions induce an explicit learning mode in a task that is not suitable for explicit learning (e.g., Schooler et al., 1993) or (b) when explicit knowledge conflicts with the implicit learning that subjects are undergoing and the implicit representations used by the subjects (Dulaney et al., 1984; Roussel, 1999). Owen and Sweller's (1985) findings that learning might be hampered when certain explicit processes (such as means–ends analysis) were encouraged also supported this idea. (However, Jimenez & Mendez, 2001, argued that implicit learning was unaffected by explicit knowledge. DeShon & Alexander, 1996, argued that explicit goal setting could not affect implicit learning.)

We know less about how implicit knowledge is used to affect explicit processes (bottom-up information). We posit that implicit knowledge is used in explicit learning and explicit performance (e.g., certain verbalization), and this reverse influence, given proper circumstances, can be strong. As indicated in the *Division of Labor* section, implicit processes often handle more complex relations (Berry & Broadbent, 1988; Lewicki et al., 1992) and can thus help explicit processes by providing them with relevant information. Implicit processes are also better at keeping track of statistical information and may use them to aid explicit processes in useful ways (Gluck & Bower, 1988; Hasher & Zacks, 1979; Lewicki et al., 1987). Bottom-up information becomes pronounced in certain brain damaged patients who are capable of certain tasks only implicitly (e.g., blind sight or amnesic patients; Schacter, 1990).

However, during explicit tasks (e.g., verbal reasoning), subjects might ignore information from implicit processes when it contradicts their explicit knowledge or explicit mental models (Seger, 1994). For example, Murphy and Medin (1985) demonstrated that concept formation was not merely a feature similarity based (implicit) process. Prior theory played an important part and could overwrite feature similarity based decisions. Rips (1989) demon-

---

[17] However, Berry and Broadbent (1988) also showed that if the cover story was completely different, there was no transfer exhibited by subjects.

strated a similar point. Wisniewski and Medin (1994) further delineated the possible ways in which such interactions might happen. Stanley et al. (1989) interpreted the difficulty their subjects had with verbalizing their knowledge used in task performance as the interference of subjects' explicit, prior domain knowledge (mental models) on the explication of their implicit knowledge.[18]

As explained in the A Model section, the interaction of explicit and implicit processes is embodied in CLARION with the two-level dual-representation framework and the interlevel interaction mechanism, which allows a proper mixture of the top and the bottom level that captures the "superimposition" of the effects of the two levels. Conflicts between the two types of knowledge occur either when the development of explicit knowledge lags behind that of implicit knowledge (see the *Bottom-Up Learning* section) or when the top level acquires explicit knowledge independent of the bottom level (when independent hypothesis testing learning methods are used; see *The Top Level* section). When conflicts occur, the interlevel interaction mechanism of CLARION may ignore the bottom level (in explicit reasoning) or the top level (in implicit skilled performance), in ways consistent with the above interpretation of human data.

Synergy

Why are there two separate (although interacting) levels? There need to be reasons other than mere redundancy (e.g., for the sake of fault tolerance). The discussion in the *Division of Labor* section concludes that there is a division of labor between explicit and implicit processes. We further hypothesize that there may be a synergy between the two types of processes (partially on the basis of dissociation between the two types of processes as discussed in the *Dissociation* section). Such a synergy may show up, under right circumstances, by speeding up learning, improving learned performance, and facilitating transfer of learned skills.

As mentioned earlier, there is indeed some evidence in support of this hypothesis. In terms of speeding up learning, Willingham et al. (1989) found that those subjects who acquired more explicit knowledge in SRT tasks appeared to learn faster. Stanley et al. (1989) reported that in a DC task, subjects' learning improved if they were asked to generate verbal instructions for other subjects during learning. That is, a subject is able to speed up his or her own learning through an explication process that generates explicit knowledge. Sun et al. (2001) showed a similar effect of verbalization in a minefield navigation task, and Reber and Allen (1978) showed a similar effect in artificial grammar learning. Mathews et al. (1989) showed that a better performance could be attained if a proper mix of implicit and explicit learning was used (in their case, through devising an experimental condition in which first implicit learning and later explicit learning was encouraged).[19]

In addition, in terms of learned performance, Stanley et al. (1989) found that subjects who verbalized while performing SRT tasks were able to attain a higher level of performance than those who did not verbalize, because the requirement that they verbalized their knowledge prompted the formation and utilization of explicit knowledge. Squire and Frambach (1990) reported that initially, amnesic and normal subjects performed comparably in a DC task and equally lacked explicit knowledge. However, with

more training, normal subjects achieved better performance than amnesic subjects and also better scores on explicit knowledge measures, which pointed to the conjecture that it was because normal subjects were able to learn better explicit knowledge that they achieved better performance. Consistent with this view, Estes (1986) suggested that implicit learning alone could not lead to optimal levels of performance. Even in high-level skill acquisition, similar effects were observed. Gick and Holyoak (1980) found that good problem solvers could better state rules that described their actions in problem solving. A. Bower and King (1967) showed that verbalization improved performance in classification rule learning. Gagne and Smith (1962) showed the same effect of verbalization in learning the Tower of Hanoi task. This phenomenon may be related, to some extent, to the self-explanation effect reported in the cognitive skill acquisition literature (Chi, Bassok, Lewis, Reimann, & Glaser, 1989): Subjects who explained examples in physics textbooks more completely did better in solving new problems. In all these cases, it could be the explication process and the use of explicit knowledge that helped the performance.

In terms of facilitating transfer of skills, Willingham et al. (1989) showed some suggestive evidence that explicit knowledge facilitates transfer of skilled performance. They reported that (a) subjects who acquired explicit knowledge in a training task tended to have faster response times in a transfer task; (b) these subjects were also more likely to acquire explicit knowledge in the transfer task; and (c) these subjects who acquired explicit knowledge responded more slowly when the transfer task was unrelated to the training task, suggesting that the explicit knowledge of the previous task might have interfered with the performance of the transfer task. Sun et al. (2001) showed similar effects. In high-level domains, Ahlum-Heath and DiVesta (1986) found that the subjects who were required to verbalize while solving Tower of Hanoi problems performed better on a transfer task after training than did the subjects who were not required to verbalize. Berry (1983) similarly showed in Watson's selection task that verbalization during learning improved transfer performance. Nokes and Ohlsson (2001) showed related results as well.

Note that synergy effects are dependent on experimental settings. They are not universal. For example, Roussel (1999) showed that under some circumstances, explicit reflection did not help to improve performance and might in fact hurt performance. Even so, it should be recognized that explicit processes play important cognitive functions. Cleeremans and McClelland (1991) and Gibson, Fichman, and Plaut (1997) both pointed to the need to include explicit processes in modeling typically implicit learning tasks. Explicit processes also serve additional functions, such as facilitating verbal communication and acting as gatekeepers (e.g., enabling conscious veto; Libet, 1985).

It has been demonstrated (Sun et al., 1998, 2001) that CLARION can produce similar synergy effects as described above

---

[18] As discussed by Nisbett and Wilson (1977), such mental models are composed of cultural rules, causal schemata of a particular culture or an individual, causal hypotheses generated on the fly, and so on.

[19] This body of evidence, considered in its entirety, cannot be explained away by factors such as attention and effort (e.g., induced by verbalization).

(as well as other effects in different settings), through comparing various training conditions, such as comparing the verbalization condition and the nonverbalization condition (whereby the verbalization condition encourages explicit learning) or comparing the dual-task condition and the single-task condition (whereby the dual-task condition discourages explicit learning; see justifications later). Although the idea of synergy has been mentioned before (e.g., Mathews et al., 1989), CLARION demonstrated in computational terms the very process by which synergy was created (Sun et al., 2001; Sun & Peterson, 1998a), which had a lot to do with differing characteristics of the two levels (see the *Differences in Representation of Resulting Knowledge; Differences in Accessibility of Resulting Knowledge;* and *Differences in Flexibility, Generalizability, and Robustness* sections). For the details of the synergy effects, see the next section (and also previous work by Sun et al., 1998, 2001; Sun & Peterson, 1998a, 1998b).

In summary, we have discussed the following issues: (a) the separation of two types of processes (in the *Dissociation* and *Division of Labor* sections) and how CLARION captures it, (b) the conjectured differing characteristics of the two types of processes (in the *Differences in Representation of Resulting Knowledge, Differences in Accessibility of Resulting Knowledge;* and *Differences in Flexibility, Generalizability, and Robustness* sections) and how they might correspond with CLARION, (c) the possible ways of interaction between the two types of processes (in the *Bottom-Up Learning* and *Knowledge Interaction* sections) and how CLARION accommodates them, and finally, (d) the possible consequences of the interaction (in the *Synergy* section) and how CLARION also generates them.

Note that in the discussions above, in most cases, our view is but one possible interpretation, and our model corresponds to that particular view. We would caution against oversimplifications of these broad issues and against any mistaken impression that these issues have been settled and our model accounts fully for them.

Simulation of Human Skill Learning Data

To substantiate the afore-discussed points, we conducted simulations using CLARION. Two kinds of tasks were chosen to be simulated: SRT tasks and DC tasks. These tasks were chosen because (a) they were representative of the tasks used in implicit learning research and (b) they showed most clearly the interaction between implicit and explicit processes.[20] We have also simulated a variety of other tasks, reported elsewhere, which are summarized briefly later (see Slusarz & Sun, 2001; Sun et al., 2001; Sun & Zhang, 2004).

In the simulations of these tasks, the same QBP and RER algorithms were used, at the bottom and the top level, respectively (although some details varied from task to task as appropriate). Although alternative encodings and algorithms (such as recurrent backpropagation in the bottom level) were also possible, we felt that our choices were reasonable ones, consistent with existing theories and data.

We focus on capturing the interaction of the two levels in the human data, whereby the respective contributions of the two levels are discernible through experimental manipulations of learning settings that place differential emphases on the two levels. We show how these data can be captured using an interactive and

bottom-up perspective. To capture manipulations, we do the following. (a) The explicit (how-to) instructions condition is modeled using explicit encoding of the given knowledge at the top level (prior to training). (b) The verbalization condition (in which subjects are asked to explain their thinking while or between performing the task) is captured in simulation through changes in parameter values that encourage more top-level activities, consistent with the existing understanding of the effect of verbalization (i.e., subjects become more explicit; Stanley et al., 1989; Sun et al., 1998). (c) The explicit search condition (in which subjects are told to perform an explicit search for regularities in stimuli) is captured through changes in parameter values that encourage more reliance on increased top-level rule learning activities, in correspondence with what we normally observe in subjects under the kind of instruction. (d) The dual-task condition is captured by changes in parameter values that reduce top-level activities, because we have reasons to believe that when distracted by a secondary task, top-level mechanisms will be less active in relation to the primary task (i.e., this condition affects explicit processes more than implicit processes; Cleeremans, 1993; Dienes & Berry, 1997; Hayes & Broadbent, 1988; Sun et al., 2001).[21] (e) Finally, given same conditions, subjects may differ in terms of the explicitness of their resulting knowledge after training (some are more aware of their knowledge, whereas others are less aware; Curran & Keele, 1993; Willingham et al., 1989). Individual differences in degree of awareness are captured by different parameter settings: More aware subjects have parameter settings that allow more rule learning to occur. (f) Many of these afore-enumerated manipulations lead to the synergy effects between implicit and explicit processes. By modeling these manipulations, we capture the synergy effects as well. Note that for modeling each of these manipulations, usually only one (or two) parameter values are changed (more details later).

Many parameters in the model were set uniformly as follows (see Table 1): Network weights were randomly initialized between -0.01 and 0.01; the Q-value discount rate was set at 0.95; the temperature (randomness parameter) for stochastic decision making was set at 0.01; the combination weights of the two levels were set at $w_1 = 0.2$ and $w_2 = 0.8$. The density parameter was set at 1/50. In this work, these are not free parameters because they were set in an a priori manner (on the basis our experience in our previous work) and not varied to match the human data. (These parameter values are not described again when discussing individual simulations later, except when deviations from the above settings are encountered.)

---

[20] In addition, these tasks are different from each other, so that different aspects of the model may be tested.

[21] Note that there have been alternative interpretations of dual-task conditions, for example, in terms of mixed modality sequences or altered organizations (Keele et al., 2003; Stadler, 1995). There have also been arguments in favor of the view that dual-task conditions dampen implicit learning or expression of implicit knowledge as well (Frensch et al., 1999; Nissen & Bullemer, 1987). However, despite such differences, it seems reasonable to assume that dual-task conditions affect negatively explicit learning more than implicit learning, judging from the literature (Hayes & Broadbent, 1988; Curran & Keele, 1993; Cleeremans, 1993; Stadler, 1995; Szymanski & MacLeod, 1996; Dienes & Berry, 1997; Cleeremans et al., 1998; Sun et al., 2001); hence, our conjectured change of learning parameters at the top level is a reasonable approach of first approximation.

Table 1
Non-Domain-Specific Parameters Used in CLARION

Parameter	Value
Initial weights	Between -0.01 and 0.01
Q-value discount	0.95
Temperature	0.01
Density	1/50
Weights of two levels	0.2–0.8

Other parameters are domain specific because they are likely adjusted in relation to domain characteristics. For example, the numbers of input, output, and hidden units were set in a domain-specific way. The same goes for the rule learning thresholds, the backpropagation learning rate, and the momentum.[22] (These parameters are described later with respect to individual tasks.) Most of these parameters were not varied to match different conditions of the same task.

A note about parameters and model fitting is in order here. At first glance, the model has too many parameters. However, on closer examination, the number of parameters is not significantly higher than it is in usual models such as backpropagation networks. In addition to parameters of backpropagation networks (as in the bottom level of the model), at the top level, there are only three significant parameters concerning rule extraction and revision. That is to say, our model is roughly comparable to backpropagation networks. Moreover, although values of all parameters affect performance, most of them were not changed throughout the simulations. Thus they are not free parameters. Even most of the domain-specific parameters were not varied to capture different experimental conditions in a task. Thus, they should not be viewed as free parameters either. They do not contribute to the degree of freedom that we have to match the change of performance of human subjects across different conditions (see McClelland, McNaughton, & O'Reilly, 1995, for a similar point). The change of performance across different conditions in a task is accounted for by changes of one or two parameters. Finally, it should be noted that our goal is to show the effect of the interaction of the two types of learning (in terms of general trends). Thus, capturing finer details of learning curves and other characteristics is not our goal here.[23] Many other researchers have taken a similar approach, for example, McClelland et al. (1995).

Simulating SRT Tasks

In this type of task, we aim to capture (a) the degree of awareness effect, (b) the salience difference effect, (c) the explicit (how-to) instructions effect, and (d) the dual-task effect. We use the data reported in Lewicki et al. (1987) and Curran and Keele (1993).

For this type of task, we used a simplified QBP in which temporal credit assignment was not used. This was because subjects predicted one position at a time, with immediate feedback, and thus there was no need for backward temporal credit assignment. $Q(x, a)$ computes the likelihood of the next position a, given the information concerning the current and past positions x. The actual probability of choosing a as the current prediction (of the next position) is determined on the basis of the Boltzmann distri-

bution (described earlier). The error signal used in the simplified QBP is as follows:

$$\Delta Q(x, a) = \alpha(r + \gamma \max_b(y, b) - Q(x, a)) = \alpha(r - Q(x, a)),$$

where x is the input, a is one of the outputs (predictions), $r = 1$ if a is the correct prediction, $r = 0$ if a is not the correct prediction, and $\gamma \max_b(y, b)$ is set to zero.[24]

Simulating Lewicki et al. (1987)

The task. The task was based on matrix scanning: Subjects were to scan a matrix, determine the quadrant of a target digit (the digit 6), and respond by pressing the key corresponding to that quadrant. Each block consisted of six identification trials followed by one matrix scanning trial. In identification trials, the target appeared in one of the quadrants, and the subject was to press the corresponding key. In matrix scanning trials, the target was embedded among 36 digits in a matrix, but the subject's task was the same. See Figure 3. In each block of seven trials, the actual location of the target in the seventh (matrix scanning) trial was determined by the sequence of the six preceding identification trials (out of which four were relevant). Twenty-four rules were used to determine the location of the target in the seventh trial. Each of these rules mapped the target quadrants in the six identification trials to the target location in the seventh trials in each block. Twenty-four (out of a total of 36) locations were possible for the target to appear. The major dependent variable was the reaction time in the seventh trial in each block.

The whole experiment consisted of 48 segments, each of which consisted of 96 blocks (so there were a total of 4,608 blocks). During the first 42 segments, the aforementioned rules were used to determine target locations. However, at the 43rd segment, a switch occurred that reversed the outcomes of the rules: The upper left was replaced by the lower right, the lower left was replaced by upper right, and so on. The purpose was to separate unspecific learning (e.g., motor learning) from prediction learning (i.e., learning to predict the target location in the seventh trial).

The data. The reaction time data of 3 subjects were obtained by Lewicki et al. (1987; see Figure 4). Each curve showed a steady decrease of reaction times up until the switch point. At that point, there was a significant increase of reaction times. After that, the curve gradually lowered again.

The model setup. In this experiment, the simplified QBP was used. The input contained (a sequence of) six elements, with each element having four possible values (for four different quad-

---

[22] The last two parameters are partially determined by other domain-specific parameters and thus also set in a domain-specific way.

[23] Another reason why finer details are not captured here is that we do not have full human data of these tasks. Thus, a detailed statistical analysis of matching between model and human data is impossible, which makes finer matching uninteresting.

[24] This simplified version is in fact similar to straight backpropagation (except that we only update one of the output at each step—the output that is used as the current prediction). Note that we also tried the full-fledged QBP instead of the simplified version described above. We obtained comparable results from this alternative approach.
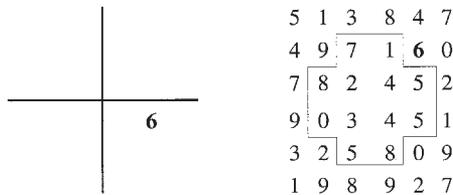
Figure 3. The displays for the identification trials (left) and the matrix scanning trials (right) in the experiments of Lewicki et al. (1987). Adapted from "Unconscious Acquisition of Complex Procedural Knowledge," by P. Lewicki, M. Czyzewska, and H. Hoffman, 1987, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* p. 525. Copyright 1987 by the American Psychological Association.

rants).[25] The output contained the prediction of the seventh element. Thus, 24 input units (representing six elements, with four values each), 24 output units (one for each possible location of the seventh element), and 18 hidden units were used. The best learning rate was 0.5, with a momentum term of 0.2. The model was trained by presenting the stimulus materials in the same way as in the human experiment of Lewicki et al. (1987), without any further embellishments or repetitions of the materials.

Because in this experiment there were a total of 6^4 possible sequences with each consisting of seven elements, the setting was too complex for subjects to discern the sequence structures explicitly, as demonstrated through various explicit tests by Lewicki et al. (1987) (although some argued otherwise; see, e.g., Perruchet & Gallego, 1993). Computationally, no rule could be extracted in the model because the large number of sequences entailed the lack of sufficient repetitions of any sequence, which prevented the model from coming up with any rule. The density parameter was set to be 1/50; that is, at least one repetition (of a sequence) was necessary every 50 blocks to maintain a rule. In this task, there were 4,608 presentations of sequences and there were $6^4 = 1,296$ possible sequences. Thus, on average, the repetition rate of any sequence was significantly below 1/50. Therefore, almost no rule could be maintained. Our simulation thus involved the bottom level of the model (with QBP).[26] See Figure 5 for a summary of parameter values used.

The match. The model was able to produce an error rate curve going downward, resembling Lewicki et al.'s (1987) reaction time curves. See Figure 5 (which was averaged over 10 runs to ensure representativeness). The model reached 100% accuracy before the switch.

The question is how we should translate error rates into reaction times. One way of translation is through a linear transformation from error rate to reaction time (as argued for by, e.g., Mehwort, Braun, & Heathcote, 1992, and as often used in existing work); that is, $RT_i = a \times e_i + b$, where RT_i is the reaction time (in ms), e_i is the error rate, and a and b are two free parameters. One interpretation of linear transformation is that it specifies the time needed by a subject to search for and then respond to a target item and the time needed by a subject to respond to a target item without searching (through correctly predicting its location); that is,

$$RT_i = ae_i + b = b(1 - e_i) + (a + b)e_i,$$

where b is interpreted as the time needed to respond to an item without searching (because $1 - e_i$ is the probability of successfully

predicting the location of a target item) and $a + b$ is interpreted as the time needed to respond to an item by first searching for it and then responding to it. So, instead of relying on an additional power function (as in Ling & Marinov, 1994), this method relies only on error rates to account for human reaction times.

Another way of generating reaction time from prediction accuracy is through adding a power function (as used in previous simulations of this task, such as in Ling & Marinov, 1994):

$$RT_i = t_1(1 - e_i) + t_2e_i + B\alpha^{-t},$$

where t_1 is the time needed to respond when there is no search (using correct predictions), t_2 is the time needed to respond when search is necessary, B is the initial motor response time, and α is the rate at which the motor response time decreases. The third term is meant to capture unspecific practice effects (mostly resulting from motor learning). In other words, in this formula, we separate the motor response time from the search time and the prediction time (as represented by t_1 and t_2, respectively). This formula takes into account the independent nature of motor learning, separate from the learning of target prediction. However, it involves two more free parameters.[27]

Using the linear transformation, we generated three sets of data from the error rate curve earlier,[28] one for matching each human subject in Lewicki et al. (1987), using different a and b values for each subject.[29] As shown in Figure 4, the model outcomes fit the human data well up to the point of switching (Segment 41). When the switch to a random sequence happened, the model's reaction times became much worsened, whereas the subject's reaction times suffered relatively slightly.[30]

Therefore, we added the power function, as done in previous simulations of this task (e.g., Ling & Marinov, 1994). The effect of adding the power function was that we reduced the

---

[25] A sequence of six elements was assumed to be within the capacity of the short-term working memory.

[26] However, we did try simulations with the top level included. We found no significant difference in terms of match with human data.

[27] Note that if we set $B = 0$, we have $t_1 = b$ and $t_2 = a + b,$ and thus this equation becomes the same as the previous one.

[28] When curve fitting, we used Microsoft Excel Solver to find the best parameter values (e.g., a and b in $a \times x + b$) such that the difference between the model data and the subjects' data was minimized. Microsoft Excel Solver uses the generalized reduced gradient nonlinear optimization algorithm (developed by Leon Lasdon, University of Texas at Austin, and Allan Waren, Cleveland State University).

[29] The error rate curve reported earlier was the best curve and happened to match all 3 subjects approximately equally well after the transformation (with different parameters for each subject). Note that Ling and Marinov (1994) also used a single error rate curve to match different subjects with different parameters for transformation.

[30] We tried many different parameters but discovered that the size of the jump tended to vary little (unless the match as a whole was bad). It is clear, from experiments with different parameter settings, that if the model learns the sequences perfectly before the switch (as is the case with our model), the model data inevitably have huge jumps. However, the more of the sequences it does not learn, the flatter the curve and the smaller the jump. Although this may model Subjects 1 and 3 satisfactorily, Subject 2 has a large drop in reaction time early on, which is best matched by having the model increase its accuracy in a rapid manner.

Subject 1

Subject 2

Subject 3
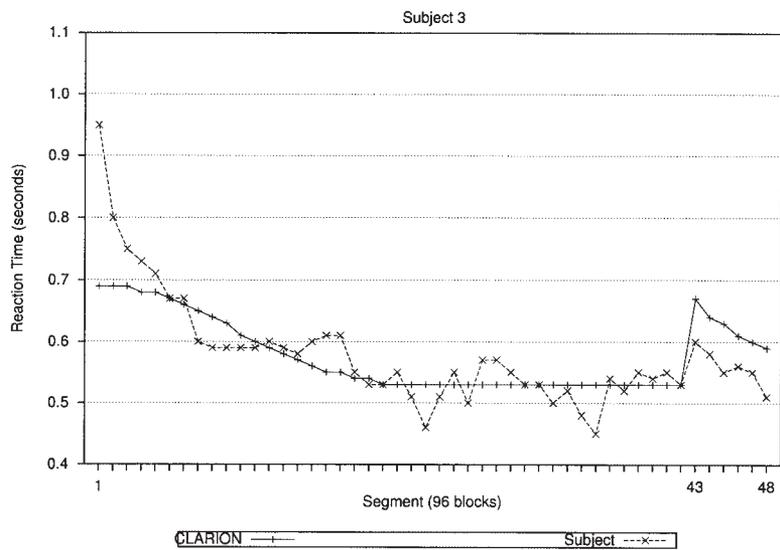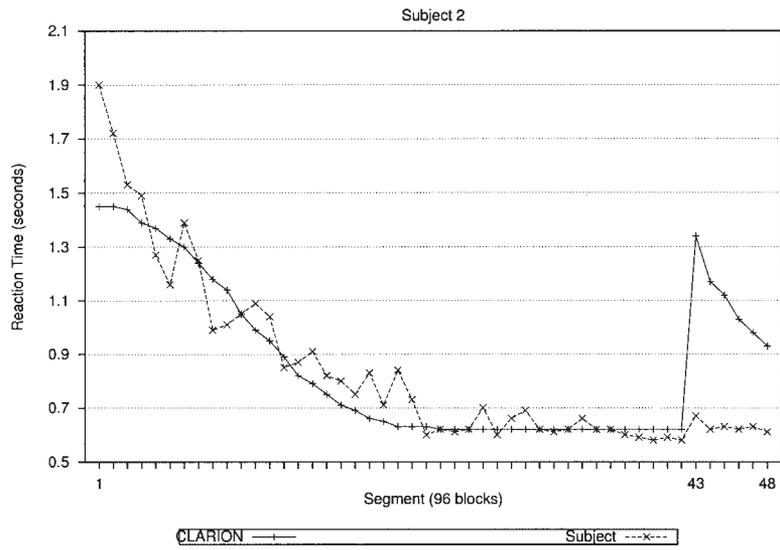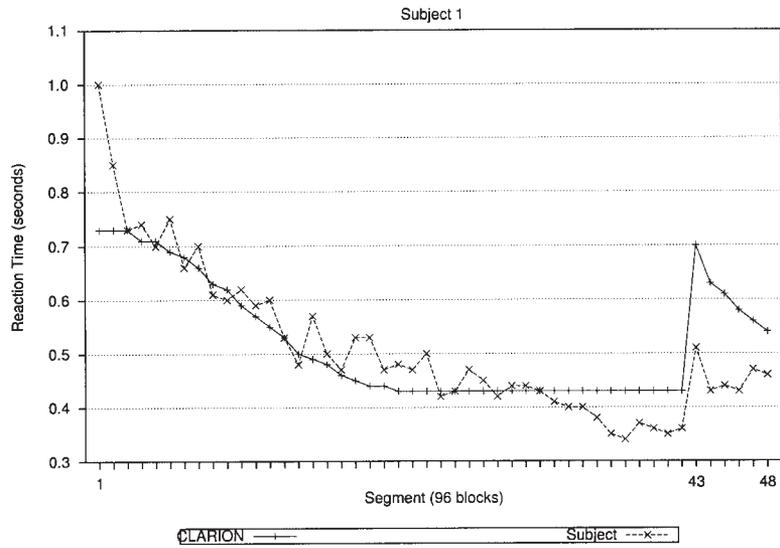
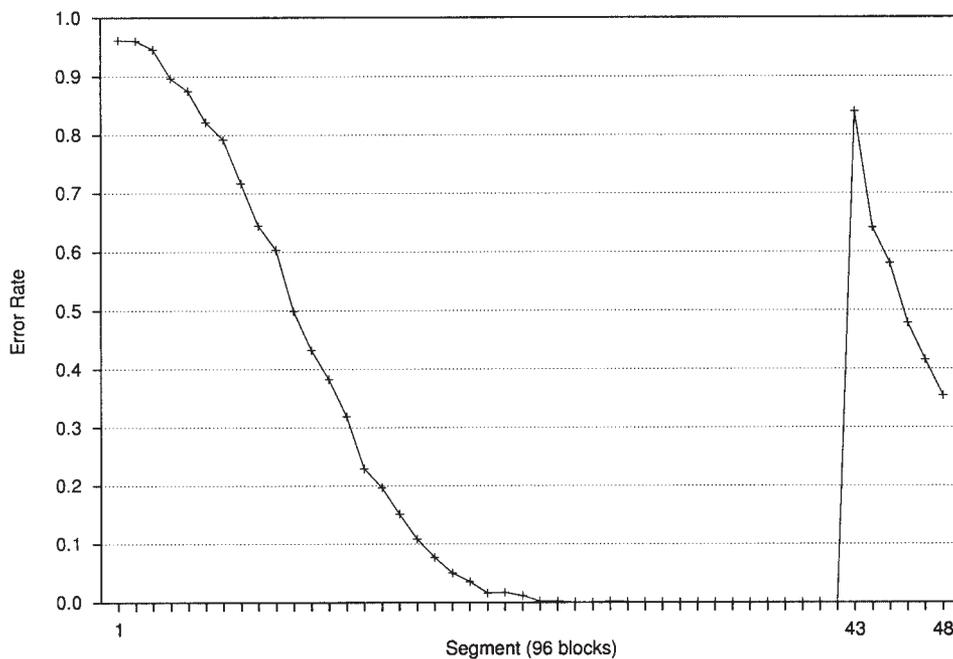| top-level learning:	RER
threshold	n/a
threshold1	n/a
threshold2	n/a
bottom-level learning:	simplified QBP
number of input units:	24
number of output units:	24
number of hidden units:	18
learning rate:	0.5
momentum term:	0.2

Figure 5.   Model parameters used in simulating data from Lewicki et al. (1987; top) and model prediction errors (bottom). RER = rule-extraction-refinement algorithm; QBP = Q-learning-backpropagation algorithm.

contribution from the model prediction (i.e., the error rate e_i) while we took into consideration the contribution from the power function. In this way, we obtained a much better match after the switch while maintaining a good match before the switch. This comparison suggested that the amount of benefit human subjects got from their predictions (i.e., by lowering e_i) was, although significant, relatively small. Significant benefit was also gained through the improvement of motor response as

represented by the power function. This is the same conclusion reached in previous simulations of this task (e.g., Ling & Marinov, 1994). See Figure 6.

Note that the match between our model and the human data was excellent as measured by the sum-squared errors. Compared with Ling and Marinov's (1994) model, CLARION (with the power function) did better for 2 of the 3 subjects, using the same parameter values for transformation as Ling and Marinov did. See Table

Figure 4.   CLARION's match of Lewicki et al.'s (1987) data with linear transformation. See Table 2 for parameter values. The graphs of the subjects' data are adapted from "Unconscious Acquisition of Complex Procedural Knowledge," by P. Lewicki, M. Czyzewska, and H. Hoffman, 1987, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* p. 527. Copyright 1987 by the American Psychological Association.

Subject 1
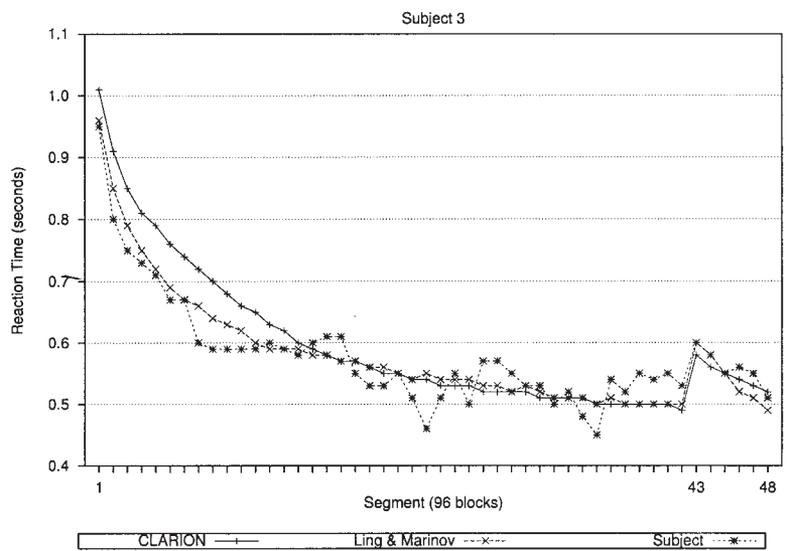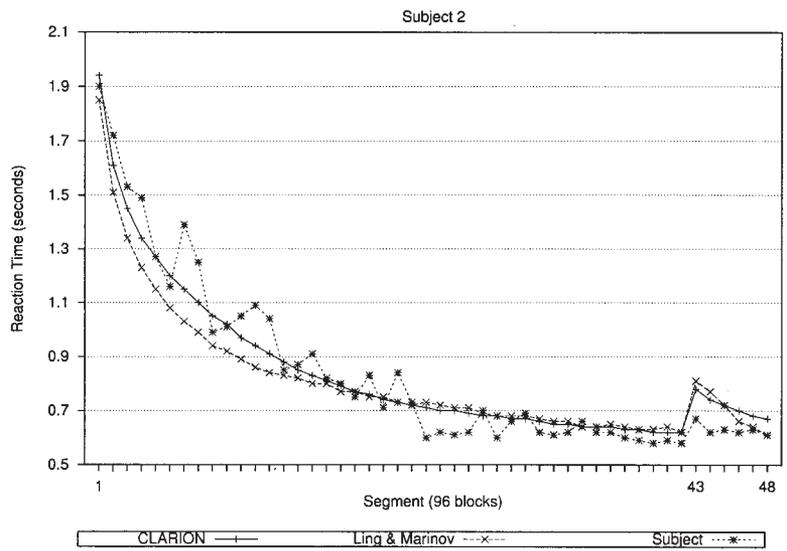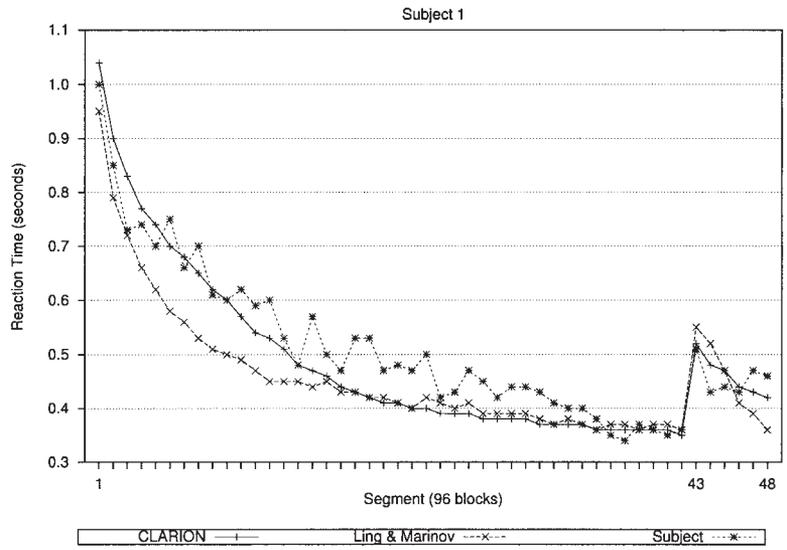
Subject 2

Subject 3

Table 2
Parameters (With and Without Power Functions) for Matching Lewicki et al.'s (1987) Reaction Time Data and Goodness of Fit for the CLARION Model (With and Without Power Functions) and Ling and Marinov's (1994) Model

| | Parameters (without power functions) | | Parameters (with power functions)[a] | | | | Goodness of fit (SSE) | | |
| | | | | | | | CLARION | | Ling & Marinov (1994) |
Subject	a	b	t_1	t_2	B	α	Without power functions	With power functions	
1	320	430	150	350	700	0.33	0.30	0.07	0.14
2	860	620	150	350	1,600	0.33	1.85	0.25	0.43
3	170	530	100	210	800	0.19	0.14	0.05	0.04

Note. SSE = sum-squared errors.
[a] As in Ling and Marinov (1994).

2 for a comparison. Note that in the table, we used exactly the same parameter settings (for *a, b, B,* and α) as did Ling and Marinov. These parameters could be further optimized, which would lead to a slightly better fit but the difference would not be significant.

Discussion. The main feature of the model that the successful match with the human data could be attributed to was the back-propagation learning at the bottom level of the model (and the associated parameters). The top level was insignificant in this task setting.

Simulating Curran and Keele (1993)

The task. The task of Curran and Keele (1993) consisted of presentation of a repeating sequence of X marks, each in one of four possible positions. The subjects were instructed to press the key corresponding to the position of each X mark. Reaction times of the subjects were recorded. The experiment was divided into three phases (in succession): dual-task practice, single-task learning, and dual-task transfer. In the first phase, which consisted of two blocks, the positions of X marks were purely random to allow subjects to get used to the task setting. Each block consisted of 120 trials. The second phase was when learning occurred: There were five sequence blocks (Blocks 3, 4, 5, 6, and 8), in which the positions followed a sequential pattern of length 6 (e.g., 1, 2, 3, 2, 4, 3), and one random block (Block 7). The third phase tested transfer to a dual-task condition (with a secondary tone counting task): Three random blocks (Blocks 9, 10, and 12) and a single sequence block (Block 11) were presented. A total of 57 subjects were tested.

The data. Three groups of subjects were identified in the analysis of data: "less aware," "more aware," and "intentional." Approximately one third of the subjects were found in each group. The intentional subjects were given explicit instructions about the exact sequence used before learning started. The more aware subjects were those who after the experiment correctly specified at least four out of six positions in the sequence used (which demonstrated their explicit knowledge), and the rest were less aware subjects. Analysis of variance (ANOVA; Intentional vs. More Aware vs. Less Aware × Sequential vs. Random) and other analyses were carried out. The analysis showed that although the intentional group performed the best, the more aware group performed close to the intentional group, and both groups performed significantly better than the less aware group. There was a significant interaction between group and block ($p < .01$), demonstrating the effect of explicit knowledge. However, during the transfer to the dual-task condition, all three groups performed poorly. There was notably no significant difference between the three groups (random vs. sequence did not interact significantly with group; $p = .67$). See Figure 7 for the reaction time data for the three groups. The finding of primary interest here is that the difference in explicit knowledge led to the difference in performance in Phase 2 and the performance difference dissipated under the dual-task condition in Phase 3 (which was known to suppress mostly explicit knowledge).

The model setup. The simplified QBP was used at the bottom level (as in the simulation of Lewicki et al., 1987). We used a network with 7 × 6 input units and 5 output units for handling both the SRT primary task and a secondary tone counting task. The seven groups of input units represented a moving window of seven steps preceding the current step (with seven being roughly the size

Figure 6. CLARION's and Ling and Marinov's (1994) matches of Lewicki et al.'s (1987) data with power functions. See Table 2 for parameter values. The graphs of the subjects' data are adapted from "Unconscious Acquisition of Complex Procedural Knowledge," by P. Lewicki, M. Czyzewska, and H. Hoffman, 1987, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* p. 527. Copyright 1987 by the American Psychological Association. The graphs of Ling and Marinov's simulations are adapted from "A Symbolic Model of the Nonconscious Acquisition of Information," by C. X. Ling and M. Marinov, 1994, *Cognitive Science, 18,* p. 600. Copyright 1994 by the Cognitive Science Society. Adapted with permission.
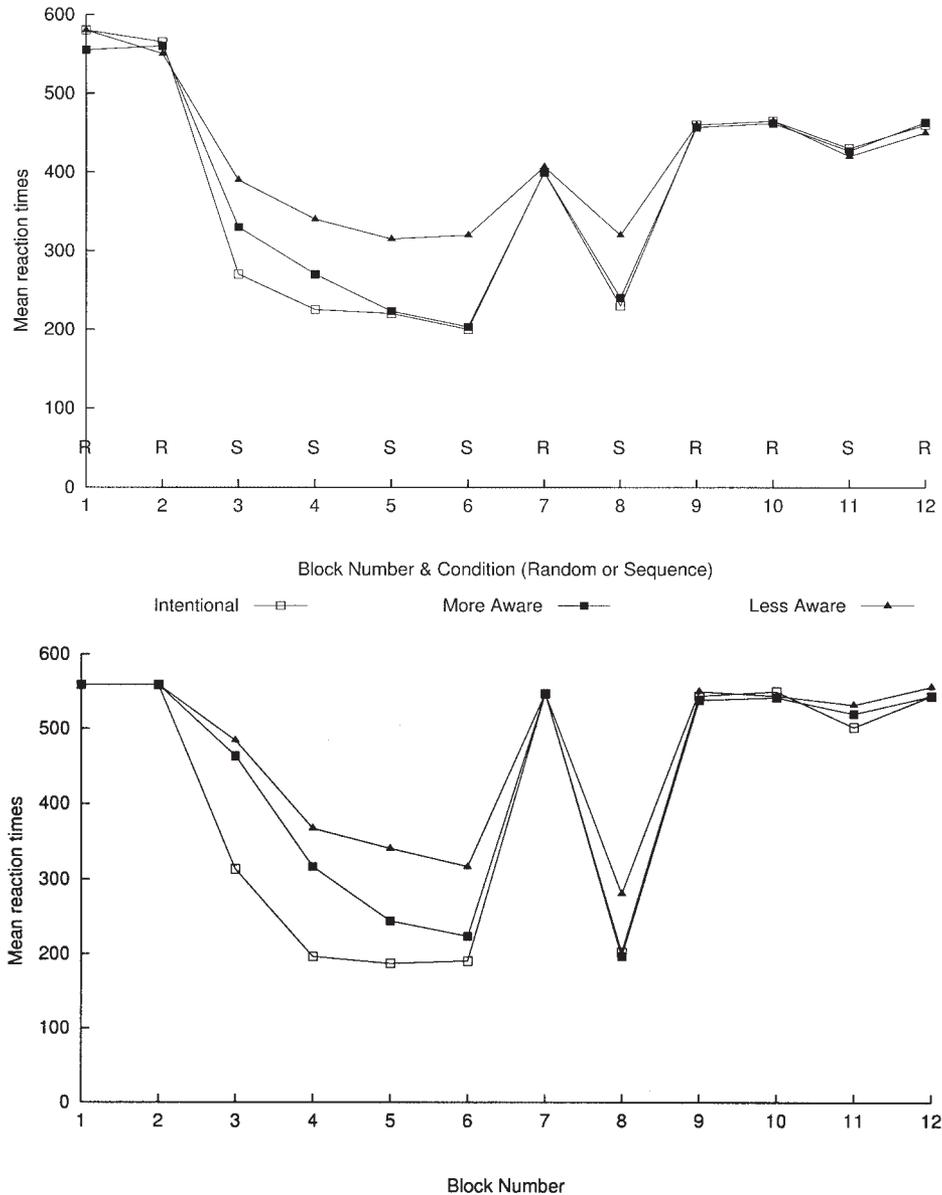
Figure 7. Top: The reaction time data (in milliseconds) from Curran and Keele (1993). Bottom: The reaction time data (in milliseconds) from CLARION's simulation. R = random; S = sequence. The top panel is adapted from "Attentional and Nonattentional Forms of Sequence Learning," by T. Curran and S. W. Keele, 1993, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* p. 192. Copyright 1993 by the American Psychological Association.

of the working memory).[31] In each group, there were 6 input units. The first four were used to encode input for the SRT task, one for each possible position of lights. The other two input units were used for the tone counting task (low tone vs. high tone). The output consisted of four units for predicting the next light position and also a unit for indicating whether a tone was detected. See Table 3 for a summary of parameter values used.

Rule learning was used at the top level. The criterion for rule extraction was based on "reward" received after each button press, which was determined by whether the prediction made was correct: If it was correct, a reward of 1 was given; otherwise, 0 was given. If 1 was received, a rule might be extracted; otherwise, no rule could be extracted (i.e., the threshold was 0.999). The positivity criterion used in calculating IG measures was set to be the same. The rule generalization threshold was set at 3.0. The rule specialization threshold was set at 1.0.

The difference between less aware and more aware subjects was captured by the difference in rule learning thresholds. To simulate

---

[31] We tried increasing the size of the working memory. The results remained essentially the same.

Table 3

Parameters Used in Simulating Data From Curran and Keele (1993)

Parameter	Value
Top-level learning: RER	
Threshold	0.999
Threshold1	3.0/6.0/6.5
Threshold2	1.0/1.5/2.0
Bottom-level learning: Simplified QBP	
No. of input units	42
No. of output units	5
No. of hidden units	20
Learning rate	0.3
Momentum term	0.15

Note. RER = rule-extraction-refinement algorithm; QBP = Q-learning-backpropagation algorithm.

less aware subjects, we used higher thresholds so that they were less likely to develop explicit knowledge: The rule generalization threshold was set at 6.0, and the rule specialization threshold was set at 1.5.[32] To simulate more aware subjects, we used the normal thresholds as given before. To simulate the intentional group, we coded the given sequence as a set of a priori rules, before the training started. The rules directly captured the given sequence.[33]

To simulate the dual-task condition, we interleaved the presentations of tones and lights (as in the Curran & Keele, 1993, experiment) to the backpropagation network (which handled both the secondary tone counting task and the primary task). Consistent with the existing knowledge of the effect of dual tasks (which lied mostly in interfering with explicit processes; see Cleeremans, 1993; Dienes & Berry, 1997; Sun et al., 2001; Szymanski & MacLeod, 1996; but see also Footnote 21 regarding other views), we hypothesized that the dual task interfered mostly with, and thus reduced, top-level activities. Thus, we increased the rule learning thresholds to reduce top-level activities (because rule learning was more difficult than rule application). The rule generalization threshold was increased to 6.5. The rule specialization threshold was increased to 2.0.[34]

We ran 19 model runs for each group, in rough correspondence with the number of subjects in the human experiment.[35] For these model subjects, we randomly set the seeds of random number generators (in initializing weights and in decision making), analogous to random selection of human subjects.

The match. We used a linear transformation that turned error rate into reaction time ($RT_i = a \times e_i + b$, where $a = 600$, $b = 100$). The results, as shown in Figure 7, captured the essential characteristics of the human data. A nonlinear transformation (with a power function as described in the previous subsection) produced an even better fit.

We performed an ANOVA (Intentional vs. More Aware vs. Less Aware × Sequential vs. Random) in correspondence with the analysis of human data in this task. The results showed that there was a significant interaction between group and block ($p < .01$), indicating a significant effect of explicit knowledge, similar to what was found in the human data. The more aware group and the intentional group performed significantly better than the less aware

group (as shown by t tests with $p < .01$), as was in the human data.[36]

For the transfer to the dual-task condition, we performed another ANOVA (Intentional vs. More Aware vs. Less Aware × Sequential vs. Random). The analysis showed that there was no significant difference between the three groups, showing the disappearance of the effect of explicit knowledge under the dual-task condition, as was in the human data.

We compared this simulation with an existing simulation by Cleeremans (1993; see Figure 8). Comparing the two simulations visually, CLARION provided a better approximation of the human data. In fact, the mean square error of the CLARION simulation was indeed lower than that of Cleeremans (1993; 73.1 vs. 79.4). The main difference between the two models lied in the inclusion of bottom-up learning in CLARION (see the *Comparisons with Other Model* section for details of Cleeremans's, 1993, model). Note that there was some mismatch of the human curves by both Cleeremans's simulation and this simulation. This was mainly because unspecific (motor) learning was not taken into account. Cleeremans's simulation did not include a power function for capturing unspecific learning, and therefore, neither did our simulation.

Discussion. The model features that were important in accounting for the data in this simulation include (a) simplified QBP at the bottom level, (b) RER at the top level,[37] and (c) rule learning thresholds (which were varied to simulate different experimental conditions). This simulation suggested that division of labor between the two levels and bottom-up learning were important for capturing human performance in this task.

In this simulation, we captured the following effects related to the interaction of the two levels: (a) the explicit instructions effect, as demonstrated by the intentional group; (b) the degree of awareness effect; (c) the synergy effect, as a result of capturing the above two effects, because explicit processes, in the forms of either given instructions or heightened awareness, led to improved performance; (d) the dual-task effect, as shown by the dual-task transfer

---

[32] Alternatively, we could assume that they had a lower probability of extracting rules (when the threshold for rule extraction was reached). This method worked as well.

[33] For example, if the sequence was $a_1, a_2, \ldots, a_6$, the following rules were used: $a_1 \rightarrow a_2, a_1 a_2 \rightarrow a_3, \ldots, a_1, a_2, a_3, a_4, a_5 \rightarrow a_6$.

[34] We tried some alternatives: reducing the weighting of the top level (in combining the outcomes of the two levels) to reduce the impact of the top level. We also tried adding noise at the top level, which served the same purpose. Both worked as well.

[35] We also tried increasing the number of runs to 100 in an effort to get a smoother curve. However, the new curve appeared essentially the same.

[36] We would like to perform a combined analysis of human and model data, for example, a 2 × 3 × 2 ANOVA (Type [human vs. model] × Group [less aware vs. more aware vs. intentional] × Block [random vs. sequence]). However, because we do not have data of individual human subjects, we cannot perform such an analysis but only a separate analysis of model data in ways similar to the reported analysis of human data in the original article. The same goes for other analyses later.

[37] Note that these two interleaved learning algorithms together created the learning curves.
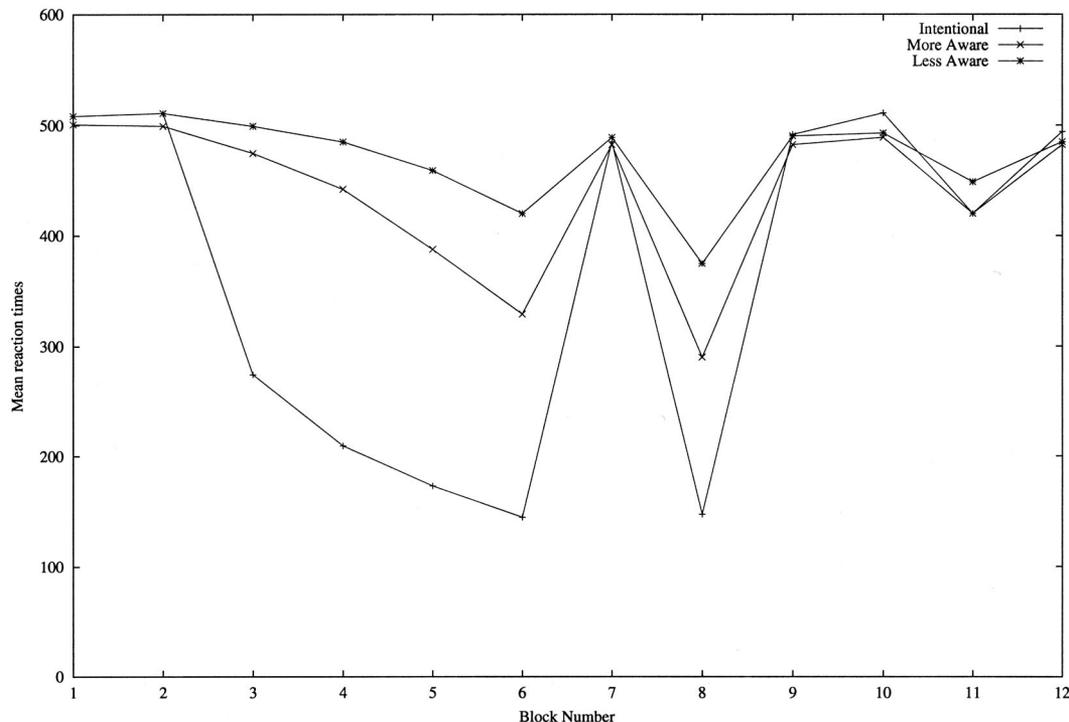
Figure 8. The simulation by Cleeremans (1993). Mean reaction times are in milliseconds. See the text for further explanation. Adapted from "Attention and Awareness in Sequence Learning," by A. Cleeremans, in *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (p. 334), 1993, Mahwah, NJ: Erlbaum. Copyright 1993 by the Cognitive Science Society. Adapted with permission.

data in which the differences due to instructions or awareness disappeared because of the interference of the secondary task.[38]

A comparison between Lewicki et al.'s (1987) task and this task demonstrated in some sense the salience difference effect: When stimuli were highly nonsalient as in Lewicki et al., the learning process was completely implicit. When stimuli were more salient as in this task, some explicit learning was involved. This difference was confirmed by our simulation.

Simulating DC Tasks

In this type of task, we aim to capture (a) the verbalization effect, (b) the explicit (how-to) instructions effect, (c) the explicit search effect, and (d) the salience difference effect. Through capturing the verbalization effect and the explicit instructions effect, we at the same time capture the synergy effect, which can be discerned through contrasting these two conditions with the standard condition (which indicates that the enhancement of the top level leads to better performance). The simulation of this task shows that the division of labor between (and the interaction of) the two levels is important, and bottom-up learning can be used to capture performance of this task.

In the bottom level, we use simplified Q-learning as described earlier. The input consists of a moving window. A window size of seven was adopted (i.e., seven pairs of system input and output of the immediate past seven steps were included).[39]

At the top level, rules for this task are mostly numerical relations, which cannot be generalized or specialized the same way as before. Therefore we use IRL, mentioned earlier—hypothesizing

rules and then testing them through experience. Initially, we hypothesize rules of a certain form to be tested. When the IG measure of a rule falls below the specialization threshold, we delete the rule. Whenever all the rules of a certain form are deleted, a new set of rules of a different form are hypothesized, and the cycle repeats itself. In hypothesizing rules, we progress from the simplest rule form to the most complex, in the order as shown below, in accordance with commonly used numerical relations (Berry & Broadbent, 1988; Stanley et al., 1989):

1. $P = aW + b$.

2. $P = aW + cP_1 + b$.

3. $P = aW_1 + b$.

4. $P = aW_1 + cP_2 + b$.

In these rule forms, $a = 1, 2$; $b = -1, -2, 0, 1, 2$; $c = -1, -2, 1, 2$; P is the desired system output level (the goal); W is the current input to the system (to be determined); W_1 is the previous input to the system; P_1 is the previous system output level (under

---

[38] The dual-task effect also lent support to the synergy effect because it showed that less involvement of explicit processes led to worsened performance (on the basis of our interpretation).

[39] The moving windows was necessary because we needed a trace of past activities to make current action decisions. Seven steps were assumed to be within the capacity of the working memory.

W_1); and P_2 is the system output level at the time step before P_1. Other rule forms can be easily added to the hypothesis testing process.

The immediate reward at each step (i.e., $r > threshold$) determines both positivity and rule extraction (where $threshold = 0.999$). The IG measure compares a rule in question and the random rule (which selects actions randomly according to a uniform distribution).[40]

To capture the explicit (how-to) instructions condition, the verbalization condition, and the explicit search condition, we used the changes of parameter values as outlined earlier. To capture the salience difference effect, however, we did not need to change any parameters: The effect falls out of the salience difference of the stimulus materials presented to subjects.

Simulating Stanley et al. (1989)

The task. Two versions of the DC task were used in Stanley et al. (1989). In the person version, subjects were to interact with a computer simulated "person" whose behavior ranged from "very rude" to "loving" (over a total of 12 levels), and the task was to maintain the behavior at "very friendly" by controlling his or her own behavior (which could also range over the 12 levels, from "very rude" to "loving"). In the sugar production factory version, subjects were to interact with a simulated factory to maintain a particular production level (out of a total of 12 possible production levels), through adjusting the size of the workforce (which also had 12 levels). In either case, the behavior of the simulated system was determined by $P = 2 \times W - P_1 + N$, where P was the current system output, P_1 was the previous system output, W was the subjects' input to the system, and N was noise. Noise was added to the output of the system, so that there was a chance of being up or down one level (a 33% chance, respectively).

There were four groups of subjects. The control group was not given any explicit how-to instruction and was not asked to verbalize. The verbalization group was required to verbalize after each block of 10 trials. Other groups of subjects were given explicit instructions in various forms, for example, "memory training," in which a series of 12 correct input–output pairs was presented to subjects, or "simple rules," in which a simple heuristic rule ("always select the response level halfway between the current production level and the target level") was given to subjects. There were 12–31 subjects in each group. All the subjects were trained for 200 trials (20 blocks of 10 trials).

The data. The exact target value ± 1 level was considered on target. The mean scores (numbers of on-target responses) per trial block for all groups were calculated. Analysis showed that the score for the verbalization group was significantly higher than that of the control group ($p < .05$). Analysis also showed that the scores for the memory training group and for the simple rule group were also significantly higher than that of the control group ($p < .0001$). See Table 4.

The model setup. The model was set up as described earlier. There were seven groups of input units, each for a particular time step, constituting a moving time window. Each group of input units contained 24 units, in which half of them encoded 12 system output levels and the other half encoded 12 system input levels at a particular step.

The rule deletion (specialization) threshold was set at 0.15 for simulating control subjects. To capture the verbalization condition, the threshold was raised to 0.35 to encourage more rule learning activities.[41] To capture the explicit instructions conditions, in the memory training condition, each of the 12 examples was wired up at the top level as a rule (in the form of $P_1 \rightarrow W$); in the simple rule condition, the simple rule (described earlier) was wired up at the top level. A reward of 1 was given when the system output was within the target range. In simulating the person task (a common, everyday task), we used pretraining of 10 blocks before data collection, to capture prior knowledge subjects likely had in this type of situation. See Table 5 for a summary of parameter values used.

The match. The simulation captured the human data well. The mean square error between the human and the model data was only 0.19. See Table 4.

Our simulation captured the verbalization effect in the human data. We used a *t* test to compare the verbalization group with the control group in the simulation data, which showed a significant improvement of the verbalization group over the control group ($p < .01$), the same as in the human data.

Our simulation also captured the explicit instructions effect. We used pairwise *t* tests to compare the memory training and simple rule groups with the control group, which showed significant improvements of these two groups over the control group, respectively ($p < .01$). Although the relative order of the two groups was reversed in the simulation of the person task, the difference between the simulation data of the two groups was not statistically significant, the same as in the human data (Stanley et al., 1989).

Simulating Berry and Broadbent (1988)

The task. The task was similar to the person task in Stanley et al. (1989). Subjects were to interact with a computer simulated person whose behavior ranged from "very rude" to "loving," and the task was to maintain the behavior at "very friendly" by controlling his or her own behavior. In the salient version of the task, the behavior of the person was determined by the immediately preceding input of the subject: It was on average two levels lower than the input ($P = W - 2 + N$). In the nonsalient version, it was determined by the input before that and was again two levels lower than that input ($P = W_1 - 2 + N$). Noise (N) was added to the output of the person so that there was a chance of being up or down one level (a 33% chance, respectively).

Four groups of subjects were tested: salient experimental, salient control, nonsalient experimental, and nonsalient control. The experimental groups were given explicit search instructions after the first set of 20 trials and, after the second set of 20 trials, were given

---

[40] This is analogous to the calculation of IG in regular RER. That is, if $IG(C, random) < threshold3$, we delete the rule C, where *random* refers to completely random actions. The match-all rule is not used here for comparison purposes, because here rules are generic: That is, in different input states, different actions may be recommended. The match-all rule thus becomes a random-action rule. This measure is equivalent to the following: If

$$IG(C) = \frac{PM(C) + 1}{PM(C) + NM(C) + 2} < threshold4,$$

we delete the rule C.

[41] Different from RER, with IRL, higher thresholds lead to more rule learning activities in the top level.

explicit how-to instructions in the form of an indication of the relevant input that determined system output. Twelve subjects per group were tested.

The data. As before, the exact target value ± 1 level was considered on target. The average number of trials on target was recorded for each subject for each set of 20 trials. Figure 9 shows the data of the four groups of subjects for the three sets of trials. An ANOVA (Salience × Condition × Set) showed significant main effects of salience, $F(1, 44) = 17.56$, $p < .01$; condition (experimental vs. control), $F(1, 44) = 6.59$, $p < .05$; and set, $F(2, 88) = 52.98$, $p < .001$, as well as significant interactions among these factors. A post hoc Newman–Keuls test showed that on the first set, neither of the two experimental groups differed significantly from their respective control groups; however, on the second set, the salient experimental group scored significantly higher than did the salient control group ($p < .01$), but the nonsalient experimental group scored significantly less than did the nonsalient control group ($p < .05$). On the third set, both experimental groups scored significantly higher than did their respective control groups ($p < .01$). The data clearly showed the explicit search effect (improving performance in the salient condition and worsening performance in the nonsalient condition) and the explicit instructions effect (improving performance in all conditions) as well as the salience difference effect (under the explicit search condition).

The model setup. The model was similar to the one described earlier for simulating Stanley et al.'s (1989) study, except for the following differences. The rule deletion threshold was set at 0.1 initially. To capture the explicit search effect (during the second training set), the rule deletion threshold was raised to 0.5 (for increasing learning activities at the top level),[42] and the weighting of the two levels was changed to 0.5/0.5 (for more reliance on the top level). To capture the explicit instructions given in this task (before the third training set), only rules that related the given critical variable to the system output were hypothesized and tested at the top level thereafter, in correspondence with the instructions (i.e., $P = aW + b$, where W is the critical variable indicated by the instructions). See Table 6 for a summary of parameter values used.

The match. We captured in our simulation of this task the following effects exhibited in the human data: the salience difference effect, the explicit search effect, and the explicit instructions effect. The results of the simulation were as shown in Figure 9. On

Table 4

The Human and Model Data for the Dynamic Control Task of Stanley et al. (1989)

Subject group	Sugar task	Person task
Human data		
Control	1.97	2.85
Verbalization	2.57	3.75
Memory training	4.63	5.33
Simple rule	4.00	5.91
Model data		
Control	2.276	2.610
Verbalization	2.952	4.187
Memory training	4.089	5.425
Simple rule	4.073	5.073

Table 5

Parameters Used in Simulating Data From Stanley et al. (1989)

Parameter	Value
Top-level learning: IRL	
Threshold	0.999
Threshold3	0.15/0.35
Bottom-level learning: QBP	
No. of input units	168
No. of output units	12
No. of hidden units	40
Learning rate	0.1
Momentum term	0.1

Note. IRL = independent rule learning; QBP = Q-learning-backpropagation algorithm.

the first set, neither of the two experimental groups differed significantly from their respective control groups; however, on the second set, the salient experimental group scored higher than the salient control group ($p < .05$), but the nonsalient experimental group scored less than the nonsalient control group ($p < .05$). On the third set, both experimental groups scored significantly higher than their respective control groups ($p < .01$).

Note that the differences during the second training set between control and experimental groups were smaller in the simulation. This was probably because we limited the simulation difference between control and experimental groups to two parameter values only, whereas in human performance more changes might have occurred. Nevertheless, we showed that our interpretation was able to capture the essential difference between control and experimental groups.

Discussion. Overall, the simulations and the human data of the dynamic control tasks confirmed the verbalization effect and the explicit instructions effect. Both effects point to the positive role of the top level. When the top level is enhanced, either through verbalization or through externally given explicit how-to instructions, performance is improved (although such improvement is not universal; see Sun et al., 2001). Thus, they both point to synergy between top-level explicit processes and bottom-level implicit processes. The simulations and the human data also showed, to some extent, the explicit search effect (improving performance in the salient condition and worsening performance in the nonsalient condition) as well as the salience difference effect.

Simulating High-Level Cognitive Skill Learning Tasks

We have simulated relevant data on the Tower of Hanoi task and a minefield navigation task as examples of high-level cognitive skill learning. Tower of Hanoi has been used extensively in cognitive skill acquisition research and is typical of the type of task addressed in such research. Therefore, addressing the task from the perspective of the implicit–explicit interaction adds to the litera-

---

[42] Different from RER, with IRL, higher thresholds lead to more rule learning activities in the top level.
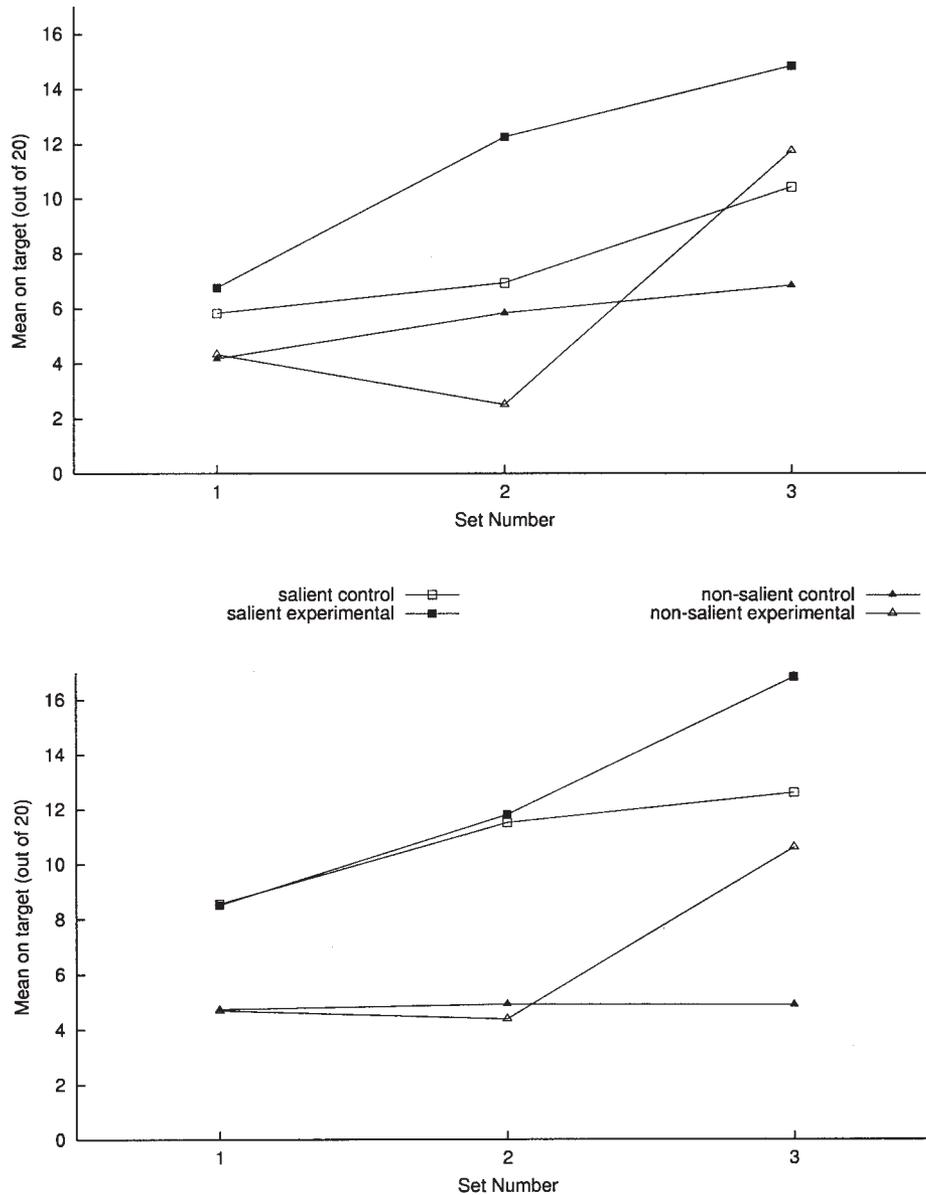
Figure 9. Top: Berry and Broadbent's (1988) data. Bottom: CLARION's simulation of Berry and Broadbent's (1988) data. The top panel is adapted from "Interactive Tasks and the Implicit-Explicit Distinction," by D. Berry and D. Broadbent, 1988, *British Journal of Psychology, 79,* p. 259. Copyright 1988 by the British Psychological Society. Adapted with permission.

ture in useful ways. The minefield navigation task serves as a further example to reinforce the idea that many complex real-world tasks involve implicit–explicit interactions (Mathews et al., 1989; Reber, 1989; Willingham, 1998). Because the details of the simulations of these two tasks have been published before, we do not repeat them here (see Sun et al., 2001; Sun & Zhang, 2004, for full details of the simulations).

We briefly summarize the results as follows: In both human data and simulations, the verbalization condition led to significantly better performance than the standard condition (in both Tower of Hanoi and minefield navigation). The standard (single-task) condition likewise led to significantly better performance than the dual-task condition (in minefield navigation). These results supported the point that explicit processes at the top level helped to improve learning (i.e., the synergy effect).

Similar to the previously discussed simulations, the model accounted for these two tasks through the combined learning process using QBP (at the bottom level) and RER (at the top level). The variations in the rule learning thresholds helped to capture the difference among different experimental conditions. Interaction of the two levels and bottom-up learning were important in capturing the two data sets.

We can briefly point out the difference between these two tasks and the previous tasks: These two tasks are more complex, in the

Table 6

Parameters Used in Simulating Data From Berry and Broadbent (1988)

Parameter	Value
Top-level learning: IRL	
Threshold	0.999
Threshold3	0.1/0.5
Bottom-level learning: QBP	
No. of input units	168
No. of output units	12
No. of hidden units	40
Learning rate	0.04
Momentum term	0
Weights of two levels	0.2–0.8/0.5–0.5

Note. IRL = independent rule learning; QBP = Q-learning-backpropagation algorithm.

sense that they have a lot more input dimensions, have a lot more different types of input, involve complex mappings from input to action decisions, often have no uniquely correct action decision, and are highly sequential. Thus, these tasks capture more real-world characteristics and tap more into real-world skill learning than did the other tasks we tackled (Sun et al., 2001).

General Discussion

Potential Controversies

Clearly, there may be alternative interpretations of the data presented here that do not involve the assumption of two levels and may be more or less equally compelling. However, alternatives notwithstanding, the two-level approach provides a consistent, theoretically motivated (Smolensky, 1988; Sun, 1994), and principled framework. The approach succeeds in interpreting many findings in skill learning that have not yet been adequately captured in computational modeling (such as bottom-up learning and synergy effects) and points to a way of incorporating such findings in a unified model, which has significant theoretical implications (see *Theoretical Implications of the Model,* below). This is where the novelty and significance of our model lie.

One-Level Models

Although it is conceivable that a one-level model may be designed so as to capture the data, we failed in our experiments to do so. However, the human data do not unambiguously point to our model. It is still possible that some one-level models may work. One may argue that if a one-level model can account for the data, then there is no need for the second level. However, it is seldom, if ever, the case that human data can be used to demonstrate the unique validity of a cognitive architecture. We need to rely on converging evidence from various sources, for example, philosophical accounts, to justify a model. By such a standard, this model fares well.

Implicit Learning

Implicit learning is admittedly a controversial topic. But the existence of implicit processes in skill learning is not in question—

what is in question is their extent and importance (Cleeremans et al., 1998; Stadler & Frensch, 1998). We allow for the possibility that both types of processes and both types of knowledge coexist and interact with each other to shape learning and performance; so, we manage to go beyond the controversies that focused mostly on the minute details of implicit learning.

For example, some criticisms of implicit learning focused on the alleged inability to isolate processes of implicit learning (e.g., Knowlton & Squire, 1994; Perruchet & Pacteau, 1990; Shanks & St. John, 1994). Such methodological problems are not relevant to our approach because we recognize that both implicit and explicit learning are present and that they are likely to influence each other in a variety of ways. Criticisms of implicit learning also focused on the degree of cognitive involvement in implicit learning tasks (e.g., Shanks & St. John, 1994). These criticisms are not relevant to our approach either because we make no claim in this regard. Yet another strand of criticisms concerned the fact that implicit learning was not completely autonomous and was susceptible to the influence of explicit cues, attention, and intention to learn (e.g., Berry, 1991; Curran & Keele, 1993; Stadler, 1995). These findings are consistent with our view of two interacting systems.

Explanations of Synergy

How is the synergy between the two separate, interacting components of the mind (i.e., the two types of processes) generated? Our model may shed some light on this issue by allowing systematic experimentations with the two corresponding levels in the model.

Sun and Peterson (1998a) did a thorough computational analysis of the source of the synergy between the two levels of CLARION in learning and in performance. Their conclusion, based on the systematic analysis, was that the explanation of the synergy between the two levels rests on the following factors: (a) the complementary representations of the two levels (discrete vs. continuous), (b) the complementary learning processes (one-shot rule learning vs. gradual Q-value approximation), and (c) the bottom-up rule learning criterion used in CLARION. (Because of length, we do not repeat the analysis here. See Sun and Peterson, 1998a, for details.)

It is very likely, in view of the match between the model and the human performance, that the corresponding synergy in human performance results also from these same factors (in the main). The analysis in the *Bottom-Up Learning; Differences in Representation of Resulting Knowledge; Differences in Accessibility of Resulting Knowledge;* and *Differences in Flexibility, Generalizability, and Robustness* sections identified distinct characteristics of the two levels in humans similar to the above three factors. It is conceivable that these same characteristics contribute to the generation of synergy in human performance (Breiman, 1996).

Comparisons With Other Models

Let us discuss existing simulation models concerning, or related to, the data sets and the effects dealt with in this work.

Connectionist Modeling of Skill Learning Tasks

Cleeremans and McClelland (1991) simulated an SRT task of their own design, which was an extension of usual SRT tasks (e.g.,

Willingham et al., 1989), by using nondeterministic grammars, and thus their task was more difficult. They used a recurrent back-propagation network that saw one position at a time but developed an internal context representation over time to help to predict next positions. Such a mechanism was more sophisticated than CLARION's general way of receiving input. The model succeeded in matching human data in terms of degrees of dependency (conditional probabilities) on preceding segments in a sequence. However, their success was obtained through introducing additional mechanisms for several types of priming (e.g., short-term weight changes and accumulating activations). They did not deal with capturing directly the reaction time data of their subjects.

Compared with CLARION, their model captured human data in the SRT task in a more fine-grained manner, whereas CLARION is a broader, more generic model that covers a broader range of tasks but at a coarser level (by necessity). Second, their recurrent network that received one input at a time was more sophisticated than CLARION's generic way of receiving input. However, the downside is that the additional mechanisms in their model, such as priming, although interesting, added to the complexity of the model in a task-specific way. Finally, their model did not deal with explicit knowledge and its learning in SRT tasks.

Dienes (1992) presented a comparison of some well-known models in the context of artificial grammar learning tasks. He compared connectionist networks (partially or fully recurrent), using the delta learning rule (which was similar to backpropagation) or the Hebb rule (which focused on direct associations), and a number of memory-array models (including instance-based models; e.g., Hintzman, 1986). He attempted to match these models with the human data on a number of measures, including percentage correct, rank ordering of string difficulty, percentage of strings on which an error was made on all the presentations, and percentage of strings on which an error was made only on some of presentations. These models were successful in terms of accounting for the human data that were examined (which, however, were not concerned with the interaction of the two types of knowledge).

One general shortcoming of these above models is that mostly, these models focused only on implicit learning and they ignored (a) the role of explicit learning in these tasks, (b) the interaction between the explicit and the implicit in learning and performing the tasks, and (c) the possibility of bottom-up learning. The only exception to the first two points above is Cleeremans's (1993) model.

Cleeremans (1993) used a simple buffer network to capture the effect of explicit knowledge. The buffer network mimicked explicit retrieval of explicitly stored items, through a backpropagation network that has a buffer as part of its input. The output of the buffer network was fed into the hidden units of the main network (a recurrent network for capturing implicit learning). The outcomes from the two networks were thus combined before the final output was produced. Dual-task conditions were simulated by adding noise, mostly to the buffer network. This is a simple and parsimonious solution for SRT tasks. However, in other, more complex types of tasks, such as DC or minefield navigation, the buffer network, as is, is inadequate for capturing explicit knowledge used in performing these tasks. Furthermore, in more complex tasks, implicit and explicit knowledge may have more complex interactions. Whereas CLARION can accommodate more complex interaction (using its combination mechanisms), this model may have trouble doing so. Thus, this model is more limited

than CLARION. See the *Simulating Curran and Keele (1993)* section for a quantitative comparison.

Nonconnectionist Modeling of Skill Learning Tasks

As mentioned before, Ling and Marinov (1994) simulated the data from Lewicki et al. (1987), using a decision tree learning algorithm (i.e., C4.5). The decision tree algorithm iteratively divides up a set of states into subsets, in an attempt to maximize output prediction consistency. A tree structure emerges through the iterative process. Their model produced data on quadrant prediction accuracy, and on the basis of the data, they succeeded in matching the human reaction time data, using a transformation that included a power function (for capturing unspecific learning). However, they did not attempt the match without such a power function (as we did). See the *Simulating Lewicki et al. (1987)* section for a quantitative comparison of the two simulations.

Dienes and Fahey (1995) developed an instance-based model for a DC task. Their model was based on acquiring successful instances from trials (as in Logan, 1988, and Hintzman, 1986, with instances being filtered according to a performance based criterion), supplemented by a set of a priori rules to start with.[43] However, it is not clear how the model can account for gradual explication of implicit knowledge, for example, as reported in Stanley et al. (1989). Dienes and Fahey (1995) also examined an alternative model, which focused instead on hypothesizing and testing rules (without using instances), accomplished through competitions among rules. They found that neither model fitted the data completely: Whereas the instance model fitted better the nonsalient version of the task, the rule model fitted better the salient version of the task. This fact, to us, suggests that it may be advantageous to include both types of learning in one unified model (as in CLARION). In such a model, the effect of salience difference results from the interaction of the two learning processes: That is, the top level handles well salient tasks, and therefore rules mostly account for the learning, but the top level cannot handle well nonsalient ones, and therefore instance-based processes (as in the bottom level) account for the learning. Thus, CLARION appears to be a more complete model, and it explains the choice between the two types of processes.

Lebiere, Wallach, and Taatgen (1998) simulated Dienes and Fahey's (1995) data using ACT–R (see discussion on ACT–R later). The simulation was based on a combination of instance-based learning implemented in ACT–R and a set of hand-coded a priori rules similar to those used in Dienes and Fahey's model. A good fit with the data was found. It was not clear, though, how the model could account for gradual explication of implicit knowledge as in Stanley et al. (1989). Along with other similar models built with ACT–R, this model was interesting in that implicit learning was accounted for in a unified framework of a cognitive architecture. However, from our perspective, a fundamental shortcoming of their model is that there was no principled distinction between, and a physical separation of, implicit knowledge and explicit knowledge because of the representational framework of ACT–R (to be discussed later).

---

[43] The use of these rules were justified on the basis of the observation of the initial set of moves made by subjects (which presumably reflected the a priori knowledge of the subjects).

Servan-Schreiber and Anderson (1987) presented a model involving a simple mechanism, which they called chunking. The model was used to account for artificial grammar learning data. In that task, chunking consists of merging fragments of letter strings into larger ones if these fragments show up often enough. Their model essentially remembers frequently seen letter string fragments (while it merges fragments according to the chunking mechanism). When the model encounters a test string, answers are produced by comparing the test string with retrieved chunks (i.e., fragments). The closer they are, the more likely it is that the test string will be recognized as familiar. Because of its simplicity, the model cannot account for the interaction of the two types of knowledge and those resulting effects that we simulated.

Top Down Versus Bottom Up

As mentioned earlier, a number of theories of skill learning assumed the distinction between procedural (implicit) knowledge and declarative (explicit) knowledge, but none of them dealt with bottom-up learning. For example, Anderson (1983, 1993) put forth two similar cognitive architectures: ACT* and ACT–R. ACT* is made up of two components: a semantic network for representing declarative knowledge and a production system for representing procedural knowledge. Productions are formed through "proceduralization" of declarative knowledge. They are modified through use by generalization and discrimination (i.e., specialization) and have strengths associated with them that are used for firing. ACT–R is a descendant of ACT*, in which procedural learning is limited to production formation through mimicking and production firing is based on log odds of success. Both models deal mostly with top-down learning.

Hunt and Lansman (1986) hypothesized another top-down learning process for explaining automatization data (which top-down learning models fit most naturally). They incorporated two separate components in their model: They used a production system for capturing controlled processes and a semantic network for capturing automatic processes. They hypothesized that through practice, production rules were assimilated into the semantic network, thus resulting in automatic processes (through spreading activation in the semantic network). They implemented this assimilation process in their model through adjusting weights of the semantic network on the basis of production firing.

The implementation of the two types of knowledge is similar between CLARION and Hunt and Lansman's (1986) model. The production system in Hunt and Lansman's model clearly resembles the top level in CLARION, in that explicit rules are used in much the same way. Likewise, the spreading activation in the semantic network in Hunt and Lansman's model resembles spreading activation in the bottom level of CLARION. However, learning directions are different across these two models. Whereas learning in CLARION is mostly bottom up but capable of being top down, learning in their model is completely top down: That is, the working of the production system is assimilated into the semantic network, but the opposite process is not available. This makes their model more limited than CLARION (which is capable of both directions).[44]

Evidently, human learning is not exclusively top down. As reviewed before, top-down approaches were contradicted by much evidence, such as in studies by Schraagen (1993), Owen and Sweller (1985), Rabinowitz and Goldberg (1995), Reber and Lewis (1977), and Willingham et al. (1989). A distinct feature of CLARION is its ability of going the other way around, capturing bottom-up learning, which makes it unique, more complete, and complementary to the afore-reviewed top-down models.

Models of Implicit Learning

Let us summarize the above comparisons in relation to modeling implicit learning. Models of implicit learning are central to this work because of our focus on the interaction between implicit and explicit learning. There are, in general, two types of computational models for implicit learning. The first type is neural network models such as that of Cleeremans and McClelland (1991), and the second type is stored data models. The second type can be instance based (e.g., Logan, 1988), rule based (e.g., the second model of Dienes & Fahay, 1995), fragment based (with only fragmentary knowledge; e.g., Servan-Schreiber & Anderson, 1987), or a combination thereof.

Although these models are quite different from each other (some of which have been reviewed earlier), they share some important commonalities (identified in Cleeremans et al., 1998): (a) Learning is incremental and ongoing (Sun et al., 2001), (b) learning is closely tied to details of instances being processed at each step, (c) learning is autonomous in the sense that it is not controlled by other processes and is generally "self-organizing," and (d) learning is sensitive to statistical structures embodied in stimuli (Stadler, 1992).

Because of these common features, we feel justified in adopting a neural network model that has these features to capture implicit learning in CLARION (while using a radically different mechanism for capturing explicit learning). Alternative models of implicit learning are also conceivable, provided that the difference in accessibility between implicit and explicit knowledge (as highlighted in the A Model section) can be accounted for somehow.

Theoretical Implications of the Model

A New Interpretation of Data and Issues

Compared with existing theories, approaches, and perspectives concerning skill learning, CLARION is distinct in its emphasis—learning through the interaction of implicit and explicit processes and in a mostly bottom-up (implicit-to-explicit) direction.

Because of this emphasis, it offers a new way of interpreting skill learning data and a new direction for theorizing about and experimentation with human cognition. Many new possibilities emerge that may be worth exploring further. For example, we may further explore the possibility of synergy and bottom-up learning in high-level cognitive skill acquisition, in which usual ways of structuring the tasks in experiments make them appear to be completely top down and (a priori) knowledge driven. If we follow a more data-driven trial-and-error approach in experiments, we may gain new insight into high-level tasks and reveal complex interactions between implicit and explicit knowledge in such tasks (cf. Shrager, 1990). As another example, this new perspective also enables us to look into developmental and aging studies in new

---

[44] Schneider and Oliver (1991) used essentially the same idea. Logan (1988), who also meant to capture automatization data, was also somewhat similar in this regard.

ways. How does the interaction between implicit and explicit processes change during development or during aging? Does the change in relative contribution from explicit processes explain the change in cognitive styles (Karmiloff-Smith, 1986; Keil, 1989)? Is the role of bottom-up learning constant throughout development and aging? Many more questions can be raised and explored along this line.

Let us explicate the relationship between the implicit–explicit distinction that we emphasize here and the procedural–declarative distinction in some other theories. In Anderson's (1983, 1993) studies, procedural knowledge is represented in an action-oriented way (using production rules that can be used only in one direction—from conditions to actions), and declarative knowledge is represented in a non-action-oriented way (i.e., with knowledge chunks that can be used in any possible direction). The difference in action-orientedness seems to be the main factor in distinguishing the two types, whereas explicit accessibility seems a secondary factor.[45] However, because CLARION is focused on skill learning (which is by definition action oriented), explicit knowledge in CLARION is mostly action oriented, and thus it is not declarative in the above sense of declarative knowledge. But it is declarative if we define declarativeness in terms of accessibility (Sun & Peterson, 1998a). The two dichotomies overlap to a large extent and can be reconciled if we adopt this alternative definition. We believe that the current view of declarative knowledge unnecessarily confounds two issues—action-orientedness and accessibility—and can be made clearer by separating the two issues (as in CLARION). As demonstrated in CLARION, action-orientedness does not necessarily go with inaccessibility (Sun et al., 2001), and non-action-orientedness does not necessarily go with accessibility either (e.g., priming and implicit memory; see, e.g., Schacter, 1987). Our perspective on this issue is close to Hunt and Lansman's (1986), because they separated two types of processes (controlled vs. automatic) on the basis of accessibility and representational differences instead of action-orientedness (in fact both types of knowledge in their model were action oriented, the same as in CLARION).

The issue of automaticity (automatic vs. controlled processes) is also relevant to CLARION. The notion of automaticity has been variously associated with (a) the absence of competition for limited resources (attention) and thus the lack of performance degradation in multitask settings (Navon & Gopher, 1979), (b) the absence of conscious control and/or intervention in processes (J. Cohen, Dunbar, & McClelland, 1990), (c) the general inaccessibility of processes (Logan, 1988), and (d) the general speed up of skilled performance (Hunt & Lansman, 1986). Although we did not focus on these issues, CLARION is certainly compatible with them. The top level can account for controlled processes (the opposite of these above properties), and the bottom level has the potential of accounting for all the aforementioned properties of automatic processes. In the foregoing discussions, we have in fact separately covered these issues: the speed up of skilled performance (see the Simulation of Human Skill Learning Data section); the direct inaccessibility of processes at the bottom level, including their ability of running without conscious intervention (see the A Model and Analysis of Interaction sections); and the lack of resource competition (due to the existence of multiple bottom-level modules that can run in parallel; see the A Model section). Thus, in CLARION, automaticity serves as an umbrella term that

describes a set of phenomena occurring in implicit processes at the bottom level.

There has also been the distinction between the unidimensional and the multidimensional system of sequence learning (Keele et al., 2003). One system may lead to awareness, and the other cannot. Examining the data reported in support of the idea, we noticed that the data were ambiguous and could also be consistent with CLARION. Specifically, it was argued that stimuli in two different modalities formed a single sequence in the multidimensional system, and if such a sequence was too complex, there was no awareness. This idea was consistent with CLARION. But Keele et al. (2003) downplayed the fact that complex single-modality sequences might also prevent awareness (e.g., Lewicki et al., 1987) and did not explain why complex single-modality sequences prevented awareness, which CLARION could explain (on the basis of its rule learning mechanism). They did not delineate conditions for awareness (beside being associated with the multidimensional system), whereas CLARION did provide some clues (in terms of its rule learning mechanism).[46]

The issue of accessibility (and the way it is accounted for) should be of major importance to theories of cognition, considering its intimate relationships to various fundamental dichotomies in cognition (such as implicit vs. explicit, subconceptual vs. conceptual, procedural vs. declarative, automatic vs. controlled, and unconscious vs. conscious). In most existing theories, the difference between accessible and inaccessible processes is simply assumed, without grounding in representational forms. In other words, the difference is not intrinsic to representational media. For example, in ACT models (Anderson, 1983, 1993; Anderson & Lebiere, 1998), both declarative and procedural knowledge are represented in an explicit, symbolic form (one with semantic networks and the other with productions, along with some numerical measures). Thus the ACT models do not explain, from a representational viewpoint, the difference in accessibility between the two types of knowledge. SOAR (Rosenbloom, Laird, & Newell, 1993) does not separate the two types of knowledge: To account for the difference in conscious accessibility, it assumes the inaccessibility of the working of individual productions, so as to distinguish implicit and explicit processes with the difference of a single production versus multiple productions. CLARION, however, accounts for this difference on the basis of the use of two different forms of representations: The inaccessibility of implicit knowledge is captured by subsymbolic distributed representations provided by a backpropagation network (Rumelhart et al., 1986). The accessibility of explicit knowledge is captured by symbolic–localist representations at the top level of CLARION. Thus, this distinction in CLARION is intrinsic instead of assumed. We sug-

---

[45] A common interpretation is that whereas procedural knowledge is inaccessible, declarative knowledge consists of both accessible symbolic representations and inaccessible subsymbolic representations.

[46] If the hypothesis of the coexistence of unidimensional and multidimensional systems is true, it can be incorporated into CLARION: That is, both systems can be viewed as parts of the bottom level, and the top level can extract explicit knowledge from only one of the two systems, the multidimensional system. Dual-task conditions tend to force, in the bottom level, the dominance by the unidimensional system because the conditions tend to confuse the multidimensional system as argued by Keele et al. (2003).

gest that this is a more principled way of accounting for the accessibility difference.

The implicit–explicit distinction bears clear relationships to the study of consciousness because this distinction involves, in its core, the issue of awareness, which is the key to consciousness no matter which philosophical position one subscribes to. The study of the implicit–explicit distinction may help us to better understand issues concerning consciousness by identifying physical mechanisms and processes correlated with consciousness (Dienes & Perner, 1999; Reber, 1989; Schacter, 1987; Sun, 1997, 1999). In this regard, CLARION may shed light on the issue of what constitutes consciousness. Our central thesis has been that direct accessibility, along with explicit manipulability (on directly accessible representations), constitutes the essence of consciousness (see Sun, 1997, 1999; cf. Dienes & Perner, 1999). CLARION naturally embodies the difference between accessibility and inaccessibility through the use of localist–symbolic and distributed representations in different levels and provides a plausible grounding for the notion of accessibility. Although there are a variety of views concerning consciousness, each on the basis of a different physical substrate,[47] Sun (1997, 1999) argued that the distinction between localist–symbolic and distributed representations provided a far superior alternative. Therefore, CLARION has bearings on theorizing on consciousness.

A New Look at Modeling Paradigms

In relation to the above issues, let us reexamine the controversy of connectionist models versus symbolic models (Fodor & Pylyshyn, 1988; Smolensky, 1988). First, there was the question of which paradigm should be adopted as a general cognitive modeling framework. This has been an issue of great controversy among theoretically minded cognitive scientists. Many claims and counterclaims have been made. CLARION sidesteps this stalemate by incorporating both paradigms, in a principled way, into its architecture. We show that the two can be combined to generate synergy in skill learning, which in turn suggests the general advantage of this combination. CLARION is one of many so-called hybrid models that started in the late 1980s and are receiving increasing attention lately (see Sun, 1994; Sun & Alexandre, 1997; Sun & Bookman, 1994).

In relation to this issue, there is also the more specific issue of the ability (or the inability) of one type of model or the other in accounting for implicit learning (i.e., subconceptual processes; Smolensky, 1988). It has been claimed, on the connectionist side, that a vast majority of human activities (i.e., implicit processes), including "perception, motor behavior, fluent linguistic behavior, intuition in problem solving and game playing—in short, practically all skilled performance" (Smolensky, 1988, p. 5), should be modeled by subsymbolic computation (connectionist models), and symbolic models can give only an imprecise and approximate explanation of these processes (Smolensky, 1988). It has also been claimed, on the symbolicist side, that "one and the same algorithm" can be "responsible for conscious and nonconscious processes alike" (Ling & Marinov, 1994, p. 619), or even that implicit learning "should be better modeled by symbolic rule learning programs" (Ling & Marinov, 1994, p. 596; see also Fodor & Pylyshyn, 1988). We believe that this issue is a red herring: Being able to simulate some data of implicit learning amounts to very little, in that any Turing equivalent computational process (i.e., any

generic computational model) should be able to simulate these data. Thus, the simulation of data by itself does not prove whether a particular model is a suitable one (Cleeremans, 1997). Other considerations need to be brought in to justify a model. We suggest that one such issue is the accessibility issue discussed above. Whereas symbolic models of implicit learning lead to explicit symbolic representations of implicit knowledge (e.g., Lebiere et al., 1998; Ling & Marinov, 1994) that are evidently accessible (without using any add-on auxiliary assumptions), connectionist models of implicit learning lead to implicit (subsymbolic) representations of resulting knowledge that are inherently less accessible (such as in the bottom level of CLARION). Thus, connectionist models have a clear advantage: Being able to match human implicit learning data (at least) as well as symbolic models, they also account for the inaccessibility of implicit knowledge more naturally than symbolic models. In this sense, they are better models. However, it is generally accepted that symbolic–localist models have their roles to play too: They are better at capturing explicit processes. This contrast lends support to the belief that because connectionist models are good for implicit processes and symbolic models for explicit processes, the combination of the two types of models should be emphasized in modeling cognition (Smolensky, 1988; Sun, 1994, 1995, 1997).

We also want to explicate the relationship between our model and instance-based models (e.g., Dienes & Fahey, 1995; Logan, 1988). Logan (1988) showed that skill learning (automatization) could be captured by the acquisition of a domain-specific knowledge base that was composed of experienced instances represented in individuated forms (Hintzman, 1986). Shanks and St. John (1994) developed a theoretical perspective in which implicit learning was viewed as nothing more than learning instances (however, this perspective has been criticized for various failings). Stanley et al. (1989) also described implicit learning and performance as mainly the result of relying on memory of past instances, which were used by being compared with a current situation and being transformed into a response to the current situation (through similarity-based analogical processes). At first glance, these models may seem at odds with CLARION. However, on a closer examination, we see that connectionist networks used in the bottom level of CLARION can be either exemplar based (essentially storing instances; Kruschke, 1992) or prototype based (summarizing instances; Rumelhart et al., 1986). The similarity-based processes in these models can also be performed in connectionist networks, which are known to excel in such processes. Instance-based models, however, generally do not handle learning of generic explicit knowledge or bottom-up learning.

Finally, as a result of its distinct emphasis, CLARION is also clearly distinguishable from existing unified theories and/or architectures of cognition, such as SOAR, ACT, and EPIC. For example, SOAR (Rosenbloom et al., 1993) is different from CLARION because SOAR makes no distinction between explicit and implicit learning, and its learning is based on specialization, using symbolic representations. EPIC (Meyer & Kieras, 1997) is also different from CLARION because it makes no implicit–explicit distinction either, although it incorporates motor and perceptual processes as part of a cognitive architecture. Although ACT* and ACT–R

---

[47] There are of course also dualistic views that rely on the assumption of nonphysical properties, which we do not deal with here.

(Anderson, 1983, 1993) make the procedural–declarative distinction, they are different from CLARION because they traditionally place more emphasis on top-down learning, not bottom-up learning.

Concluding Remarks

This work highlights the importance of the interaction of implicit and explicit processes in skill learning (instead of focusing on minute details of implicit learning). It points to the usefulness of incorporating both explicit and implicit processes in theorizing about cognition in general.

It demonstrates the interaction through a model that captures both types of processes, with a particular emphasis on their interaction. We developed a unified model, CLARION, through examining a variety of relevant data (mostly from the implicit learning literature) and through capturing the data in our simulations (based on the CLARION model). This work reveals something new in the existing data. With fairly detailed comparisons between the human and the model data, these simulations shed light on plausible causes of these data.

The contribution of this model lies not only in capturing a range of human data in skill learning through the interaction of the two types of processes but also in demonstrating the computational feasibility and psychological plausibility of bottom-up learning, which complements the extensive existing treatment of top-down learning in the cognitive modeling literature and fills a significant gap in that literature. We show the possibility of synergy (as well as detrimental effects) that may result from interactions.

References

Ahlum-Heath, M., & DiVesta, F. (1986). The effect of conscious controlled verbalization of a cognitive strategy on transfer in problem solving. *Memory & Cognition, 14,* 281–285.

Aizenstein, H., MacDonald, A., Stenger, V., Nebes, R., Larson, J., Ursu, S., & Carter, C. (2000). Complementary category learning systems identified using event-related functional MRI. *Journal of Cognitive Neuroscience, 12,* 977–987.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Ben-Zur, H. (1998). Dimensions and patterns in decision-making models and the controlled/automatic distinction in human information processing. *European Journal of Cognitive Psychology, 10,* 171–189.

Berry, D. (1983). Metacognitive experience and transfer of logical reasoning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 35*(A), 39–49.

Berry, D. (1991). The role of action in implicit learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 43*(A), 881–906.

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

Berry, D., & Broadbent, D. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology, 79,* 251–272.

Bower, A., & King, W. (1967). The effect of number of irrelevant stimulus dimensions, verbalization, and sex on learning biconditional classification rules. *Psychonomic Science, 8,* 453–454.

Bower, G. (1996). Reactivating a reactivation theory of implicit memory. *Consciousness and Cognition, 5,* 27–72.

Bowers, K., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22,* 72–110.

Boyd, L., & Weistein, C. (2001). Implicit motor-sequence learning in humans following unilateral stroke: The impact of practice and explicit knowledge. *Neuroscience Letters, 298,* 65–69.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24,* 123–140.

Bruner, J., Goodnow, J., & Austin, J. (1956). *A study of thinking.* New York: Wiley.

Busemeyer, J., & Myung, I. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General, 121,* 177–194.

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

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, P. (1989). Self-explanation: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind & Language, 8,* 487–519.

Cleeremans, A. (1993). Attention and awareness in sequence learning. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 330–335). Mahwah, NJ: Erlbaum.

Cleeremans, A. (1997). Principles for implicit learning. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 195–234). Oxford, England: Oxford University Press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2,* 406–416.

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

Cohen, A., Ivry, R., & Keele, S. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 17–30.

Cohen, J., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: Toward an evolutionarily rigorous cognitive science. *Cognition, 50,* 41–77.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 189–202.

DeShon, R., & Alexander, R. (1996). Goal setting effects on implicit and explicit learning of complex tasks. *Organizational Behavior and Human Decision Processes, 65,* 18–36.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review, 8,* 343–350.

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

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4,* 3–23.

Dienes, Z., & Fahey, R. (1995). The role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 848–862.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22,* 735–808.

Dominowski, R. (1972). How do people discover concepts? In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 257–288). Potomac, MD: Erlbaum.

Dreyfus, H., & Dreyfus, S. (1987). *Mind over machine: The power of human intuition.* New York: The Free Press.

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

Elman, J. (1990). Finding structures in time. *Cognitive Science, 14,* 179–211.

Estes, W. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115,* 155–174.

Faulkner, D., & Foster, J. (2002). The decoupling of explicit and implicit processing in neuropsychological disorders: Insights into the neural basis of consciousness. *Psyche, 8*(2). Retrieved February 1, 2003 from http://psyche.cs.monash.edu.au/v8/psyche-8-02-faulkner.html

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. In S. Pinker & G. Mehler (Eds.), *Connections and symbols* (pp. 3–71). Cambridge, MA: MIT Press.

Frensch, P., Wenke, D., & Ruenger, D. (1999). A secondary tone-counting task suppresses expression of knowledge in the serial reaction task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 260–274.

Gagne, R., & Smith, E. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology, 63,* 12–18.

Gibson, F., Fichman, M., & Plaut, D. (1997). Learning in dynamic decision tasks: Computational model and empirical evidence. *Organizational Behavior and Human Decision Processes, 71,* 1–35.

Gick, M., & Holyoak, K. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355.

Gluck, M., & Bower, G. (1988). From conditioning to category learning. *Journal of Experimental Psychology: General, 117,* 227–247.

Halford, G., Wilson, W., & Philips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803–865.

Hasher, J., & Zacks, J. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–358.

Hayes, N., & Broadbent, D. (1988). Two modes of learning for interactive tasks. *Cognition, 28,* 249–276.

Haygood, R., & Bourne, L. (1965). Attribute and rule learning aspects of conceptual behavior. *Psychological Review, 72,* 175–195.

Heidegger, M. (1962). *Being and time.* New York: Harper and Row. (Original work published 1927)

Hintzman, D. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93,* 528–551.

Howard, J., & Ballas, J. (1980). Syntactic and semantic factors in classification of nonspeech transient patterns. *Perception and Psychophysics, 28,* 431–439.

Hunt, E., & Lansman, M. (1986). Unified model of attention and problem solving. *Psychological Review, 93,* 446–461.

Jimenez, L., & Mendez, C. (2001). Implicit sequence learning with competing explicit cues. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54*(A), 345–369.

Johnstone, T., & Shanks, D. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 42,* 61–112.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition, 23,* 95–147.

Keele, S., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110,* 316–339.

Keil, F. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Kemler-Nelson, D. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior, 23,* 734–759.

Kersten, A., & Billman, D. (1992). The role of correlational structure in learning event categories. In *Proceedings of the 14th Annual Meeting of the Cognitive Science Society* (pp. 432–437). Mahwah, NJ: Erlbaum.

Knowlton, B., & Squire, L. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 79–91.

Kruschke, J. (1992). ALCOVE: An examples-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Lavrac, N., & Dzeroski, S. (1994). *Inductive logic programming.* New York: Ellis Horword.

Lebiere, C., Wallach, D., & Taatgen, N. (1998). Implicit and explicit learning in ACT–R. In *Proceedings of the European Conference on Cognitive Modeling 1998* (pp. 183–189). Nottingham, England: Nottingham University Press.

Lee, Y. S. (1995). Effects of learning contexts on implicit and explicit learning. *Memory & Cognition, 23,* 723–744.

Lewicki, P., Czyzewska, M., & Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 523–530.

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *American Psychologist, 47,* 796–801.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8,* 529–566.

Ling, C. X., & Marinov, M. (1994). A symbolic model of the nonconscious acquisition of information. *Cognitive Science, 18,* 595–621.

Logan, G. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Mandler, J. (1992). How to build a baby. *Psychological Review, 99,* 587–604.

Mathews, R., Buss, R., Stanley, W., Blanchard-Fields, F., Cho, J., & 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., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

Mehwort, D., Braun, J., & Heathcote, A. (1992). Response time distributions and the stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 872–887.

Meyer, D., & Kieras, D. (1997). A computational theory of executive cognitive processes and human multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104,* 3–65.

Michalski, R. (1983). A theory and methodology of inductive learning. *Artificial Intelligence, 20,* 111–161.

Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, J. L. McGaugh, & N. M. Weinberger (Eds.), *Neurobiology of human learning and memory* (pp. 65–77). New York: Guilford Press.

Mitchell, T. (1998). *Machine learning.* New York: McGraw-Hill.

Moscovitch, M., & Umilta, C. (1991). Conscious and unconscious aspects of memory. In R. G. Lister & J. H. Weingartner (Eds.), *Perspectives on cognitive neuroscience* (pp. 229–266). New York: Oxford University Press.

Muente, S., Panning, B., Piepenbrock, S., & Muente, T. (2001). No implicit memory under propoful-alfentanil-anesthesia: The lexical decision task. *Current Opinion in Clinical Experimental Research, 3,* 63–70.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86,* 214–255.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 1977.

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

Nokes, T., & Ohlsson, S. (2001). How is abstract, generative knowledge acquired? A comparison of three learning scenarios. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 710–715). Mahwah, NJ: Erlbaum.

Nosofsky, R., Palmeri, T., & McKinley, S. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Experimental Psychology, 77,* 272–284.

Perruchet, P., & Gallego, J. (1993). Association between conscious knowledge and performance in normal subjects: Reply to Cohen and Curran (1993) and Willingham, Greeley, and Bardone (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1438–1444.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 118,* 264–275.

Plaut, D., & Shallice, T. (1994). *Connectionist modeling in cognitive neuropsychology: A case study.* Philadelphia, PA: Psychology Press.

Poldrack, R. A., Clark, J., Par-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001, November 29). Interactive memory systems in the human brain. *Nature, 414,* 546–550.

Pollack, J. (1991). The induction of dynamic recognizers. *Machine Learning, 7,* 227–252.

Posner, M., DiGirolamo, G., & Fernandez-Duque, D. (1997). Brain mechanisms of cognitive skills. *Consciousness and Cognition, 6,* 267–290.

Premack, D. (1988). Minds with and without language. In L. Weiskrantz (Ed.), *Thought without language* (pp. 46–65). Oxford, England: Clarendon Press.

Proctor, R., & Dutta, A. (1995). *Skill acquisition and human performance.* Thousand Oaks, CA: Sage Publications.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227–270). Cambridge, MA: MIT Press.

Rabinowitz, M., & Goldberg, N. (1995). Evaluating the structure-process hypothesis. In F. Weinert & W. Schneider (Eds.), *Memory performance and competencies* (pp. 225–242). Hillsdale, NJ: Erlbaum.

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

Reber, A. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 88–94.

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

Reber, A., & Allen, R. (1978). Analogy and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition, 6,* 189–221.

Reber, A., Kassin, S., 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 and Memory, 6,* 492–502.

Reber, A., & Lewis, S. (1977). Toward a theory of implicit learning: The analysis of the form and structure of a body of tacit knowledge. *Cognition, 5,* 333–361.

Rips, L. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge University Press.

Rosenbloom, P., Laird, J., & Newell, A. (1993). *The SOAR papers: Research on integrated intelligence.* Cambridge, MA: MIT Press.

Roussel, L. (1999). *Facilitating knowledge integration and flexibility: The effects of reflection and exposure to alternative models.* Unpublished doctoral dissertation, Louisiana State University.

Rumelhart, D., McClelland, J., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition.* Cambridge, MA: MIT Press.

Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518.

Schacter, D. (1990). Toward a cognitive neuropsychology of awareness: Implicit knowledge and anosagnosia. *Journal of Clinical and Experimental Neuropsychology, 12,* 155–178.

Schneider, W., & Oliver, W. (1991). An instructable connectionist/control architecture. In K. VanLehn (Ed.), *Architectures for intelligence* (pp. 113–145). Hillsdale, NJ: Erlbaum.

Schooler, J., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122,* 166–183.

Schraagen, J. (1993). How experts solve a novel problem in experimental design. *Cognitive Science, 17,* 285–309.

Seger, C. (1994). Implicit learning. *Psychological Bulletin, 115,* 163–196.

Servan-Schreiber, E., & Anderson, J. (1987). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 592–608.

Shallice, T. (1972). Dual functions of consciousness. *Psychological Review, 79,* 383–393.

Shanks, D. (1993). Human instrumental learning: A critical review of data and theory. *British Journal of Psychology, 84,* 319–354.

Shanks, D., & St. John, M. (1994). Characteristics of dissociable learning systems. *Behavioral and Brain Sciences, 17,* 367–394.

Shrager, J. (1990). Commonsense perception and the psychology of theory formation. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation* (pp. 437–470). San Mateo, CA: Morgan Kaufmann Publishers.

Siegler, R., & Stern, E. (1998). Conscious and unconscious strategy discovery: A microgenetic analysis. *Journal of Experimental Psychology: General, 127,* 377–397.

Slusarz, P., & Sun, R. (2001). The interaction of explicit and implicit learning: An integrated model. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 952–957). Mahwah, NJ: Erlbaum.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1–74.

Squire, L., & Frambach, M. (1990). Cognitive skill learning in amnesia. *Psychobiology, 18,* 109–117.

Stadler, M. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 318–327.

Stadler, M. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 674–685.

Stadler, M., & Frensch, P. (1998). *Handbook of implicit learning.* Thousand Oaks, CA: Sage.

Stanley, W., Mathews, R., Buss, R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 41*(A), 553–577.

Sun, R. (1992). On variable binding in connectionist networks. *Connection Science, 4,* 93–124.

Sun, R. (1994). *Integrating rules and connectionism for robust commonsense reasoning.* New York: Wiley.

Sun, R. (1995). Robust reasoning: Integrating rule-based and similarity-based reasoning. *Artificial Intelligence, 75,* 241–296.

Sun, R. (1997). Learning, action, and consciousness: A hybrid approach towards modeling consciousness. *Neural Networks, 10,* 1317–1331.

Sun, R. (1999). Accounting for the computational basis of consciousness: A connectionist approach. *Consciousness and Cognition, 8,* 529–565.

Sun, R., & Alexandre, F. (Eds.). (1997). *Connectionist symbolic integration.* Hillsdale, NJ: Erlbaum.

Sun, R., & Bookman, L. (Eds.). (1994). *Computational architectures*

integrating neural and symbolic processes. Norwell, MA: Kluwer Academic.

Sun, R., Merrill, E., & Peterson, T. (1998). A bottom-up model of skill learning. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 1037–1042). Mahwah, NJ: Erlbaum.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science, 25,* 203–244.

Sun, R., & Peterson, T. (1998a). Autonomous learning of sequential tasks: Experiments and analyses. *IEEE Transactions on Neural Networks, 9,* 1217–1234.

Sun, R., & Peterson, T. (1998b). Some experiments with a hybrid model for learning sequential decision making. *Information Sciences, 111,* 83–107.

Sun, R., & Zhang, X. (2004). Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research, 5,* 63–89.

Szymanski, K., & MacLeod, C. (1996). Manipulation of attention at study affects an explicit but not an implicit test of memory. *Consciousness and Cognition, 5,* 165–175.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

Vokey, J., & Brooks, L. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 328–344.

Vygotsky, L. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Warrington, E., & Weiskrantz, L. (1982). Amnesia: A disconnection syndrome? *Neuropsychologia, 20,* 233–248.

Watkins, C. (1989). *Learning with delayed rewards.* Unpublished doctoral thesis, Cambridge University, Cambridge, England.

Willingham, D. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105,* 558–584.

Willingham, D., Nissen, M., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1047–1060.

Wisniewski, E., & Medin, D. (1994). On the interaction of data and theory in concept learning. *Cognitive Science, 18,* 221–281.