References

Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence, 42*, 7-49.

Anderson, J. R., & Lebiére, C. (Eds.). (1998). *Atomic components of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Matessa, M., & Lebiére, C. (1997). ACT-R: A theory of higher-level cognition and its relation to visual attention. *Human-Computer Interaction, 12*(4), 439-462.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66-80.

Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson & C. Lebiére (Eds.), *The atomic components of thought* (pp. 167-200). Hillsdale, NJ: Erlbaum.

Gray, W. D. (in preparation). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and microworlds in basic and applied cognitive research. In R. Mahan, D. Serfaty, S. Kirschenbaum, M. McNeese, & L. Elliott (Eds.), *Scaled Worlds (working title)* . Hillsdale, NJ: Erlbaum.

Gray, W. D. (in press). The nature and processing of errors in interactive behavior. *Cognitive Science.*

Gray, W. D., & Altmann, E. M. (in press). Cognitive modeling and human-computer interaction. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* . New York: Taylor & Francis, Ltd.

Gray, W. D., & Boehm-Davis, D. A. (1999). Milliseconds Matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Manuscript submitted for publication.*

Gray, W. D., Schoelles, M. J., & Fu, W.-t. (1999). Modeling microstrategies in a continuous dynamic task. *Manuscript submitted for publication.*

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12*(4), 391-438.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.

Ritter, F. E., Baxter, G. D., & Jones, G. (2000). Cognitive models as users. *ACM Transactions on Computer-Human Interaction, in press.*

Learning of event sequences: An architectural approach

Dieter Wallach
Institut für Psychologie
Universitaet Basel
Bernoullistr. 16
CH-4056 Basel
dieter.wallach@unibas.ch

Christian Lebiere
Psychology Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
cl+@cmu.edu

Abstract

The paper summarized an integrated account of sequence learning, implemented in ACT-R, which can successfully account for experimental data from classic papers in the field (Lebiere & Wallach, in prep.). After providing evidence for the empirical adequacy of the approach, the results from two experiments are outlined that were conducted to test objections to a central assumption of the model proposed.

1. Introduction

Sequencing information and learning of event contingencies are fundamentally important processes without which adaptive behavior in dynamic environments would hardly be possible. From sequencing continuous speech to learning operating sequences of technical devices or acquiring the skill to play an instrument, learning of event sequence seems to be an essential capability of human cognition. It is therefore not surprising that the last decade witnessed a tremendous interest in sequence learning in cognitive psychology and the neurosciences.

In a typical sequence learning experiment subjects are exposed to visuospatial sequences in a compatible response mapping serial reaction time task. Participants are asked to react to each of a small number of events as quickly and accurately as possible with a discriminative response. Usually these events consist of asterisks that are presented in one of several horizontally aligned positions on a computer screen. Responses most frequently require participants to press keys that spatially match these positions. Unbeknownst to the subjects, the presented sequence of visuospatial signals follows a well-defined systematicity resulting in a continuous stream of structured events to which subjects have to respond. Sequence learning is said to have occurred when (a) subjects exposed to systematic event sequences show faster response latencies than those responding to random event sequences, or (b) response times of subjects increase significantly when systematic sequences are temporary switched to random sequences. The faster response times are interpreted as resulting from acquired knowledge about the event pattern that allows subjects to prepare their responses. Learning of the systematicity of event sequences is hence accessed *indirectly* by contrasting the response latencies to structured sequences with the reaction times to randomly presented events. A broad range of studies report to have demonstrated dissociations between sequence learning (as revealed indirectly through decreasing latencies in the serial reaction time task) and conscious awareness of the sequence (as assessed by *direct* post-task measures), providing evidence for *implicit* learning processes (Stadler & Frensch, 1998).

investigations in sequence learning have put forth a large variety of research questions that are documented in more than 100 scientific publications over the past ten years. However, despite the availability of an extraordinary wide body of experimental data on serial pattern learning, theory development to explain the phenomena empirically observed has been surprisingly scarce. With the exception of Cleeremans' work using simple recurrent networks (Cleeremans, 1993; Cleeremans & Jiménez, 1998) or the connectionist model proposed by Keele and Jennings (1992), computational approaches have been rare in the field. While the mentioned connectionist models had some success in qualitatively reconstructing selected subject data in serial pattern learning, they have yet to be applied to a wider range of experiments (and experimental conditions) to judge their theoretical scope. At present, a major limitation of these approaches seems to be that they "do not account for explicit learning or the difference between it and implicit learning" (Stadler & Roedinger, 1998, p. 126).

2. An integrative theory of sequence learning

Recently, Lebiere and Wallach (1998; in prep.) have proposed a comprehensive theory of sequence learning that is based on the ACT-R cognitive architecture (Anderson & Lebiere, 1998). ACT-R is a hybrid production system that distinguishes between a permanent procedural memory and a permanent declarative memory. Procedural knowledge is encoded in ACT-R using modular condition-action-rules (*productions*) to represent potential actions to be taken when certain conditions are met. So-called *chunks* are used to store factual knowledge in declarative memory. Chunks encode knowledge as structured, schema-like configurations of labeled slots with associated values that can be organized hierarchically. A representation of goals in a goal stack is utilized to control information processing whereby exactly one chunk is designated to be the active goal of the system. Knowledge represented symbolically by chunks and productions is associated with *subsymbolic* (i.e. real-valued) numerical quantities that control which productions are used and how they apply to chunks. These subsymbolic quantities reflect past statistics of use of the respective symbolic knowledge structures and are learned by Bayesian learning mechanisms derived from the *rational analysis* of cognition (Anderson, 1990). Generally, subsymbolic learning allows ACT-R to adapt to the statistical structure of an environment.

Central assumptions of the ACT-R theory of sequence learning

Declarative knowledge: A basic assumption of the theory is that the mappings between stimulus locations and response keys are encoded as chunks. Each chunk associates the respective stimulus location on the screen to the desired response key. These declarative representations essentially represent a straightforward explicit encoding of the experimental instructions informing the subjects of the stimulus-response mappings. When a stimulus is observed, the chunk representing the mapping between that stimulus location and the associated response key will be retrieved. Each retrieval results in the immediate reinforcement of that chunk through ACT-R's subsymbolic base-level learning mechanism that strengthens a chunk to reflect its frequency of use. These reinforcements will lead to higher activation levels for the mapping chunks, which result in faster response latencies since the time to retrieve a chunk is inversely proportional to (the exponential of) its activation. This speedup will occur independently of whether the stimulus sequence is systematic or random because it only depends upon the frequency of each retrieval.

A fundamental assumption of our model is the *persistence of (working) memory*. ACT-R states that the components of the current goal are sources of activation. If the new goal is to respond to a particular stimulus with a certain response, we assume that the previous stimulus (i.e. the central element of the previous goal) remains in the encoding of the new goal. This assumption has two important implications. First, since every goal contains both the previous stimulus and the next one, when that goal is popped and becomes a chunk in declarative memory, it contains a record of a small fragment of the sequence. The set of these chunks constitute the model's *explicit* knowledge of the sequence. The second implication is that when the chunk encoding the mapping between the current stimulus and the proper response key is retrieved, both the current stimulus and the previous one are components of the goal and thus sources of activation. The co-occurrence between previous stimulus (as a source of activation) and current stimulus (as a component of the mapping chunk being retrieved) is automatically learned by ACT-R in the association strengths between source stimulus and mapping chunk and facilitates further processing. The subsymbolic strengths of associations constitute the model's *implicit* knowledge of the sequence.

Procedural knowledge: Table 1 describes the procedural knowledge that uses these declarative encodings, both implicit and explicit, to perform the task. The basic goal, as expressed in the instructions to the sequence learning experiments, is to map the location of a screen stimulus to a key and press that key as a response. Production *Input* checks that no stimulus has been encoded yet and checks if one is present and if so encodes it and places its location in the current goal. Production *Map-Location* then retrieves the chunk from declarative memory that maps this location to the proper response key, and places the key in the goal. Production *Type-Key* then types the key and pops the goal. Before a stimulus has appeared, production *Guess* attempts to retrieve a chunk that holds a piece of the sequence starting with the current context, and if successful uses that chunk to anticipate which stimulus will appear and which key to press. Once the stimulus appears, if the anticipation was correct then the key can be typed directly without the need for mapping location to key. Otherwise, the production *Bad-Guess* withdraws the prepared response, which requires mapping the actual stimulus location to a different key.

Input IF the goal is to respond to a stimulus and no stimulus has been encoded yet THEN check if a *stimulus* is present and encode it	Guess IF the goal is to respond to a stimulus given a previous *context* And there is no stimulus present and no guess has been made And *context* was in the past most often followed by *stimulus* responded to by *key* THEN guess that the next stimulus will be *stimulus* and get ready to respond with *key*
Map-Location IF the goal is to respond to a stimulus and no key has been prepared to respond to *stimulus* and *key* is associated to *stimulus* THEN get ready to respond with *key*	
Type-Key IF the goal is to respond to a stimulus and a response with *key* has been prepared THEN type *key* and pop the goal	Bad-Guess IF the goal is to respond to a stimulus And a *guess* was made that is different from the encoded *stimulus* THEN withdraw the prepared response with *key*

Table 1: Productions to perform the mapping of a stimulus to a response key

3. Comparing model predictions and empirical data

The ACT-R theory of sequence learning precisely specifies mechanisms and memory structures on a level that is sufficient to implement a computational model that generates qualitative and quantitative predictions that can be compared to empirical data. To explore the integrative scope of the theory we have modeled experiments that were conducted by researchers with opposite theoretical orientations and which are regarded as providing the most direct evidence for their respective positions (Lebiere & Wallach, 1998; in prep.). The experimental conditions modeled vary widely with regard to the type of sequence used (unique, hybrid), the length of the sequence, the distribution of systematic vs. random blocks, the number and length of blocks, the response-to-stimulus interval used and whether the serial reaction task was presented as a single task or in combination with a secondary task[1] ("tone counting"). The validation of the model was not restricted to comparing model-generated and empirical data on a single dimension, but comprised a comparison of latencies, learning trajectories, errors, stimulus anticipations, individual differences as well as the structure of acquired chunks. We claim to have successfully modeled experiments from the following studies:

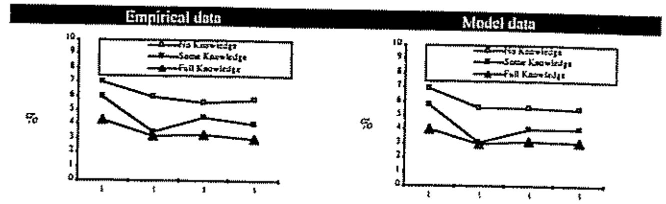
- *Willingham, Nissen, & Bullemer (1989)* investigated the temporal dependence of declarative and procedural memory and report that the "development of knowledge in one system seems not to depend on the other" (p. 1047).
- *Perruchet & Amorim (1992)* claim to have provided evidence "that undermine the most direct experimental support for the widespread view that conscious knowledge and performance tap two independent knowledge basis in normal subjects (see Willingham et al. (1989))" (p. 785).
- *Curran & Keele (1993)* investigated the role of attention for sequence learning and report results that "suggest that attentional and nonattentional learning operate independently, in parallel, do not share information, and represent sequential information in qualitative different ways" (p. 189).

Although a comprehensive presentation of all modeling results is beyond the scope of this paper, Figure 1 gives an impression of the precision with which the model reconstructs the data from the three seminal studies (see Lebiere & Wallach, in prep.).

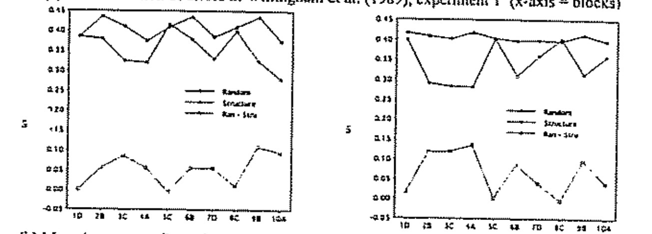
The graphs (a-c) in figure 1 show that the model outlined in previous section provides good fits to the empirical data, both on a qualitative and on a quantitative level. This is especially remarkable since the basic model was not fit individually to each experiment, but provides a truly integrative account of the experiments modeled.

Since the model predictions seem to be largely satisfying in their correspondence to empirical data, the question arises how to challenge the model in a next step. Simon and Wallach (1999) have argued that the scientific value of a cognitive model should not only be judge by looking at the goodness of fit, but also by exploring how the model gives rise to new, theoretically motivated experiments. The next section outlines the results of two experiments that were conducted to test a central assumption of the model.

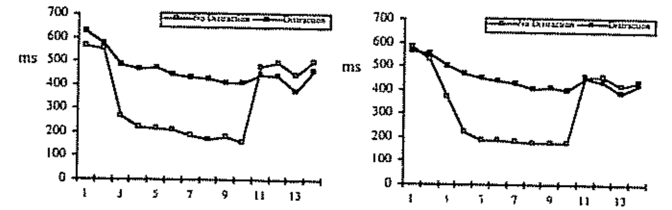
[1] Note that productions necessary for modeling dual task situations are not given in Table 1. For details on these productions see Lebiere & Wallach (in prep.).



(a) Reconstruction of errors in Willingham et al. (1989), experiment 1[2] (x-axis = blocks)



(b) Mean latencies in Perruchet & Amorim (1992), experiment 1 (x-axis = serial item position)



(c) Mean reaction time in Curran &Keele (1993), experiment 2[3] (x-axis = blocks)

Figure 1: Comparison of empirical and model data

[2] It should be noted that the ACT-R model uses one parameter *(amount of source activation, W)* to account for subjects that are fully, partially or not aware of the sequence (see Figure 1a).

[3] *Distraction* in Fig. 1/c refers to a condition in which subjects had to respond to stimuli in addition to a secondary task in blocks 3-10, while the *no distraction* group was not exposed to a secondary task in these blocks. In block 11-14 both groups worked in a dual-tasking situation.

4. A challenge for the ACT-R view on sequence learning

A basic assumption of the ACT-R theory of sequence learning is that subjects acquire (declarative) *Stimulus-Stimulus*-fragments[4] (S-S learning) of the sequence that are associated by subsymbolic learning. This assumption of an S-S type of learning has recently been challenged by authors who propose that what subjects learn in a sequence learning task can be described more appropriately as building up relations between consecutive *reactions* (R-R-learning): "The critical information on which sequence learning is based is contained in the sequence of responses rather than in the stimulus sequence" (Nattkemper & Prinz,1997, p. 110).

Empirical evidence on the question of whether sequence learning is a type of S-S learning as opposed to R-R learning is rather inconsistent. Cohen, Ivry, and Keele (1990) conclude from an experiment where subjects were required to switch from using three fingers to respond to using one finger, that the knowledge of the sequence pattern is independent of the particular motor system. Howard, Mutter, and Howard (1992) found evidence for S-S learning by showing that subjects who only observed the stimulus sequence in the first three blocks of an experiment showed no difference in a fourth block to a group that had to respond to the sequence in all four blocks. However, Hoffmann and Koch (1997), Nattkemper and Prinz (1997) and Zießler (1994), using variants of sequence learning tasks, stress the impact of motor responses on serial pattern learning.

To investigate the question of what kind of contingencies are learned in sequence learning, we conducted an experiment that was targeted to explore whether sequence knowledge can be transferred across different response effectors, i.e. from using the index and middle finger of each hand to respond by pressing keys to using mouse clicks to respond by pressing on menu items. If a significant transfer to using the mouse for responses after being trained with pressing keys with fingers could be empirically substantiated, this would be incompatible with a view of sequence learning being a type of pure R-R learning, but would be in line with S-S learning. Before being able to investigate this research question, we first had to ensure that the typical sequence learning pattern would generalize to using the mouse as an input device. In experiment 1 we therefore compared the effects of using the mouse in contrast to using key presses.

Experiment 1:

50 subjects (25 male, 25 female) participated in the experiment for course credit. Subjects were assigned to one of two experimental groups: the *key-group* (n=25) had to respond to signals on four horizontally aligned locations on the screen by using the index and the middle finger of each hand on a computer keyboard. The *mouse-group* (n=25) responded by using the mouse to click on four physically separated locations on the computer screen that were presented as menu items. Both groups were exposed to the same hybrid 6-trial sequence "*A-D-C-A-C-B*", used by Curran and Keele (1993). The sequence was repeated 20 times per block, for a total of 12 blocks. A response-to-stimulus interval of 250 ms was used. Subjects were tested individually in single sessions and were not informed about the systematicity of the pattern. Figure 2 shows the latencies (presented as means of median reaction times) observed in experiment 1.

[4] More precisely, the model learns S-SR fragments because the response key is included in the goal chunk and is thus used to prevent having to map the predicted stimulus.

Results: Contrasting the latencies of random block R3 (the first random block after four systematic blocks) with systematic block S5 shows clear evidence for sequence learning in the key-group ($t(24)=13.37$, $p<.00001$) as well as in mouse-group ($t(24)=7.68$, $p<.00001$). Comparing S5 with R4 in the key groups shows again significant faster responses in the systematic block in the key-group ($t(24)=9.78$, $p<.00001$) as well as in the mouse-group ($t(24)=6.85$, $p<.00001$). The same statistically significant pattern holds for the comparison of R5/S6 and S6/R6 in both groups. On a qualitative level, the latency graphs of the key-group and the mouse-group show very similar trajectories — however, due to the more complex motor action, latencies in the mouse-group are generally significantly longer than in the key-group ($p<.00001$).

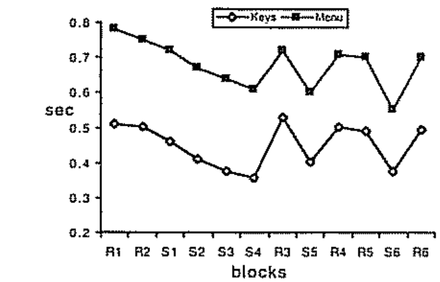


Figure 2: Latencies in experiment 1 (x-axis shows succession of random [R] vs. systematic [S] blocks)

Experiment 1 has shown that the typical pattern found in sequence learning experiments when using keys to respond can also be found when mouse responses — and thus a much more complex motor reaction — are required. We have hence provided evidence that sequence learning can be generalized to situations where the mouse is used as an input device. In experiment 2 we can now test whether there is a de facto transfer of sequence knowledge when switching the input modalities as predicted by ACT-R's S-S view of serial pattern learning.

Experiment 2:

25 subjects (9 male, 16 female) participated in the experiment for course credit. All subjects were exposed to the same hybrid sequence as in experiment 1, using a response-to-stimulus interval of 250ms. Again, the sequence was repeated 20 times per block, with 12 successive blocks in total. All subjects responded with the index and middle finger of each hand to the succession of the sequence in the first 6 blocks and switched to the mouse to respond for the last 6 blocks. Figure 3 shows the latencies (presented as means of median reaction times) observed in experiment 2.

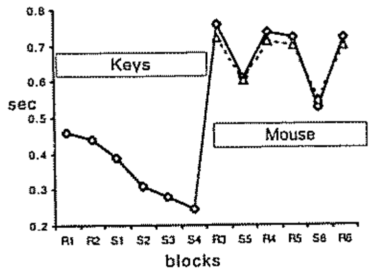


Figure 3: Latencies in experiment 2 (x-axis shows succession of random [R] vs. systematic [S] blocks)

Results: Contrasting the latencies of random block R3 with systematic block S5 shows clear evidence for a transfer of sequence knowledge despite a change of the response type ($t(24)=8.75$, $p<.00001$), the same holds for a comparison of the latencies of block S5 with R4 ($t(24)=7.02$, $p<.00001$). Comparing S6 with R5 ($t(24)=10.9$, $p<.00001$) and S6 with R6 ($t(24)=10.37$, $p<.00001$) provides further evidence for a positive transfer from keys to the mouse. The amount of transfer becomes especially evident when comparing the performance of subjects in experiment 2 to latencies of the mouse-group in experiment 1 who responded with the mouse in all 12 blocks (dashed line in Fig. 3): both groups of subjects seem to perform at essentially the same level.

The results of experiment 2 are clearly incompatible with a view of sequence learning as the association of (effector-specific) responses and provide direct support for the S-S type of learning proposed in the ACT-R theory of sequence learning. In addition to the latency data shown in Figure 3, fine-grained analyses of anticipatory mouse movements and error data further strengthen the view of sequence learning as the acquisition of S-S contingencies. From a rational analysis point of view, we have to agree to Goschke (1998) who concludes: "It appears prima facie to be important for adaptive systems to acquire knowledge about event contingencies independently of specific motor responses" (p. 415).

4. Conclusions

In this paper we have outlined a detailed ACT-R theory of the mechanisms and representational structures underlying learning in serial reaction tasks. We have provided evidence for the validity of our approach by modeling classic papers in the field and by experimentally investigating alternative theoretical views of sequence learning in the form of R-R learning. Our next challenge in evaluating the scope of the ACT-R theory of sequence learning will be the modeling of the results found in the experiments reported in this paper.

References

Anderson, J.R. (1990). *The adaptive character of thought.* Hillsdale: LEA.
Anderson, J.R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah: LEA.
Cleeremans, A. (1993). *Mechanisms of implicit learning.* Cambridge, MA: MIT Press.

Cleeremans, A. & Jiménez, L. (1998). Implicit sequence learning: The truth is in the details. In M.A. Stadler & P.A. Frensch (Eds.). *Handbook of implicit learning* (p. 323-364). Thousand Oaks, CA: Sage.
Cohen, A., Ivry, R., & Keele, S.W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 17-30.
Curran, T. & Keele, S.W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 189-202.
Goschke, T. (1998). Implicit learning of perceptual and motor sequences. In M.A. Stadler & P.A. Frensch (Eds.). *Handbook of implicit learning* (p. 401-444). Thousand Oaks, CA: Sage.
Hoffmann, J. & Koch, I. (1997). Stimulus-response compatibility and sequential learning in theserial reaction task. *Psychological Research, 60,* 87-97.
Howard, J.H., Mutter, S.A., & Howard, D.V. (1992). Serial pattern learning by event observation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 1029-1039.
Keele, S.W. & Jennings, P.J. (1992). Attention in the representation of sequence: Experiment and Theory. *Human Movement Science, 11,* 125-138.
Lebiere, C., & Wallach, D. (1998). Eine deklarative Theorie des Sequenzlernens. Paper presented at the *Forty First Conference of the German Psychological Society*, Dresden, Germany.
Lebiere, C., & Wallach, D. (in prep.). Implicit does not imply procedural: A declarative theory of sequence learning. Manuscript in preparation.
Nattkemper, D. & Prinz, W. (1997). Stimulus and response anticipation in a serial reaction task. *Psychological Research, 60,* 98-112.
Perruchet, P. & Amorim, M.-A. (1992).Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 785-800.
Simon, H.A. & Wallach, D. (1999). Cognitive modeling in perspective, *Kognitionswissenschaft, 8, 1-5.*
Stadler, M.A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 318-327.
Stadler, M.A. & Frensch, P.A. (1998). *Handbook of implicit learning* (p. 105-132). Thousand Oaks, CA: Sage.
Stadler, M.A. & Roediger, H.L. (1998). The question of awareness in research on implicit learning. . In M.A. Stadler & P.A. Frensch (Eds.). *Handbook of implicit learning* (p. 105-132). Thousand Oaks, CA: Sage.
Willingham, D.B., Nissen, M.J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 1047-1060.
Zießler, M. (1994). The impact of motor responses on serial learning. *Psychological Research, 57,* 30-41.