# Attention and Implicit Learning

Edited by
Luis Jiménez

# Attention and Implicit Learning

# Advances in Consciousness Research

Advances in Consciousness Research provides a forum for scholars from different scientific disciplines and fields of knowledge who study consciousness in its multifaceted aspects. Thus the Series will include (but not be limited to) the various areas of cognitive science, including cognitive psychology, linguistics, brain science and philosophy. The orientation of the Series is toward developing new interdisciplinary and integrative approaches for the investigation, description and theory of consciousness, as well as the practical consequences of this research for the individual and society.

Series B: Research in progress. Experimental, descriptive and clinical research in consciousness.

**Volume 48**

Attention and Implicit Learning
Edited by Luis Jiménez

# Attention and Implicit Learning

*Edited by*

Luis Jiménez
University of Santiago de Compostela, Spain

**John Benjamins Publishing Company**
Amsterdam / Philadelphia

**Library of Congress Cataloging-in-Publication Data**

Attention and Implicit Learning / edited by Luis Jiménez.
p. cm. (Advances in Consciousness Research, ISSN 1381–589X ; v. 48)
Includes bibliographical references and indexes.
1. Implicit learning. 2. Attention. I. Jiménez, Luis, 1964- II. Series.

BF319.5.I45 A88 2002
153.1'5-dc21 2002034215
ISBN 90 272 5175 4 (Eur.) / 1 58811 335 3 (US) (Hb; alk. paper)
ISBN 90 272 5176 2 (Eur.) / 1 58811 336 1 (US) (Hb; alk. paper)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

# Table of contents

**I. The cognitive debate**

**II. Neuroscientific and computational approaches**

## III. Reciprocal influences: Implicit learning, attention, and beyond

# Acknowledgement

The edition of this volume was supported by Grant XUGA 21106B98 from the Consellería de Educación e Ordenación Universitaria da Xunta de Galicia (Spain) and by the DGES Grant PB97-0525 from the Ministerio de Educación y Cultura (Spain). I thank Maxim Stamenov for the invitation to prepare this volume. I also wish to thank the contributors to this volume for having accepted to take part in the project, and the editors at the Amsterdam office of John Benjamins, for their continuous support. Finally, I also wish to acknowledge the help of Gustavo Vázquez, who took care of some of the most time-consuming tasks concerning the production of the Index, and who did it quite efficiently and thoroughly.

# Contributors

**F. Gregory Ashby**
University of California, Santa Barbara
(Santa Barbara, USA)
email: ashby@psych.ucsb.edu

**Alan S. Brown**
Southern Methodist University
(Dallas, USA)
email: abrown@mail.smu.edu.

**Michael Casale**
University of California, Santa Barbara,
(Santa Barbara, USA)
email: casale@psych.ucsb.edu

**Marvin Chun**
Vanderbilt University
(Nashville, USA)
email: marvin.chun@vanderbilt.edu

**Axel Cleeremans**
Université Libre de Bruxelles
(Brussels, Belgium)
email: axcleer@ulb.ac.be

**Arnaud Destrebecqz**
Université Libre de Bruxelles
(Brussels, Belgium)
email: adestre@ulb.ac.be

**Peter F. Dominey**
CNRS
(Bron, France)
email: dominey@nimbus.isc.cnrs.fr

**Peter A. Frensch**
Humboldt University at Berlin
(Berlin, Germany)
peter.frensch@psychologie.hu-berlin.de

**Hilde Haider**
University of Köln
(Köln, Germany)
email: hilde.haider@uni-koeln.de

**Eliot Hazeltine**
NASA - Ames Research Center.
(Moffet Field, CA, USA)
email: ehazeltine@mail.arc.nasa.gov

**Richard B. Ivry**
University of California, Berkeley
(Berkeley, CA USA)
email: ivrylab@socrates.berkeley.edu

**Yuhong Jiang**
MIT
(Cambridge, MA, USA)
email: yuong@mit.edu

**Luis Jiménez**
University of Santiago
(Santiago, Spain)
email: jimenez@usc.es

**Tony Lambert**
University of Auckland
(Auckland, New Zealand)
email: t.lambert@auckland.ac.nz

**Christian Lebiere**
Carnegie Mellon University
(Pittsburg, USA)
email: el@andrew.cmu.edu

**Neil Mulligan**
Southern Methodist University
(Dallas, USA)
email: mulligan@post.cis.smu.edu

**Uwe Neugebauer**
University of Köln
(Köln, Germany)
email: Uwe.Neugebauer@uni-koeln.de.

**Dennis Rünger**
Humboldt University at Berlin
(Berlin, Germany)
email: dennis@brainpan.info

**David Shanks**
University College London
(London, UK)
email: d.shanks@ucl.ac.uk

**Sabine Voigt**
Humboldt University at Berlin
(Berlin, Germany)
email: sabine.voigt.1@rz.hu-berlin.de

**Dieter Wallach**
University of Applied Sciences
(Kaiserslautern, Germany)
wallach@informatik.fh-kl.de

**Jana Werg**
Humboldt University at Berlin
(Berlin, Germany)
email: janawerg@gmx.de

# Introduction

## Attention to implicit learning

Luis Jiménez
University of Santiago

During the last few years, the empirical study of implicit cognition has become increasingly integrated with the conceptual and philosophical debates concerning the nature and functions of consciousness (Baars 1997; Flanagan 1997; French & Cleeremans 2002). This progressive integration can be seen as part of an overall effort designed to implement what Flanagan (1992) called the *natural method* for studying consciousness, i.e. a triangulated approach that aims to take advantage of the combined powers of phenomenology, psychology and neuroscience, to provide a naturalistic framework within which to explain consciousness. In this context, the study of implicit cognition plays an important role in the search for the functional correlates of consciousness (Atkinson, Thomas, & Cleeremans 2000), and is specially needed to make functional sense of the results obtained through the search for its neural correlates (Block 1996; Chalmers 1998). This volume is intended to contribute toward this goal, by bringing together a selection of the current research on implicit learning and, specifically, by reviewing the current knowledge about the functional relation that exists between implicit learning and attention, about its neural correlates, and about the implications that this information may have on the conceptual debate about the nature and functions of consciousness.

This volume on *Attention and Implicit Learning* provides a comprehensive overview of the research conducted in this area. It is conceived as a multidisciplinary forum of discussion on the question of whether implicit learning – that is often defined as the learning that takes place without intention and awareness – may also be depicted as a process that runs independently of attention or whether, on the contrary, it may rely on the same type of attentional mediation that is often considered to govern explicit learning processes. The answer to this question will obviously depend on the detailed meaning conveyed by these expressions and, hence, one of the first conclusions to be reached from the present debate is that there is not a quick answer to this overall question. However, after going through all these contributions in detail, I expect that the reader may end up with the impression that good answers are beginning to arise from the joint effort of researchers addressing these issues simultaneously from different standpoints, and becoming increasingly aware of the advantages provided by the sharing of their perspectives.

## An overview of the volume

This volume consists of eleven chapters that address this key question from a blend of cognitive, neuroscientific, and computational approaches. Chapters 1 and 2 set the stage for the cognitive debate, presenting it from a functional and empirical perspective. In Chapter 1, Shanks provides a sceptical overview of the claim that implicit learning can proceed without making extensive demands on attentional resources, and independently from awareness. He presents new results that are taken to demonstrate that, under dual-task conditions, not only the expression, but also the acquisition of sequence learning is impaired. In addition, he also shows that the knowledge expressed through the indirect measures of performance is closely associated with that manifested through direct measures such as those taken from recognition and free generation tasks, which are usually considered to rely on explicit knowledge.

Consistently with this warning against the potential contamination of the measures of implicit learning with explicit influences, I point out in Chapter 2 that most of the alleged measures of implicit learning may be sensitive to some explicit learning as well, and thus that most of the effects of attention on implicit learning measures might inadvertently have a bearing on the effects of attention on this residual sensitivity to explicit knowledge. Hence, I propose to use complex, probabilistic structures as a way to circumvent this problem, and review some results that have shown that when the structure is complex, implicit sequence learning does not appear to depend on the amount of attentional resources available. These results also indicate that learning in these conditions requires participants to pay selective attention to the relevant dimensions and, therefore, I conclude that this implicit learning might be taken to be an automatic side-effect of processing, which would associate all the features of the environment that undergo enough processing, but only those features that are being selectively attended, as determined by their relevance for the orienting task.

After having set the stage for this discussion in its cognitive terms, Chapters 3 to 5 address similar issues in neuroscientific terms, by reviewing a set of neurophysiological and neuropsychological approaches to the role of attention in implicit learning. In Chapter 3, Hazeltine and Ivry review the neural structures that are believed to support implicit sequence learning, and separate two groups of structures which have been shown to respond differently to attentional manipulations. On one hand, the supplementary motor area, along with a number of areas within the parietal lobe and the basal ganglia, are taken to constitute a non-attentional, implicit learning system, that could encode a series of responses even under conditions of distraction. On the other hand, a second learning system comprising anterior regions, including the prefrontal and premotor cortex, is posited to produce implicit learning effects only when full attention can be devoted without interruption to the main task.

Ashby and Casale undertake a similar localizing task in Chapter 4, for the paradigm of category learning. They start by recognizing the use of a different notion of implicit learning in these tasks, in which participants are typically instructed to learn, and in which they receive feedback about their categorization performance. However, they still consider that learning under these circumstances can be taken to be implicit if participants gain no conscious access to the system that mediates such learning, and if they remain unable to verbalize the underlying knowledge. Within this paradigm, Ashby and Casale report on different sources of behavioral and neuropsychological evidence, and propose to distinguish between three different learning subsystems. First, a rule-based, explicit categorization system is taken to be mediated by the prefrontal lobe and the anterior cingulate, and would be specially useful to learn about single-dimension categorization tasks. Second, an implicit, procedural memory system is characterized as a reward-mediated system that would be modulated by the release of dopamine from the *substantia nigra*, and in which the caudate nucleus within the basal ganglia would play a major role. This learning is depicted as resistant to distraction, and would affect performance specially in what the authors call "information-integration tasks". Finally, a third implicit learning mechanism is described to rely on a perceptual representation memory system, and is taken to play a major role in prototype distortion tasks. Unlike the procedural learning system, this perceptual learning system is described as being based on a form of long-term potentiation, and it could be implemented through a non-competitive, Hebbian learning algorithm.

After this brief journey through the categorization tasks, Chapter 5 takes us back to sequence learning. In this chapter, Dominey analyzes the effects of temporal, serial, and abstract structure in the standard serial, reaction-time (SRT) task, and he reports on behavioral results, neuropsychological evidence, and computational simulations that allow him to distinguish between two different learning systems in terms of their attentional requirements. The implicit learning system is taken to be independent from attention and is computationally described as a temporal recurrent network that learns to predict the next output based on its current encoding of the temporal context. Just as the procedural learning system depicted by Ashby and Casale, this implicit learning system is described as performing a reinforcement algorithm that would involve the connections between the caudate nucleus and the prefrontal cortex, and that would be modulated by the dopaminergic input coming from the *substantia nigra*. On the other hand, the abstract learning subsystem resembles the rule-based learning system described by Ashby and Casale, in that it also depends on the integrity of attentional resources, and relies on structures that are closely related to those involved in language processing.

The computational analysis undertaken by Dominey is continued and deepened throughout Chapters 6 and 7, that put more emphasis on the functional implications of the models, and less on the neural implementation details. In Chapter 6, Destrebecqz and Cleeremans investigate the temporal dynamics of sequence learning, showing that an increase in the response-to-stimulus interval (RSI) increases explicit sequence learning. To account for these results, the authors use a new model based on the Simple Recurrent Network (SRN). Within the framework provided by this recurrent structure, the new model incorporates the cascade algorithm to capture the time course of processing during a single trial, and combines the prediction responses provided by the SRN with the identification responses produced by an auto-associator. The temporal competition established between these two subsystems provides a straightforward account for the fact that long RSIs allow for the development of higher-quality, and potentially conscious representations within the SRN whereas, on the contrary, short SRIs lend a comparatively more important role to the auto-associator, and hence reduce the role of the recurrent structure to that of providing a certain amount of implicit facilitation.

To close this computational section, Wallach and Lebiere present the ACT-R symbolic architecture and report on a series of simulations through which they illustrate their main proposal: that explicit learning can be identified with the learning of declarative chunks, whereas implicit learning could be based on the ACT-R's subsymbolic learning algorithms. Specifically, they present simulations of both the process control task and the sequence learning paradigm, and show that a mechanism of blending between chunks can account for the production of new control responses to non-trained contexts, whereas the subsymbolic processes of chunk activation can account for the effects of different RSIs as observed, for instance, in the results of Destrebecqz and Cleeremans.

Throughout the chapters overviewed so far, the reader may thus find various cognitive, neuroscientific, and computational perspectives on the questions of how implicit learning works and how attention may affect these effects of implicit learning. However, the volume also includes two chapters concerned with the complementary question of whether and how implicit learning affects the dynamics of attention. Chapter 8, by Lambert, characterizes this issue by borrowing the Jamesian notion of "derived attention", and explores the role of implicit learning in attention through the *spatial cueing paradigm*. In this paradigm, different cues are contingently related with the location of a target, and the propensity of these cues to capture attention is observed to change by virtue of the learned associations. Jiang and Chun describe a similar strategy in Chapter 9, in which they introduce the *contextual cueing paradigm*, and investigate the reciprocal influences that hold between attention and implicit learning. In this paradigm, a visual search task is manipulated by including distractors that co-vary with the spatial location of the target, and the authors demonstrate that participants do implicitly learn about these spatial correlations, but only to the extent that the distractors cannot be efficiently ignored – i.e., only when the target cannot be preattentively segregated from the distractor set. Thus, the results of these experiments do strongly indicate that implicit learning and attention are related in complex and reciprocal ways, so that selective attention does modulate what can be learned implicitly, and implicit learning shapes the deployment of attention.

After all this discussion concerning the relations between attention and implicit learning, two closing chapters go beyond the limits of this topic, by fixing its relations with two neighboring topics, such as those of implicit memory and explicit learning. In Chapter 10, Mulligan and Brown provide an overarching review of the effects of attention in implicit memory, and they conclude that attentional manipulations can affect both conceptual and perceptual implicit memory. Finally, in Chapter 11, Frensch, Haider, Rünger, Neugebauer, Voigt, and Werg, analyze the relation between implicit learning and consciousness, and propose an integrated view according to which consciousness of an environmental regularity experienced in the course of an incidental learning task can be taken to be a consequence of implicit learning. They suggest that the effects of implicit learning may give place to a number of unexpected events –

for instance, learners may notice an increase in the efficiency of their responding – and that these unexpected effects may be instrumental in triggering an intentional search, that could eventually lead to the conscious discovery of the underlying regularities. This framework is consistent with claims, such as those made by Lambert or by Jiang and Chun, in the sense that the effects of implicit learning may drive attentional strategies that, in turn, may further modulate the effects of learning, either implicit or explicit. From this perspective, therefore, implicit and explicit learning would no longer be conceived as the product of two completely independent learning modules, but rather as different results of a single implicit *learning* system, that would be continuously modulated by an explicit *reasoning* sytem, that would be in charge of managing the attentional functions according to the learner's current goals.

Overall, the view of implicit learning that arises from these chapters is surely not that of a mysterious faculty of learning without even knowing it, or that of an experimental curiosity that arises exclusively under heavily controlled experimental settings. Rather, implicit learning is taken to be an elementary ability of the cognitive systems to extract the structure existing in the environment, regardless of their intention to do so. Implicit learning, thus, may produce pervasive effects on the whole dynamics of cognition, and may continuously shape not only our behavior, but also our representations of the world, our cognitive processes, and possibly our conscious experience as well. In fact, the right question to be raised from this viewpoint is not whether implicit learning has been demonstrated to run without awareness beyond any reasonable doubt, but rather whether "explicit" learning mechanisms exist, and whether they may be identified with the action of some other learning processes, different from those subserving implicit learning. Several years ago, Barsalou (1995: p. 412) advanced a negative answer to this question, by claiming that there is no such thing as a goal-driven learning mechanism that could be deliberately switched on and off: "people do not have the ability to turn the storage of information on and off depending on its relevance to their goals". Hence, if this proposal is sound, then we should probably accept that implicit and explicit learning effects do not rely on different storage mechanisms, but merely result from the effects of different strategic processing operations on the same basic learning mechanisms. This may allow for a deep change in focus, by which both implicit and explicit learning might be merely called "learning" (Cleeremans 1997), but it would not amount to the complete sidestepping of the issue of implicit learning. Indeed, the analysis of the effects of attention on different learning paradigms could still tell us something very important about which kind of regularities our cognitive systems are prepared to capture immediately, and which other contingencies can be grasped exclusively by relying on a series of strategic, resource-demanding, and conscious recoding operations. Of course, this may not be as mysterious as the existence of a learning ability that may run completely independent from awareness, but this is not to be expected if consciousness really does have a function in the overall cognitive economy. Hence, despite the change in focus that may be perceived throughout these chapters, the question about the role of attention in learning still stands as an essential topic that is worth pursuing, and which should be of interest to anyone concerned with improving our current understanding of the dynamics of cognition, and of the overall role that consciousness could play in this dynamics.

## References

Atkinson, A. P., Thomas, M. S. C., & Cleeremans, A. (2000). Consciousness: Mapping the theoretical landscape. *Trends in Cognitive Sciences, 4*, 372–382.

Baars, B. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. New York: Oxford University Press.

Barsalou, L. (1995). Storage side effects: Studying processing to understand learning. In A. Ram and D. B. Leake (Eds.), *Goal-Driven Learning* (pp. 407–419). Cambridge, MA: MIT Press.

Block, N. (1996). How can we find the neural correlates of consciousness? *Trends in Neuroscience, 19*, 456–459.

Chalmers, D. (1998). On the search for the neural correlate of consciousness. In S. Hameroff, A. Kaszniak and A. Scott (Eds.), *Toward a Science of Consciousness II: The Tucson Discussions and Debates* (pp. 219–229). Cambridge, MA: MIT Press.

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

Flanagan, O. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Flanagan, O. (1997). Prospects for a unified theory of consciousness. In N. Block, O. Flanagan, and G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.

French, R. & Cleeremans, A. (Eds.) (2002). *Implicit Learning and Consciousness: An Empirical, Computational and Philosophical Consensus in the Making?* Hove, UK: Psychology Press.

# Part I

# The cognitive debate

# Attention and awareness in "implicit" sequence learning

David R. Shanks*
University College London

In this chapter I examine two ideas about "implicit" learning, that it can proceed normally without making demands on central attentional resources and that it can proceed independently of, and be dissociated from, awareness. Traditionally, implicit learning has been defined as learning which takes place incidentally, in the absence of deliberate hypothesis-testing strategies, and which yields a knowledge base that is inaccessible to consciousness. From this sort of conception the two aforementioned claims follow fairly directly.

The first claim – that implicit learning can occur with minimal demands on attention – has recently been defended by a number of authors (e.g., Frensch 1998; Frensch, Lin, & Buchner 1998; Frensch, Wenke, & Rünger 1999; Hayes & Broadbent 1988; Heuer & Schmidtke 1996; Jiménez & Méndez 1999; Schmidtke & Heuer 1997; Stadler 1995). The idea is that implicit learning, unlike explicit learning, can proceed normally in the presence of concurrent resource-demanding tasks and therefore qualifies as an automatic process. The present chapter scrutinizes some of the key evidence supportive of this conception. The second claim, concerning awareness, is related but tends to focus on apparent demonstrations of learning accompanied by chance-level performance on direct tests of awareness such as recognition and generation tests. The first part of this chapter concentrates on attention and the second on awareness.

## 1. Attention and implicit sequence learning

At first glance the “attentional” claim about implicit learning faces a number of problems. A large literature has cast doubt on the general notion of automaticity (e.g., Cheng 1985; Kahneman & Chajczyk 1983; McCann, Remington, & Van Selst 2000; Styles 1997). Genuinely automatic cognitive processes which make no demands on central capacity have been very hard to find. For instance, on the basis of the Stroop effect, word reading is often assumed to be a prototypical automatic process, but Kahneman and Chajczyk (1983) presented evidence that Stroop interference is diluted by the presence of additional words in the display and concluded that even word reading is therefore not fully automatic.

Moreover, several implicit learning studies appear to have shown that the addition of a secondary task has an adverse effect on learning. For example, consider the sequential reaction time (SRT) task which is the focus of the present chapter. This task is especially well-suited for the study of the role of attention in implicit learning as it involves quite low-level perceptual-motor skill learning and can readily be combined with a variety of concurrent attention-demanding tasks. In this task, a target such as a dot appears in one of several possible locations on a computer display and the participant presses as fast as possible a response key assigned to that location. Instead of appearing at random across a series of trials, however, the target follows a predictable sequence of locations and the issue is whether participants learn (implicitly) this sequence. Learning is measured chronometrically by changing the sequence after a number of training blocks; an increase in RTs on this transfer sequence is evidence that participants have learned something about the training sequence and were using their knowledge to anticipate the target location on each trial, thus achieving rapid RTs. Using this task, Cohen, Ivry, and Keele (1990: Exp. 4) obtained evidence suggesting that a concurrent tone-counting task reduced sequence-learning. That is, switching from the training sequence to the transfer sequence (which was in fact a random sequence) had a small effect on RTs in a dual-task group, whereas in a single-task group the switch led to a more substantial increase in RTs. This seems to imply that implicit sequence learning is attention-demanding, contrary to the proposal. The tone-counting task required attentional resources and left participants with insufficient capacity to learn the target sequence. Other studies have confirmed that the RT increase on transfer trials is smaller under dual-task than single-task conditions (e.g., Frensch & Miner 1994; Stadler 1995).

## 1.1 Serial reaction time studies with a tone-counting secondary task

In a striking study, Frensch (1998; Frensch et al. 1998) offered an alternative interpretation of these findings. In the experiments mentioned above, participants in the dual-task condition performed the tone-counting task both during the training blocks and during the transfer block. Thus, Frensch argued, it is possible that the results reflect a performance effect rather than a learning deficit. Participants may learn as much about the sequence under dual- as under single-task conditions, but may be less able to express that knowledge when tested with a concurrent task. This "suppression hypothesis" – the hypothesis that dual-task testing conditions adversely affect the expression of sequence knowledge – is supported by the following finding: Suppose participants are trained on a sequence under single- or dual-task conditions and are then tested under both single- and dual-task conditions. The suppression hypothesis predicts that the measure of sequence learning (the RT increase on the transfer block) will be lower on the dual-task than on the single-task test, regardless of training conditions, since the former but not the latter will suppress the expression of sequence knowledge. Experiments testing this prediction have been somewhat contradictory (see Curran & Keele 1993; Frensch et al. 1999), but to cut a long story short, there is now a fair amount of evidence in support of the prediction (though see below). For example, in participants trained under dual-task conditions, Frensch, Wenke, and Rünger (1999) obtained significantly lower transfer scores on a dual-task than on a single-task test.

The obvious way to avoid the difficulty created by suppression is to train some participants under dual-task conditions and others under single-task, and then test all participants under identical conditions (e.g., under single-task conditions). Frensch, Lin, and Buchner (1998, Exps. 1a & 1b) report a pair of experiments essentially of this sort the results of which indicate that a concurrent task during the training stage has no effect on sequence learning per se.

In Frensch et al.'s Experiment 1a, participants trained for 7 blocks of trials on a repeating sequence, with each block comprising 16 repetitions of a 9-location sequence. The sequence was ABCDEADFC for some participants and ABCDECFBE for others, where A-F refer to 6 screen locations (the assignment of A-F to the actual screen locations was varied across participants). On blocks 8 and 9 the structured sequence was replaced by a quasi-random sequence (in which the frequency of each location was the same as in the structured sequence), and then on blocks 10 and 11 the original sequence was reinstated. Frensch et al. computed the difference in mean RTs between quasi-random blocks 8 and 9 versus sequence blocks 7 and 10 and took this transfer score as their measure of sequence knowledge.

For all participants, the test blocks (7–11) were conducted under single-task conditions. The major independent variable was the presence of a secondary task during the training stage. This task, which has been used in many similar experiments (e.g., Cohen et al. 1990; Nissen & Bullemer 1987), involved presentation of a high- or low-pitched tone in the interval between the visual targets. Participants were required to count the number of high-pitched tones during each block and report the number at the end of the block. For some participants the training stage (blocks 1–7) was conducted mainly under single-task conditions whereas for others most of the training blocks included a secondary task. Specifically, for Group 2-DT/5-ST the first 2 blocks were run under dual-task (DT) conditions but the remaining 5 blocks were single-task (ST); for Group 4-DT/3-ST the first 4 blocks were run under dual-task conditions and the remaining 3 were single-task; and for Group 6-DT/1-ST the first 6 blocks were run under dual-task conditions and the remaining block was single-task. Hence the two extreme groups (2-DT/5-ST, 6-DT/1-ST) compare conditions of mainly single-task and mainly dual-task training.

The key question is whether this manipulation of training conditions affects sequence learning in circumstances where testing is conducted under identical (single-task) conditions. The results were clear: Transfer scores were very nearly identical (approx. 85 msec) in the 3 groups. Thus participants learned the sequence equally well regardless of the inclusion of a secondary task. In their Experiment 1b, Frensch et al. (1998) replicated this pattern but in a situation where sequence knowledge was now assessed under dual-task conditions. Here the learning effect was smaller, with transfer scores of about 55 msec, but again the scores did not vary as a function of how many training blocks included the secondary task. The fact that the scores were lower overall in this experiment supports the suppression hypothesis: The inclusion of a secondary task during the testing phase tends to reduce transfer scores.

Similar results have been reported by other researchers. Seger (1997) and Cleeremans and Jiménez (1998) found nearly identical transfer scores in participants trained under single- or dual-task conditions when they were tested under identical conditions. Also, some data reported by McDowall et al. (1995, Exp. 3) support the same conclusion. These authors trained one group of subjects for 5 blocks under single-task conditions and another group for 4 blocks under dual-task followed by a final block under single-task conditions. On block 5, the mean RT of the two groups was comparable. The absolute level of RTs is probably a poor measure of sequence knowledge, compared to the effect of transfer to a random sequence, but nevertheless these results are consistent with the view that sequence learning under single- and dual-task conditions does not differ.

Schvaneveldt and Gomez (1998, Exp. 3) found evidence consistent with Frensch's hypothesis, albeit with one important proviso. These authors used probabilistic rather than deterministic sequences, in which each trial had a 90% chance of being consistent with an underlying sequence and a 10% chance of being inconsistent. The difference in RTs to these probable and improbable stimuli provided a continuous measure of sequence knowledge. Schvaneveldt and Gomez obtained an RT difference of 51 msec at the end of the training stage in a single-task group and a difference of 56 msec in a group trained under dual-task conditions and then switched to single-task testing. Again, sequence learning (measured by RT) under single- and dual-task conditions did not differ noticeably. The proviso is that error rates (an error being an incorrect keypress) were higher in the single-task group at the end of the training stage than in the dual-to-single task group during the test stage. If we assume that better sequence knowledge generates more errors with this version of the SRT task (because a participant who knows the underlying sequence is more likely to incorrectly anticipate the "consistent" location on an inconsistent trial), then the error data suggest that sequence learning was after all somewhat better in the single-task group.

A study by Heuer and Schmidtke (1996, Exp. 1) which again used tone-counting as the secondary task did obtain a small but reliable difference between groups trained under single- and dual-task conditions and then tested under single-task conditions. However, compared to the designs used by Frensch and his colleagues, this study is not ideal. The single-task test phase immediately followed training for participants trained under single-task conditions, whereas the comparable test for participants trained under dual-task conditions occurred somewhat later in the experiment, after a dual-task test. The possible contaminating effects of the prior test in the group trained under dual-task conditions are unknown. Thus, although the results of these various studies are contradictory, the experiments reported by Frensch, Lin, & Buchner (1998, Exps. 1a and 1b) seem to come closest to the ideal of a design specifically intended to allow the performance and learning accounts to be distinguished.

These studies are important because they tend (putting aside Heuer and Schmidtke's data) to support a conception of implicit learning in which the role of attention is rather different from that seen in more typical (explicit) learning tasks: Full attention seems not to be necessary for implicit sequence learning to proceed normally. On the other hand, there are some reasons why the results should be regarded with a certain amount of caution. For example, Frensch et al. (1998) gave all of their groups both single- and dual-task training, rather than giving one group just single-task training and another group just dual-task. In addition, the training conditions of even the most extreme groups (2-DT/5-ST vs. 6-DT/1-ST) only differed on 4 blocks of trials. The design Frensch et al. adopted therefore tends to reduce the likelihood of obtaining a group difference in transfer scores and their study may therefore constitute a fairly conservative test of the experimental hypothesis.

Moreover, Frensch et al. included in their analysis all participants whose average tone-counting error on the dual-task training blocks was 20% or less. This is a very liberal criterion and means that participants were included in the analysis who may have been allocating minimal attention to the secondary task. Such participants would be expected to show large transfer scores since, functionally, they are performing the task just like single-task participants. Naturally, a strong test of the experimental hypothesis requires some evidence that dual-task participants were indeed concentrating to an adequate level on the secondary task. It is not clear why Frensch et al. adopted this liberal criterion rather than the more common criterion of 10% (e.g., Cohen et al. 1990).

Thirdly, Frensch et al. used training and transfer sequences which have a number of undesirable properties. For instance, inspection of the training sequences (ABCDEADFC and ABCDECFBE) reveals immediately that they contain no reversals, that is, occasions on which the target moves back to the location it occupied on the last-but-one trial (e.g., ABA). In contrast, the quasi-random sequence presented in the test stage does contain reversals. Suppose participants learn the abstract feature of the training sequences that they contain no reversals. At any moment during the training phase the participant knows that the target will not appear in 2 of the 6 possible locations: the location of the last trial (since there are no immediate repetitions) and the last-but-one location. In the test phase, the target does sometimes appear in the reversal location, and RTs would be expected to be particularly slow on such trials. Hence the transfer scores Frensch et al. obtained may have been inflated: In fact, it is possible that many participants had no specific sequence knowledge at all. In that case, the fact that the transfer scores did not differ is uninformative.

The presence versus absence of reversals is only one feature that differs between the training and test sequences Frensch et al. used. Reed and Johnson (1994) have identified several such factors (e.g., *rate of coverage*, the mean number of trials required to see the target appearing in each of the possible locations) and have provided an elegant method for avoiding these difficulties. Rather than switching participants to a quasi-random sequence, they are transferred to a sequence that is structurally identical to the training sequence but which is instantiated differently in terms of assignment to screen locations.

Shanks and Channon (2002), therefore, conducted a conceptual replication of Frensch et al.'s Experiment 1a, but presented one group with only single-task training blocks and another with only dual-task blocks, and we used Reed and Johnson's sequences to avoid the problems described above. In Experiment 2 we tested participants under dual-task conditions (as in Frensch et al.'s Experiment 1b) as well as under single-task conditions.

The training and test sequences in our experiments were A = 1-2-1-3-4-2-3-1-4-3-2-4 and B = 4-2-4-3-1-2-3-4-1-3-2-1, where 1-4 are screen locations. These sequences are structurally identical and are related by the transformation 1↔4. They are balanced for simple location and transition frequency. Each location (e.g. 1, 2, 3, 4) occurs three times in each 12 trial sequence, and each possible transition (e.g., 1-2, 1-3, 1-4, etc.) occurs once. But at the level of three (or more) consecutive locations the two sequences differ. Reed and Johnson (1994) gave sequences of three locations the name *second order conditionals* (SOCs), which refers to the fact that the next location in the sequence of dot movements can be predicted from the last two locations. For example in sequence A, 1-2 is always followed by 1, whereas in sequence B, it is always followed by 3. Because the sequences are structurally identical, any increase in RTs in the test block must reflect sequence knowledge rather than the confounding of structural properties such as the frequency of reversals.

Participants were randomly assigned to two critical groups: Single or Dual. All participants performed 14 blocks of 96 trials in the training phase. During blocks 1–10, the dot followed sequence SOC1. Participants in group Single performed the RT task alone, while participants in group Dual performed the secondary task as well. On blocks 11–14, both groups were treated identically. On block 11, sequence SOC1 was used under single-task conditions. Participants in group Dual were informed prior to this block that they were no longer required to perform the tone-counting task, but that they should continue to respond to the target as rapidly as possible. On block 12 sequence SOC2 was used, and on blocks 13–14 sequence SOC1 was re-introduced. The relative slowing down on block 12 compared to blocks 11 and 13 provided the main index of sequence knowledge. In the SRT task, four boxes were presented along the bottom of the computer screen. A dot (2 mm in diameter) appeared in the center of one of these boxes on each target location trial. Participants were instructed to indicate locations 1–4 as quickly as possible by using the V, B, N, and M keys located across the bottom of the keyboard, respectively.

Each block of target-location trials began at a random point in the sequence, and thereafter targets appeared according to the sequence that corresponded to the particular condition and block type. A target-location trial ended when a participant pressed the correct key, at which time the target was erased. The next trial began 200ms later. Response latencies were measured from the onset of the targets to the completion of correct responses.

For approximately half the participants in the Single and Dual groups, the training sequence (designated SOC1) was A and the test sequence (SOC2) was B. For the remaining participants these were reversed. Each of these 12-item sequences was repeated 8 times in each block of 96 trials.

Details of the secondary task are as follows. In each block of dual-task RT trials, a 100-ms computer generated tone was emitted 100-ms after each correct target location response. Each tone was randomly determined to be either low (1000Hz) or high (2000Hz), and participants were instructed to count the number of high tones emitted during each block of trials. At the end of each block, participants were asked to provide their count. Feedback presented at the end of the block encouraged them to count the tones accurately. Participants were excluded from the analysis if they made more than 10% errors on average.

Figure 1 shows mean RTs for each group across blocks. Participants in group Single rapidly reached a stable level of short RTs which they maintained across the training blocks. Participants in group Dual were slower initially, but on blocks 8–10 RTs were equivalent, suggesting that participants in the latter group had developed the skill of combining the 2 tasks with minimal interference of tone-counting on RTs. On block 11, all participants performed the SRT task under single task conditions, and no RT difference was present.

The principal data concern the changes in RTs on block 12. For group Single, the introduction of sequence SOC2 was accompanied by a very substantial increase in reaction times, but RTs returned to their earlier level on blocks 13–14. For group Dual, a very small increase in RTs occurred on block 12, with RTs again returning to their earlier level on blocks 13–14. To assess sequence knowledge, we computed a difference (D) score based on the difference between the RT on block 12 and the average RT on blocks 11 and 13. The mean D scores are shown on the left of Figure 2. Plainly, there was less evidence of sequence learning under dual-task conditions[1].

Shanks and Channon's (2002) findings are straightforward: Under common testing conditions, sequence knowledge is substantially greater in a group trained under single-task conditions than in one trained under dual-task conditions. We thus failed to replicate the null effect reported by Frensch, Lin, and

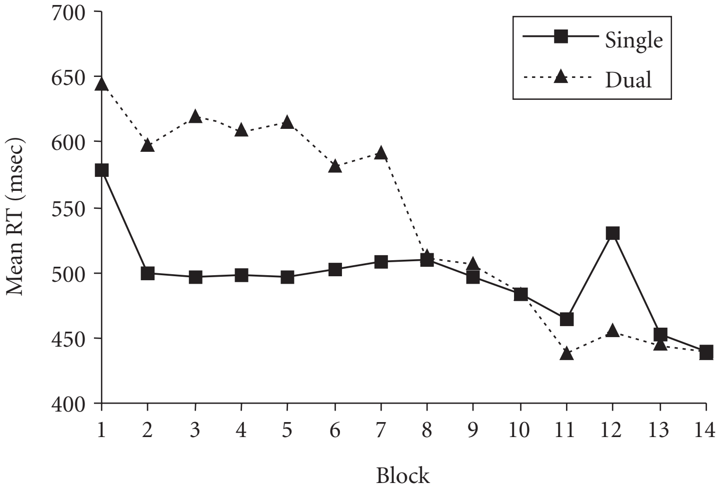


**Figure 1.** Mean reaction times across blocks of trials in Shanks and Channon's (2002) Experiment 1. Both groups were trained on sequence SOC1 on blocks 1–10. Group Single performed under single-task conditions in all blocks while group Dual performed under dual-task conditions in blocks 1–10 prior to the removal of the secondary task on block 11. Sequence SOC1 was used for both groups on blocks 11 and 13–14 while sequence SOC2 was used on the transfer block, block 12.

Buchner (1998). At variance with the attentional hypothesis of implicit learning, the results suggest that the division of attention impairs sequence learning.

In a second experiment we (Shanks & Channon 2002: Exp 2) predicted that transfer scores would again be lower in a dual- than in a single-task training group even if testing were conducted for both groups under dual-task conditions (contrasting with the results obtained by Frensch, Lin, and Buchner 1998, Exp. 1b). Although D scores might be lower overall under dual- than under single-task testing conditions (because of suppression), we still anticipated a group difference. Participants were randomly assigned to four groups constructed according to whether training took place under single- (groups Single/Single and Single/Dual) or dual-task (groups Dual/Single and Dual/Dual) conditions. During blocks 1–8, participants in groups Single/Single and Single/Dual performed the RT task alone while participants in groups Dual/Single and Dual/Dual performed the secondary task as well. The sequence (SOC1) was A for roughly half the participants in each group and B for the remainder. On blocks 9–11, groups Single/Dual and Dual/Dual performed the SRT task combined with the tone-counting task whereas the other two groups per-

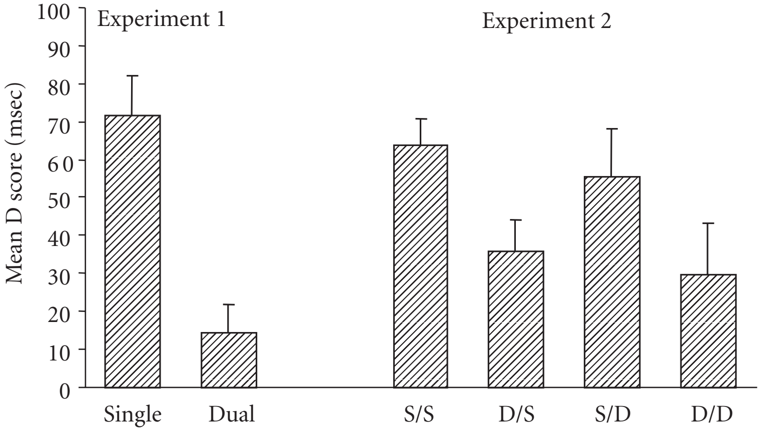


**Figure 2.** Mean (+ s.e.m.) D score (RTs on the transfer block minus the average of RTs on the preceding and subsequent blocks) in each group of Shanks and Channon's (2002) Experiments 1 and 2. S/S: Single/Single, D/S: Dual/Single, S/D: Single/Dual, D/D: Dual/Dual.

formed it alone. Participants in group Single/Dual were informed prior to this block about the tone-counting task. On block 10 sequence SOC2 was used, and on block 11 sequence SOC1 was re-introduced. Only one block with sequence SOC2 followed the transfer block.

Figure 3 presents mean RTs for each group across blocks. Participants in groups Single/Single and Single/Dual rapidly reached a stable level of short RTs which they maintained across blocks 1–8. Participants in groups Dual/Single and Dual/Dual were considerably slower. On block 9 the new conditions came into effect and RTs were now considerably longer in the two groups receiving dual-task conditions (Groups Single/Dual and Dual/Dual). Between blocks 8 and 9 there was an almost perfectly symmetrical relationship between the speed-up of RTs in group Dual/Single and the slowdown in group Single/Dual. Block 9 also reveals a form of behavioral contrast: single-task responding is slower after single- than dual-task training (also evident in Experiment 1) while dual-task responding is faster after dual- than single-task training.

The principal data concern the change in RTs on block 10. Contrasting with the results of Frensch et al. (1998, Exp. 1b), the increase was largest in groups Single/Single and Single/Dual than in the other two groups, for whom the in-

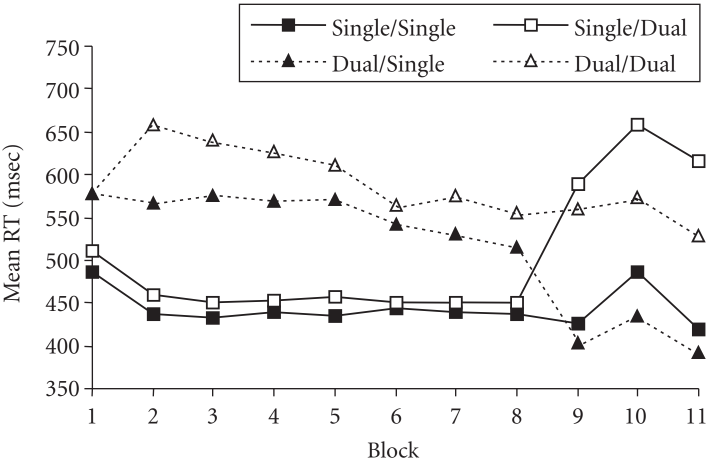


**Figure 3.** Mean reaction times across blocks of trials in Shanks and Channon's (2002) Experiment 2. Groups Single/Single and Single/Dual performed under single-task conditions in blocks 1–8 while groups Dual/Single and Dual/Dual performed under dual-task conditions. The secondary tone-counting task was performed concurrently with the RT task on blocks 9–11 in groups Single/Dual and Dual/Dual, while groups Single/Single and Dual/Single performed these blocks under single-task conditions. Sequence SOC1 was used for both groups on blocks 1–9 and 11 while sequence SOC2 was used on the transfer block, block 10.

crease was very small. That is to say, there was more disruption in responding in the groups trained under single-task conditions than in those trained under dual-task conditions, regardless of testing conditions, and this is consistent with the secondary task interfering with sequence learning. Figure 2 shows this pattern more clearly in terms of difference scores.

Overall these results are very straightforward: they confirm that under the conditions used by Shanks and Channon (2002), sequence learning is impaired by a secondary task. We replicated the results of Experiment 1, with D scores being larger in group Single/Single than in group Dual/Single, but we also found the same pattern under dual-task testing conditions. Although testing conditions had an overall effect on RTs (which were longer under dual- than single-task conditions), they had no detectable effect on the expression of sequence knowledge which continued to be greater for those participants trained under single-task conditions.

The findings of Shanks and Channon's experiments are consistent in suggesting that attention cannot be divided without detrimentally affecting implicit sequence learning. This is most clear in Experiment 2 where dual-task training conditions impaired sequence learning, independently of testing conditions. Our results are in conflict with Frensch et al.'s in two respects. First, in both experiments we obtained greater learning scores in groups trained under single-task conditions than in groups trained under dual-task conditions, regardless of the testing conditions: in their comparable experiments (Frensch et al. 1998, Exps. 1a and 1b), no such difference was evident. Secondly, our findings do *not* lend support to the suppression hypothesis. Recall that the suppression hypothesis states that dual-task testing conditions suppress the expression of sequence knowledge: group differences are flattened out by the secondary task. The evidence for this hypothesis comes from a number of experiments: for instance, Frensch, Lin, and Buchner (1998, Exps. 2a & 2b) trained participants on a repeating sequence under single- or dual-task conditions and then tested them under both single- and dual-task conditions. Transfer scores were generally lower on the dual- than on the single-task test, regardless of training conditions. In another study, Frensch, Wenke, and Rünger (1999) trained participants under dual-task conditions and tested them first under dual- and then single-task conditions, and again found that transfer scores were lower on the dual- than on the single-task test.

In contrast, sequence knowledge in Experiment 2 was not better expressed under single- than under dual-task testing conditions: There was no overall effect of testing conditions in the ANOVA described above. Indeed, in one specific comparison we find evidence of a "reverse" suppression effect, in that D scores were numerically greater in group Single/Dual than in group Dual/Single. This is contrary to the suppression hypothesis because, according to Frensch et al., the two groups should have learned the sequence equally but the former group should have suffered suppression in the test stage.

Why do our results conflict with those of Frensch and his colleagues? The experiments differ in many ways but we contend that the use of within-subjects designs in most of the critical suppression studies (e.g., Frensch et al. 1999) is a significant concern. If participants are first tested under (say) dual-task conditions and then under single-task conditions, the possibility arises of contamination of the later test by the earlier one. We have very little reason to discount the possibility of such contamination. In Experiments 1 and 2 this issue was circumvented by the use of between-subjects designs. The suppression hypothesis predicts larger D scores under single- than dual-task testing conditions. Yet the pattern of results was the exact converse of this. We contend that Frensch et al.'s conclusion – that tone-counting has no effect on transfer scores provided that common testing conditions are used – is not in general correct. Our results therefore challenge the idea that implicit learning can be usefully distinguished from explicit learning on the basis of its attentional requirements, as Frensch (1998; Frensch et al. 1998; Frensch et al. 1999) and others (e.g., Cleeremans 1997; Hayes & Broadbent 1988; Heuer & Schmidtke 1996; Jiménez & Méndez 1999; Schmidtke & Heuer 1997; Stadler 1995) have suggested.

## 1.2 Other secondary tasks

The secondary task of tone-counting does appear to affect sequence learning. I now turn to a consideration of other secondary tasks. As a number of researchers have noted (Frensch et al. 1998; Heuer & Schmidtke 1996; Schmidtke & Heuer 1997; Stadler 1995), even if a secondary task such as tone-counting does affect sequence learning, the locus of this need not be at the level of competition for attentional resources. The effects of a secondary task may be due, for example, to specific interference rather than competition for central capacity. There is now a sizable body of work attempting to isolate the exact mechanisms by which different secondary tasks might affect performance.

Stadler (1995) used a memory-load secondary task in the expectation that this would be a "purer" attention-demanding task than tone-counting. Compared to a single-task group, participants who memorized a 7-letter string at the outset of each block of SRT trials and who recalled it at the end of the block showed a significantly reduced transfer effect when shifted to a random sequence, although the effect was much smaller than that caused by tone-counting. Stadler (1995, Exp. 2) downplayed this finding because of a post-hoc reanalysis of the data according to whether participants were aware or not of the sequence and concluded that implicit sequence learning is not attention-demanding. In unaware participants, the difference in sequence learning between the memory-load group and the single-task control group was reduced. However, the difference was not eliminated and loss of statistical power makes the reduction hard to interpret. There remains clear evidence of an overall disruption of sequence learning as a result of the memory load. Interpretation is made additionally problematic, though, because Stadler's experiments confounded learning with performance: The secondary task was present in both the training and transfer blocks. Furthermore, Reed and Johnson (1994) have documented a number of problems with the sequences Stadler used, and Willingham, Greenberg, and Thomas (1997) were unable to replicate some of his findings. Thus it is difficult to draw firm conclusions from this study.

Another study which employed a secondary task other than tone-counting was conducted by Heuer and Schmidtke (1996). These authors pointed out that the tone-counting task has two components, memorizing the current number of tones and classifying each tone as high or low. In contrast to the findings of Stadler (1995), they (Heuer & Schmidtke 1996, Exp. 2) found that sequence learning was completely unaffected by 2 secondary tasks (the verbal and visuo-spatial tasks of Brooks 1967) which impose a memory load without additional stimulus processing, whereas it was affected by a task (pressing a foot pedal in response to a high-pitched but not a low-pitched tone) requiring stimulus processing without a memory load. On the assumption that the Brooks secondary tasks were to some degree attention-demanding, Heuer and Schmidtke's data represent quite strong evidence that sequence learning in the SRT task does not require central attentional resources: So long as an appropriate secondary task is used (i.e., one that does not require stimulus processing in the response-stimulus interval of the main task), no interference of sequence learning will be observed. On the other hand, Heuer and Schmidtke's studies can again be criticized on the grounds that they used training and transfer sequences which were not structurally identical and hence which did not control for factors such as the frequency of targets at each location or rate of reversals.

These memory-load studies, in which participants maintain a memory load across an entire training block with no trial-by-trial secondary task events, do hint that implicit sequence learning does not require attention. But some improvement in methodology seems warranted. In the experiment reported next I essentially replicate Shanks and Channon's (2002) Experiment 1 but using a memory-load rather than a tone-counting secondary task in an attempt to clarify this issue. There were 24 participants in the experiment, 12 per group. Those in group Single performed the SRT task alone on blocks 1–10 with sequence SOC1 while those in group Dual performed the SRT task in combination with a memory load task. A string of 7 different consonants (excluding Y) was presented for 10 sec at the beginning of each block, with a different string used for each block. Strings were written on separate white cards and participants were instructed to treat the two tasks as being of equal importance. Immediately after the string was removed, participants commenced the SRT task for that block, and at the end of the block they reported the string to the experimenter. All participants were tested under single-task conditions on blocks 11–13 with the memory load task removed. Participants in group Dual were informed prior to block 11 that there would be no memory load. On block 12 sequence SOC2 was used, and on block 13 sequence SOC1 was re-introduced.

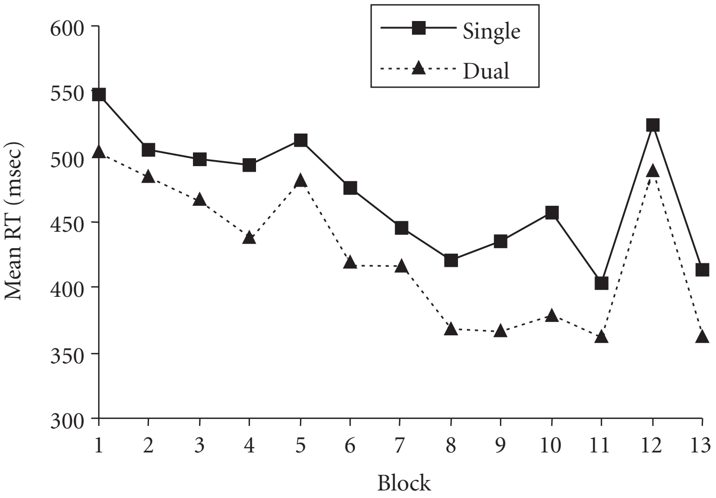


**Figure 4.** Mean reaction times across blocks of trials in an experiment in which one group of participants performed a concurrent memory load task. Both groups were trained on sequence SOC1 on blocks 1–10. Group Single performed under single-task conditions in all blocks while group Dual performed under dual-task conditions in blocks 1–10 prior to the removal of the secondary task on block 11. Sequence SOC1 was used for both groups on blocks 11 and 13 while sequence SOC2 was used on the transfer block, block 12.

The memory-load task was performed with a high degree of accuracy, participants recalling a mean of 6.10/7 letters correctly in position. Performance did not vary systematically across blocks and no participant achieved an overall mean of less than 5.10/7 correct.

Figure 4 presents mean RTs for each group across blocks. Participants in group Dual were somewhat faster than those in group Single across blocks 1–10. On block 11 all participants performed under single task conditions and RTs were fairly close. The principal data again concern the change in RTs on block 12. The increase was very similar, and greater than zero, in the two groups. The mean D scores, which did not differ, $t < 1$, were 115 msec in Group Single and 130 msec in Group Dual. Both scores were significantly greater than zero, $t(11) > 6.52, p < .001$.

In contrast to the data obtained when tone-counting was the secondary task, the present results appear to support the conjecture that sequence learning does not place significant demands on attentional resources. The findings endorse the conclusions of Heuer and Schmidtke's (1996) study and suggest that the effect of tone-counting may arise from one of its nonattentional properties. However, the obvious counterargument is that these memory load tasks may simply not have been sufficiently taxing. Unlike tone-counting, they do not necessarily impose a task requirement on every single SRT trial. Perhaps participants were able to rehearse the letter string sufficiently in the first few trials of each block to commit it to memory with only occasional "refreshing" being necessary to maintain it until the end of the block. In that case, full attentional resources would have been available for the SRT task. Indeed, the fact that RTs were no slower in the Dual than the Single group (if anything, they were faster) is supportive of this conjecture.

In response to this alternative viewpoint, Jiménez and Méndez (1999, see Chapter 2, current volume) developed a secondary task which they argued would avoid the twin problems of being insufficiently demanding and of introducing stimuli (i.e., tones) irrelevant to the primary task. They used a probabilistic sequence learning task in which the target stimulus could be one of 4 symbols; as well as reacting to the location of each target, dual-task participants had to count the frequency of 2 of the symbols. The attraction is that symbol-counting imposes a continuous attentional demand but uses stimuli (i.e., the targets themselves) to which the participant already has to attend to carry out the primary task. This secondary memory-load task had no detectable effect on sequence learning. Jiménez and Méndez speculated that the use of a probabilistic sequence was critical in their study for revealing a form of learning which is independent of attention. In a conceptual replication of this experiment, we (D. Shanks & S. Banfield, unpublished data) obtained a rather different outcome, however. We used the probabilistic generation procedure of Schvaneveldt and Gomez (1998) described previously and gave participants 9 blocks of trials (100 trials/block), either with or without the symbol-counting task, prior to a test block without the secondary task. Whereas the single-task group showed good sequence learning (indexed by faster responses to consistent than to inconsistent targets) on the test block, the dual-task group showed almost no sequence learning. Hence it is not the case that probabilistic sequences necessarily invoke a form of learning that makes no demands on attention.

But why did we get results different from those of Jiménez and Méndez? The major difference between the studies is that their one monitored learning over many thousands of trials whereas ours looked at learning over only a few hundred trials. This raises the possibility that if we re-ran our experiment with a longer training stage we would now find no difference between the single- and dual-task groups. This is exactly what we found when we doubled the amount of training. Under these circumstances, both groups showed a healthy consistent/inconsistent difference on the final single-task test block. We therefore offer the following perspective on Jiménez and Méndez' results: because participants had had so much training at combining the two tasks, by the time they began to show evidence of sequence learning the secondary task had become largely automated and hence made little demand on attention. Thus the dual-task group showed the same degree of sequence learning as the single-task group. But if Jiménez and Méndez had used a more easily learnable sequence and had tested their participants much earlier (as we did in our study), they would have observed a dual-task decrement. Just because participants can eventually combine two tasks with minimal interference does not mean that *learning* to perform those tasks efficiently does not require attention.

In the next experiment we used two secondary tasks which, like symbol-counting, required processing on every trial. These were mental arithmetic and articulatory suppression. As with symbol-counting, the important element of these secondary tasks is that although they require trial-by-trial processing, no external stimuli are presented to the participant in the intertrial intervals and thus they do not require the sort of categorization process which Heuer and Schmidtke (1996) suggested was important.

The procedure was very similar to the previous experiment. There were 36 participants, 12 per group. Participants in the Mental Arithmetic group were given a number, randomly chosen between 500 and 800 (from the set 667, 796, 632, 504, 732, 591, 674, 800, 555, and 694), at the beginning of each block. They were then required to subtract 3 from that number and say it aloud before making a keypress in the primary SRT task and to do this on every trial. Participants in the Articulation group were also given a number from the set above at the beginning of each block but in this case were required to say that number aloud once before making each keypress in the primary task. All participants were again tested under single-task conditions on blocks 11–13 with the memory load tasks removed.

Figure 5 presents mean RTs across blocks. RTs were very slow in the Articulation group and even slower in the Mental Arithmetic group. On block 11 all participants performed under single task conditions and RTs were very close. The RT increase on block 12 were much greater in the single-task group than in the other groups. The mean D scores were 86 msec in Group Single, 24 msec in Group Articulation, and –1 msec in Group Mental Arithmetic. These scores differ [$F(2,33) = 9.38, p < .05$] and plainly suggest that both mental arithmetic and articulation impair sequence learning ($p < .05$ in each case).

It might be argued that the detrimental effects of these secondary tasks on learning are due not to the presence of the secondary task *per se* but rather are a

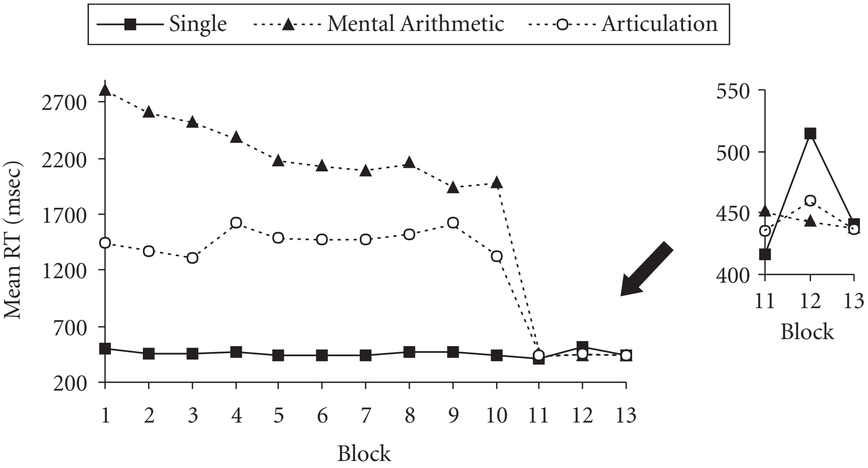


**Figure 5.** Mean reaction times across blocks of trials in an experiment in which one group of participants performed a concurrent mental arithmetic task and another performed an articulation task. All groups were trained on sequence SOC1 on blocks 1–10. Group Single performed under single-task conditions in all blocks while groups Mental Arithmetic and Articulation performed under dual-task conditions in blocks 1–10 prior to the removal of the secondary task on block 11. Sequence SOC1 was used for all groups on blocks 11 and 13 while sequence SOC2 was used on the transfer block, block 12. The inset figure shows the results from blocks 11–13 in larger scale.

by-product of the change in the timing of the trials that they create. Specifically, these tasks lead to an increase in the stimulus-stimulus interval of up to 2500 msec in the case of mental arithmetic. However, careful experiments by Willingham, Greenberg, and Thomas (1997) tested this claim directly by looking at the effects of lengthened response-stimulus intervals on sequence learning. Although they did not examine increases as large as 2500 msec, their results provided no evidence whatsoever that increasing the effective stimulus-stimulus interval, or making it more variable, affected learning.

## 1.3 Summary

Where do these results leave us? There are several related conclusions. It is not the case that tone-counting only has an effect on performance and not on learning. Shanks and Channon's experiments seem to clearly falsify this claim and to show instead that there is an effect on learning. But that does not mean that attention is necessary for implicit sequence learning as the critical effects of tone-counting might be on some other process (e.g., stimulus processing). This argument can hardly be made about memory load tasks, which do not affect learning. However, such tasks may only make minimal attentional demands. Symbol-counting (in some circumstances), mental arithmetic, and articulation, tasks in which no stimuli are presented in the inter-trial intervals, do affect sequence learning.

It does not appear, I conclude, that the attentional independence of implicit learning has been satisfactorily established. To endorse such a claim, one would have to argue either that memory load tasks are sufficiently demanding to comprehensively reduce the resources participants have available for learning the stimulus sequence, or that the adverse effects of symbol-counting, mental arithmetic, and articulation are secondary to their effects on sequence organization and timing. Neither of these claims seems to be strongly supported by the available evidence.

## 2. Awareness and implicit sequence learning

In the SRT task, participants' sequence knowledge may be expressed either indirectly, via reduction in response latency to predictable targets (priming), or directly, via recall, recognition, prediction, or generation tests thought to require conscious knowledge.

It has often been claimed that direct and indirect measures of sequence knowledge can be dissociated, and such dissociations have been taken to support the existence of an "implicit" learning process which is independent of explicit learning. Since implicit sequence learning is proposed to be unrelated to and dissociable from consciously-accessible knowledge, it is conjectured to be an unconscious process. In this section I examine in detail whether such an unconscious learning process needs to be postulated or whether sequence learning can be understood from the perspective of a unitary learning system.

The case for implicit learning depends crucially on the validity of the tests used to index awareness. A common distinction is drawn between "subjective" and "objective" tests, where the former ask the participant to report his/her state of awareness while the latter demand some forced-choice discrimination. There is absolutely no doubt that participants' verbal reports about training sequences in SRT experiments fail to incorporate all of the information and serial dependencies that can be detected chronometrically in their primed keypresses (e.g., Shanks & Johnstone 1998; Willingham, Greeley, & Bardone 1993). Another way to elicit subjective reports is to ask participants to generate the sequence they were trained on and then give a "metacognitive" assessment of their confidence in their generation accuracy. For instance, Shanks and Johnstone (1998) initially trained participants on SOC sequences and then asked them to freely generate those sequences. Participants who reported that they were guessing in this test nonetheless generated their training sequence far better than would be expected by chance.

Although such results establish that in at least one sense sequence learning can be implicit, they may have a rather mundane explanation: As many authors have pointed out (e.g., Merikle, Smilek, & Eastwood 2001; Reingold & Merikle 1990), an adequate test of awareness must be exhaustive which means that the test must be sensitive to all of the conscious knowledge of which the participant is in possession. The exhaustiveness criterion is a problem for subjective tests of awareness because there is little to guarantee in such tests that the participant has indeed reported all available knowledge. For example, he or she may simply choose to withhold conscious knowledge held with low confidence: in signal detection terms, the participant's response criterion may be very strict. If that happens, then an implicit measure may dissociate from a subjective measure simply because the former is more sensitive to conscious knowledge. To avoid this problem it would be necessary for the experimenter to induce and motivate the participant to report all relevant knowledge, including hunches and so on. This has not often been attempted. One way of achieving it is to force the participant to report a given number of pieces of information, a procedure which, when compared to unforced recall, can significantly improve performance (e.g., Schmidt & Dark 1998). In fact, some studies which have probed quite thoroughly for all available verbalizable knowledge have even ended up finding that *all* knowledge is accessible for report (e.g., Marescaux 1997).

As a reaction to this problem in the interpretation of subjective tests, it has been widely accepted that objective rather than subjective tests provide the best measures of awareness, on the grounds that they are more likely to be exhaustive. However, prior examples of dissociations between direct and indirect tests when the former are objective are rather equivocal. Key results appearing to demonstrate dissociations have not been replicated (see Curran 1997; Shanks & Johnstone 1999) or have been criticized (Dienes & Berry 1997; Perruchet & Amorim 1992; Perruchet & Gallego 1993; Perruchet, Gallego, & Savy 1990; Shanks & Johnstone 1998; Shanks & St. John 1994) on a variety of methodological grounds. To take just one example, it is almost always the case that the direct and indirect measures are taken at different times in distinct test phases and this creates a number of potential difficulties: for instance, if the direct test is administered some time after the indirect test they may be differentially affected by forgetting. Moreover, with distinct test phases participants may be inclined to adopt different response sets, response biases, strategies, levels of motivation and so on which may significantly affect the relative levels of performance obtained in the tests. The availability of a testing method which enables concurrent direct and indirect knowledge assessment would allow many methodological problems of this sort to be finessed. Shanks and Perruchet (2002) have introduced and exploited such a method which I describe below.

Many of the key methodological issues at the heart of the debate over implicit learning can be illuminated by reference to a recent study by Destrebecqz and Cleeremans (2001: see pp. 181–212, current volume). These authors trained participants on an SRT task in which the target moved according to a repeating SOC sequence like those used in the experiments described in the previous section. The learning phase consisted of 15 blocks of 96 trials for a total of 1440 trials. For half the participants there was a response-stimulus interval (RSI) of 250 msec between the execution of one response and the appearance of the next target while for the remainder the RSI was 0 msec. RTs reduced somewhat across blocks 1–12 in both groups. Then on block 13 the sequence was changed to a different SOC sequence with the original sequence being reintroduced on blocks 14 and 15. Destrebecqz and Cleeremans found that RTs were significantly greater in the transfer block (block 13) and concluded that their participants had learned something about the sequence which permitted them to anticipate, perhaps unconsciously, where each successive target would appear and hence make fast, “primed,” responses.

To ascertain whether this sequence knowledge was conscious or unconscious, Destrebecqz and Cleeremans presented two tests following block 15. First, they informed participants that there had been a repeating sequence and asked them to generate a sequence of keypresses under both “inclusion” and “exclusion” conditions following the logic of opposition developed in the process dissociation procedure (Jacoby, Toth, & Yonelinas 1993). In the inclusion test participants were to try to reproduce the sequence they saw in training while in the exclusion test they were to *avoid* reproducing the training sequence or any of its parts. The key finding was that, at least for participants in the RSI = 0 msec group, the sequence generated under exclusion conditions contained more chunks from the training sequence than would be expected by chance. Thus participants’ sequence knowledge, Destrebecqz and Cleeremans argued, was unconscious in the sense that they could not exert voluntary control over it when explicitly required to exclude it in generating a sequence[2].

The second assessment of awareness comprised a recognition test. Participants were shown short sequences of 3 targets half of which came from the training sequence (e.g., 342, 423) and half of which were new (in fact they weren't completely new since they came from the block 13 transfer sequence). Participants executed each sequence just as in the training stage and then made an old/new rating on a 6-point confidence scale. The key finding was that recognition was above chance in the RSI group but not in the no RSI group. If recognition is indeed a measure of conscious sequence knowledge then the results of the no RSI group seem to suggest, in line with the exclusion generation data, that participants' knowledge of the sequence was entirely implicit. Note however that the measure of conscious knowledge was obtained in a test conducted some period of time after the transfer test of implicit knowledge. Thus to conclude that participants possessed implicit but not explicit knowledge requires assuming, amongst other things, that their state of knowledge had not altered (e.g., by interference or forgetting) during the lengthy interval prior to the recognition test in which they performed the inclusion and exclusion generation tests.

Note also that Destrebecqz and Cleeremans tested recognition with 3-item sequences. Shanks and Johnstone (1999) and Shanks and Perruchet (2002) have shown that recognition is far superior with 6- than with 3-item sequences. Moreover, in two of the dual-task experiments described above (one with a memory load secondary task, the other with mental arithmetic and articulation), recognition was above chance in all groups when tested with 6-item sequences. After the final SRT block, participants in those experiments were presented with test trials comprising 6-location sequences which they responded to exactly as in the training stage (without a secondary task). Half of these test sequences were fragments of the training sequence (i.e., old) and others were not (new). Figure 6 shows the mean recognition ratings for old and new sequences for each of the 5 groups and reveals clear old/new discrimination in all groups, with $p < .05$ in each case. Consistent with previous research (Perruchet & Amorim 1992; Perruchet, Bigand, & Benoit-Gonin 1997; Shanks & Johnstone 1999; Shanks & Perruchet 2002), this implies that participants do have at least some conscious access to their knowledge of the sequence and that if implicit learning is defined in terms of a lack of awareness, then knowledge acquired in the SRT task is not unconscious. I conjecture that the lengthy delay that Destrebecqz and Cleeremans' interposed prior to their recognition test, together with their use of 3-item test sequences, contributed to their (spurious) null result. This conjecture is supported by the results of experiments reported by Shanks, Wilkinson, and Channon (2002) who again found better recognition with 6- than with 3-item test sequences and who also found that, even with 3-item sequences, participants performed above-chance in recognition if

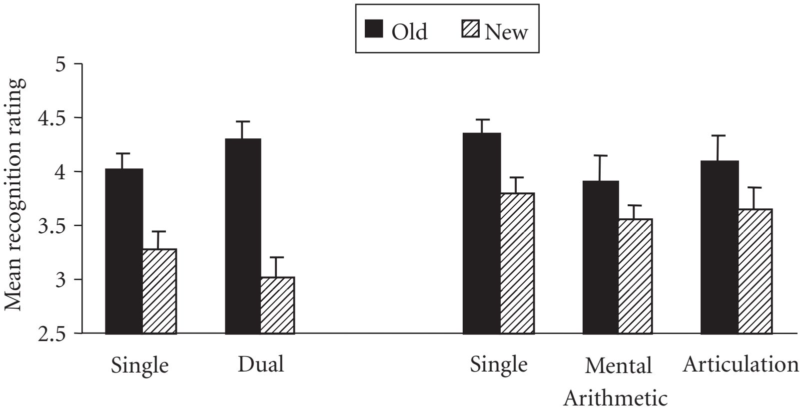


**Figure 6.** Mean (+ s.e.m.) recognition ratings to old and new test sequences in the experiments described previously (see Figs 4 and 5). After the final block of the SRT task, participants were presented in a random order with 24 6-item test sequences, 12 of which were from the SOC1 training sequence (old) and 12 of which were from the SOC2 sequence (new). They executed each of these sequences prior to making a recognition rating (1 = *certain new*, 6 = *certain old*). In each group, the old/new difference was statistically significant.

the test phase followed the study phase immediately. Shanks et al. were also unable to detect any qualitative difference between RSI and no RSI conditions.

As described above in regard to Destrebecqz and Cleeremans' (2001) study, both generation and recognition tests are amongst those which have been extensively studied. If participants can be shown by some priming measure to possess knowledge of sequential structure, but fail to perform above chance on an objective test, then it is hard to argue that this is simply a problem of sensitivity or exhaustiveness. In a recognition test, for example, the retrieval context is identical to the learning context and a forced-choice old/new response is required. Although I have raised concerns over Destrebecqz and Cleeremans' (2001) results, there is little doubt that priming can be dissociated from performance on an objective test. I now present an example of such a dissociation from a study by Shanks and Perruchet (2002). However, far from establishing the validity of the implicit/explicit distinction, this example can instead be used to undermine it. It demonstrates the limited usefulness of dissociations in inferring mental processes.

Shanks and Perruchet used a recognition test similar to that of Destrebecqz and Cleeremans and found that participants were able to discriminate old from new sequences overall (the recognition ratings were similar to those shown in

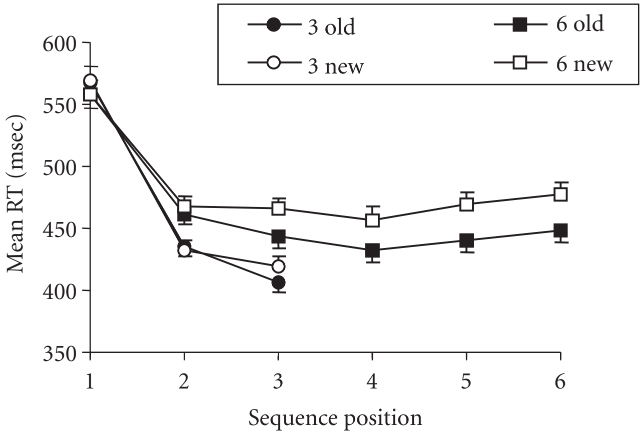


**Figure 7.** Mean (± s.e.m.) reaction time to targets 1–6 of old and new test sequences in Shanks and Perruchet's (2002) experiments. For some participants test sequences were of length 3 while for others they were of length 6. Targets 3–6 are predictable from the preceding targets whereas targets 1–2 are not. It can be seen that RTs are faster for old than new sequences after target 2.

Figure 6). However, we also found that participants responded faster in executing the old sequences. These data are shown in Figure 7. Here test sequences were of length 3 for some participants and length 6 for others. The graph plots RT for each position in the test sequences. Priming did not occur for the first and second targets (i.e., there was no old-new RT difference) because 2 elements of context are required to discriminate old from new SOC sequences. After position 2, however, RTs were reliably faster for old than for new sequences. The fact that old and new sequences were discriminated both in their (direct) recognition ratings and in their (indirect) speed of execution allowed us to look in more detail at the correlation between these measures. This question was addressed by computing the relative execution speeds for old and new test trials for which participants gave identical recognition ratings. If, for example, priming and recognition are dependent on distinct knowledge sources, then priming might be expected even when old and new sequences are not differentially recognized. Alternatively, if priming and recognition are tightly coupled then old and new test sequences given identical recognition ratings should be executed with equivalent latencies.

These results are shown in Figure 8. It is plain that the predictable targets in old sequences elicited more rapid responses than the corresponding unpredictable targets in new sequences at each recognition score. Thus the

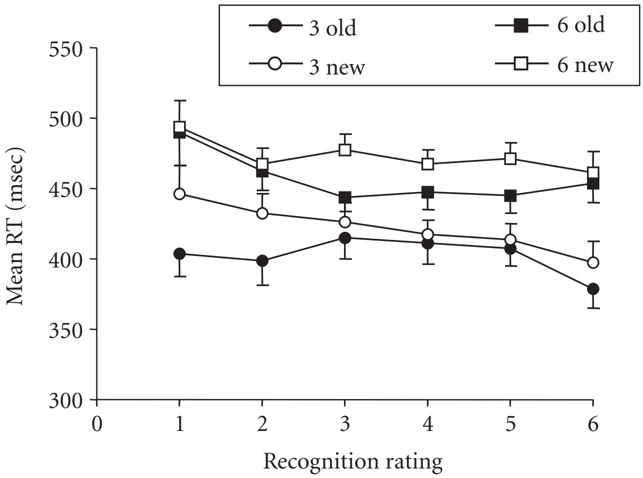


**Figure 8.** Mean (± s.e.m.) response latency (msec) to targets in old and new test sequences in Shanks and Perruchet's (2002) experiments is plotted on the ordinate as a function of recognition rating (1 = *certain new*, 6 = *certain old*) on the abscissa. Data are based on RTs to target 3 of each sequence in the group that received test sequences of length 3 and on mean RTs across targets 3–6 of each sequence in the group that received 6-element test sequences. Targets 1–2 were unpredictable in both old and new sequences. The critical result is that old sequences were executed faster than new ones even when they received the same recognition rating.

direct (recognition) and indirect (priming) measures were dissociable: even when old and new test sequences received identical recognition ratings, old sequences were nevertheless executed more rapidly than new ones. This effect was quite consistent across participants: averaging across recognition ratings, 47/69 (68%) of participants presented with 3-item sequences and 63/79 (80%) of those presented with 6-item ones had shorter response latencies overall to old than new sequences which they did not discriminate in recognition.[3] These results therefore demonstrate response priming of old relative to new sequences which is not accompanied by differential recognition. Indeed, the priming effect was significant with 3-item test sequences even for sequences which were not recognized at all (i.e., which received a rating of 1).

This is the first demonstration of implicit priming of a sequentially-structured response chain under conditions in which priming and recognition are assessed concurrently. Previous research (Cleeremans & McClelland 1991; Destrebecqz & Cleeremans 2001; Frensch et al. 1998; Honda et al. 1998; Jiménez, Méndez, & Cleeremans 1996; Nissen & Bullemer 1987; Perruchet &

Amorim 1992; Perruchet et al. 1997; Reber & Squire 1998; Reed & Johnson 1994; Stadler 1995; Willingham et al. 1993) has not allied contiguous measurement on direct and indirect tests with the analytic procedure of examining performance on the indirect test at different levels of performance on the direct test, which means that fine-grained comparison of measures across test items has not been possible. Shanks and Perruchet's study shows that such fine-grained comparison is essential as global measures of priming and recognition were strongly associated (i.e., both priming and recognition were significantly above-chance).

One attractive conclusion from these results is that there exists a form of learning (i.e., implicit) which is independent of explicit learning and which can occur in the absence of awareness (i.e., recognition). An alternative possibility, however, is that a single knowledge source underlies performance on both types of test, with subtle differences between the retrieval processes recruited by the tests being responsible for the observed dissociations (Kinder & Shanks 2001; Nosofsky & Zaki 1998). I next present such a model which reveals that priming can be dissociated from recognition in the manner found in the present results even if the two measures depend on the same underlying memory variable.

## 2.1 A model of priming and recognition

This model (Shanks & Perruchet 2002), which is conceptually very similar to standard signal detection theory models for recognition judgments and their latencies (Pike 1973; Ratcliff & Murdock 1976), starts with the simple assumption that new and old test items are associated with a memory strength variable which we will call familiarity $f$. Greater degrees of familiarity lead to higher recognition judgments and faster RTs, and familiarity can be thought of as some composite but unidimensional function of the perceptual familiarity of the stimulus sequence and the motor fluency of the executed response sequence. In the model $f$ is a uniformly distributed random variable in the interval [0,.8] for new items and in the interval [.2,1] for old items. Thus the mean familiarity of old items, $f_{\text{old}}$, is slightly higher (by .2) than the mean for new items, $f_{\text{new}}$. For each participant a single value of familiarity is independently sampled for new and old items from these distributions. Next, we assume that RT is a decreasing function of $f$ but with the addition of some random error:

$$RT_{\text{old}} = 200 + 100(1 - f_{\text{old}}) + 300e \tag{1}$$

$$RT_{\text{new}} = 200 + 100(1 - f_{\text{new}}) + 300e \tag{2}$$

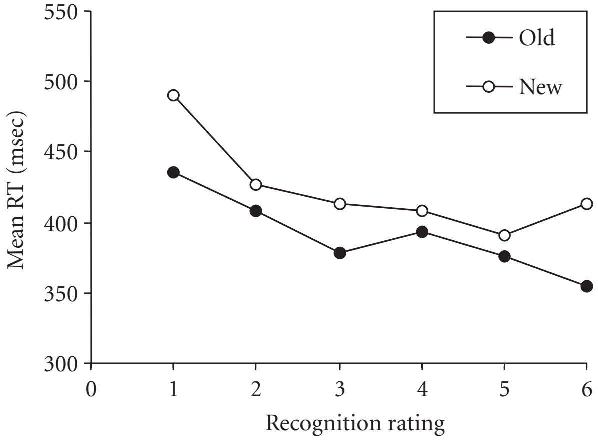


**Figure 9.** Mean simulated RTs (msec) to targets in old and new test sequences as a function of recognition rating. From Shanks and Perruchet (2002).

where $e$ is uniformly distributed random error in the interval [0,1]. The numbers in these equations are simply chosen to ensure that RTs are generated between a maximum of 600 msec when the familiarity of the test item is zero and a minimum of 200 msec when familiarity is 1. These correspond roughly to observed response times.

Recognition judgments (J) are also based on familiarity, but include another (independent) source of error:

$$J_{old} = 2f_{old} + 3e + 1 \tag{3}$$

$$J_{new} = 2f_{new} + 3e + 1 \tag{4}$$

where $e$ is again uniformly distributed random error in the interval [0,1]. J is rounded to the nearest integer value. These equations generate recognition ratings between a maximum of 6 when the familiarity of the test item approaches 1 and a minimum of 1 when familiarity is 0.

Despite the fact that RTs and recognition judgments depend on the same variable ($f$) in this model, and depend on nothing else apart from noise, the model nevertheless generates a pattern of data strikingly similar to that shown in Fig. 8. Figure 9 presents the mean RTs to old and new items at each recognition judgment based on 1000 simulated subjects. RTs are faster to old than new items simply as an automatic by-product of the fact that the two measures are affected by the random variation and measurement error that plague any experimental measure. More specifically, for old and new sequences to be rated equally in recognition a larger value of $e$ in Eqn. 4 compared to Eqn. 3 is necessary, on average, to offset the larger average value of $f_{\text{old}}$ compared to $f_{\text{new}}$. However, when these same $f$ values are used to determine RTs in Eqns. 1 and 2 they will be combined with independently-generated values of $e$. Since the latter are uncorrelated with the $e$ values incorporated in the recognition judgments, on average they will not differ for old and new items. Hence, as $f_{\text{old}}$ is on average greater than $f_{\text{new}}$, $\text{RT}_{\text{old}}$ will be lower than $\text{RT}_{\text{new}}$, as observed empirically in the participants' behavior.

When error is not included in the model the old-new difference is zero. This confirms that it is the imperfect relationship between measures of priming and recognition which is responsible for the effect.

### 2.2 Summary

The present section supports an empirical conclusion and a theoretical one. Empirically, previous research has failed to demonstrate convincingly that above-chance sequence knowledge can be accompanied by null awareness when the latter is indexed by objective measures such as recognition. However, Shanks and Perruchet showed that practiced sequences of responses are executed faster than unpracticed ones even when the sequences are given identical recognition ratings. This is a clear confirmation that priming and recognition can be dissociated. But the theoretical conclusion is that this dissociation is to be expected from any pair of measures which are less than perfectly correlated and is not inconsistent with a model in which priming and recognition depend on the same underlying memory structure.

## 3. Concluding comments

It is not clear that either of the claims about implicit sequence learning described at the outset of this chapter has been established. If the goal is to demonstrate the existence of a form of learning which is both functionally and neurally separate from explicit learning then I would argue that such a goal has not yet been achieved. But perhaps a more important message is that researchers may have been misguided in devoting so much effort to demonstrating dissociations. Whether it be the dissociation of implicit learning from attention or from explicit knowledge that is the object of study, the problem arises that dissociations provide only the weakest constraint on cognitive structure and process. As the model I have described demonstrates, it is not difficult to generate dissociations from single-system theories. Perhaps a better goal of implicit learning research is to try to develop more adequate computational models of behavior in so-called "implicit" learning tasks.

## Notes

* David R. Shanks, Department of Psychology, University College London, London, England. The research described here was supported by grants from the Leverhulme Trust, the Economic and Social Research Council, and the Biotechnology and Biological Sciences Research Council. The work is part of the programme of the ESRC Centre for Economic Learning and Social Evolution, University College London. I thank Andrea Leigh, Simon Li, and Sophie Wood for their assistance in collecting data and Ben Newell and Richard Tunney for their helpful comments.

1. Our participants were to some extent aware of the training sequence in that they were able to perform above chance in a free generation test (see below). One might therefore object that our results do not speak to the issue of whether *implicit* sequence learning requires attention. For learning to be implicit, surely participants must *not* be in possession of the sort of sequence knowledge we detected in free generation? In fact, Shanks and Channon (2002) showed that the dual/single learning difference was independent of explicit knowledge in that even participants who performed poorly in free generation possessed more sequence knowledge if they had been trained under single- than dual-task conditions. Similarly, in the other experiments reported in this section participants were able to recognize the training sequence (as described in the next section). But again, the critical single-/dual-task differences were not related to this recognition ability.

2. One might raise concerns over these findings, however. For example, it seems quite likely that the exclusion task is cognitively very demanding. How can one rule out the possibility that some participants ignore the instructions and give up trying to exclude known sequence continuations?

3. The old/new difference was largest at the end-points of the rating scale with 3-item sequences and around the mid-point of the scale with 6-item ones. Whether anything should be read into this pattern must remain a question for future studies. The model presented below sometimes yields one pattern and sometimes the other, but these variations are merely due to sampling error.

## References

Brooks, L. R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology, 19*, 289–299.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review, 92*, 414–423.

Cleeremans, A. (1997). Sequence learning in a dual-stimulus setting. *Psychological Research, 60*, 72–86.

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* (pp. 323–364). Thousand Oaks, CA: Sage.

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

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

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience, 9*, 522–533.

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

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

Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of Implicit Learning* (pp. 47–104). Thousand Oaks, CA: Sage.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological Research, 61*, 83–98.

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition, 22*, 95–110.

Frensch, P. A., Wenke, D., & Rünger, 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.

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

Heuer, H., & Schmidtke, V. (1996). Secondary-task effects on sequence learning. *Psychological Research, 59*, 119–133.

Honda, M., Deiber, M.-P., Ibáñez, V., Pascual-Leone, A., Zhuang, P., & Hallett, M. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning: A PET study. *Brain, 121*, 2159–2173.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122*, 139–154.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 236–259.

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 948–969.

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 497–509.

Kinder, A., & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience, 13*, 648–669.

Marescaux, P.-J. (1997). Can dynamic control task knowledge be communicated? *Psychologica Belgica, 37*, 51–68.

McCann, R. S., Remington, R. W., & Van Selst, M. (2000). A dual-task investigation of automaticity in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1352–1370.

McDowall, J., Lustig, A., & Parkin, G. (1995). Indirect learning of event sequences: The effects of divided attention and stimulus continuity. *Canadian Journal of Experimental Psychology, 49*, 415–435.

Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition, 79*, 115–134.

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

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247–255.

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.

Perruchet, P., Bigand, E., & Benoit-Gonin, F. (1997). The emergence of explicit knowledge during the early phase of learning in sequential reaction time tasks. *Psychological Research, 60*, 4–13.

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., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology, 22*, 493–516.

Pike, R. (1973). Response latency models for signal detection. *Psychological Review, 80*, 53–68.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190–214.

Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience, 10*, 248–263.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 585–594.

Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind and Language, 5*, 9–28.

Schmidt, P. A., & Dark, V. J. (1998). Attentional processing of "unattended" flankers: Evidence for a failure of selective attention. *Perception & Psychophysics, 60*, 227–238.

Schmidtke, V., & Heuer, H. (1997). Task integration as a factor in secondary-task effects on sequence learning. *Psychological Research, 60*, 53–71.

Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological Research, 61*, 175–190.

Seger, C. A. (1997). Two forms of sequential implicit learning. *Consciousness and Cognition, 6*, 108–131.

Shanks, D. R., & Channon, S. (2002). Effects of a secondary task on "implicit" sequence learning: Learning or performance? *Psychological Research, 66*, 99–109.

Shanks, D. R., & Johnstone, T. (1998). Implicit knowledge in sequential learning tasks. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learning* (pp. 533–572). Thousand Oaks, CA: Sage.

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1435–1451.

Shanks, D. R., & Perruchet, P. (2002). Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomic Bulletin & Review, 9*, 362–367.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367–447.

Shanks, D. R., Wilkinson, L., & Channon, S. (2002). Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Manuscript submitted for publication.*

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

Styles, E. A. (1997). *The Psychology of Attention.* Hove, England: Psychology Press.

Willingham, D. B., Greeley, T., & Bardone, A. M. (1993). Dissociation in a serial response time task using a recognition measure: Comment on Perruchet and Amorim (1992). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1424–1430.

Willingham, D. B., Greenberg, A. R., & Thomas, R. C. (1997). Response-to-stimulus interval does not affect implicit motor sequence learning, but does affect performance. *Memory & Cognition, 25*, 534–542.

# Intention, attention, and consciousness in probabilistic sequence learning

Luis Jiménez[*]
University of Santiago

## 1. Introduction

María is a student of Psychology at the University of Santiago. One Monday, she enters the laboratory and is instructed to perform two different tasks simultaneously. First, she is asked to respond as fast as possible on each trial by pressing a key that corresponds to the current location of a stimulus, and second, she simultaneously needs to keep a running count of the number of times that this target stimulus has the shape of either an "x" or an "*". By the following Friday, after about eight hours and several thousands of trials of practice with these two tasks, she has learned to respond faster and more accurately to the locations that are statistically more likely to appear in the context defined by the previous locations, and to respond more efficiently to those locations that are predictable by relying on the current response in the counting task. At that point, however, she still doesn't believe that there is any predictive relation between shapes and locations, and neither is she able to use her knowledge about the sequence of locations to generate the next one when she is directly told to do so, and under conditions that, otherwise, resemble those of training with the Serial Reaction-Time (SRT) task.

This specific pattern of results has been reported repeatedly (e.g., Jiménez & Méndez 1999, 2001), and raises a host of questions of both empirical and theoretical interest on the relations between learning, intention, attention, and consciousness. To wit: Was María aware of the fact that there existed a structure in the series of locations? Was she aware of the specific contingencies she had been learning, or even of the fact that she has been learning? Was she deliberately trying to learn this structure, or would her learning be larger if she had been instructed to do so? Would learning be either larger or better expressed if she had been trained on the SRT task alone? If the shapes were not relevant for the counting task, would she still have learned about them? Reciprocally, if she had been informed about the existence of reliable predictive contingencies between shapes and locations, would she still have learned the redundant information provided by the sequence of locations? Finally, if she had been explicitly informed about the predictive value of the shapes, but had to keep performing the shape-counting task, how could this complex setting have affected the expression of shape learning, as well as the implicit acquisition of knowledge about the sequence of locations?

The main purpose of this chapter is to summarize the results of a number of experiments conducted with the probabilistic sequence learning paradigm, a paradigm that I believe has the potential to provide satisfactory answers to all these questions. Most of these results have been reported elsewhere, but their joint review here will allow to draw a broader picture of some of their theoretical implications. I will devote the next section to highlight the advantages of this probabilistic sequence learning paradigm, and then I will use the three following sections to review some results that are relevant to questions raised above. In Section 3, I discuss whether or not the intention to learn and the intention to use what has been learned plays a relevant role in the acquisition and in the expression of probabilistic sequence learning. In Section 4, I review the debate on the role of attention in implicit sequence learning, considering both the selective and resource meanings of attention. In Section 5, finally, I close with a discussion of what I take to be a reasonable, although admittedly speculative, framework within which to think about the relations that may hold between learning and consciousness. By the end of the chapter, I hope to have convinced the reader that implicit sequence learning is just learning in its most elementary guise, that it can be conceived as an automatic side-effect of processing (Barsalou 1995), and that it may shape consciousness but is only indirectly caused by it. Implicit learning, therefore, is viewed as the obligatory product of attending to any set of structured events, and is caused not directly by the learners' conscious intention to learn but, rather, by the way in which their perceptual skills, attentional priorities, and motivational states, affect the effective encoding and perception of the information provided by the environment (see Logan 1998; Logan & Etherton 1994; Logan, Taylor, & Etherton 1996; 1999 for a similar perspective).

## 2. The paradigm of probabilistic sequence learning

Implicit learning has been investigated with many different procedures (see Seger 1994, for an exhaustive list). However, during the last fifteen years, the sequence learning paradigm, first devised by Nissen and Bullemer (1987), has become dominant. The SRT task has been adapted in a number of ways and is particularly well-suited to explore many of the issues of interest in the area of implicit learning. Some of the most recent summaries of research conducted with this paradigm can be sampled in this volume, and may be found, for instance, in Buchner and Frensch (2000), Buchner, Steffens, and Rothkegel (1998), Destrebecqz and Cleeremans (2001), Hoffmann, Sebald, and Stöcker (2001), Jiménez and Méndez (1999, 2001), Koch and Hoffmann (2000), Meulemans, Van der Linden, and Perruchet (1998), P. J. Reber and Squire (1998), Remillard and Clark (2001), Schvaneveldt and Gomez (1998), Willingham (1999), Willingham, Wells, Farrell, and Stemwedel (2000), Ziessler (1998), or Ziessler and Nattkemper (2001).

The description of María's task, presented in the Introduction to this chapter, may be taken as a rough description of this paradigm. Thus, participants in these experiments are told to respond as fast as possible to the location of a stimulus that appears on each trial at one of several possible locations on a computer screen (typically, between 3 and 6 locations). The series of locations follows a regularity that is often repeated over many cycles; participants are usually not informed about the existence of the pattern. Indeed, the lack of information about the learning situation can be construed as one of the main strengths of the paradigm as compared, for instance, with those of the grammar learning paradigm (A. Reber 1967; A. Reber & Allen 1978) or the dynamic systems' control paradigm (e.g., Berry & Broadbent 1984, 1988). The sequence learning paradigm is thus unique in allowing the experimenter not only to design truly incidental learning conditions, but also to assess the amount of learning without revealing the existence of any learnable structure, by comparing responses to structured and random trials.

Furthermore, the use of probabilistic sequential structures provides the researcher with some additional advantages over the original version designed by Nissen and Bullemer (1987). Conceptually, the addition of noise into a sequential pattern can be taken as a step toward the goal of designing more realistic replications of what occurs outside the laboratory when one acquires a perceptual and motor skill. Indeed, under natural conditions, skills are not typically acquired through a single session of practice with a fixed and repetitive pattern, but they accrue rather gradually as a result of a great amount of practice with materials that are only loosely structured, and in which the existing regularities are continuously interspersed with random noise.

Methodologically, the use of probabilistic sequences also provides the experimenter with some additional advantages. First, the fact that these structures incorporate stimuli that either conform or do not conform to the target sequence allows the experimenter to administer large periods of practice while minimizing the risk of having participants discover and memorize the sequence. Thus, even in the case where participants might be led to believe that there exist regularities in the observed sequence, this belief would not lead participants to systematically base their performance on such knowledge, given that any hypothesized rule would be often falsified by the continuous appearance of exceptions. Furthermore, the fact that structured and random events are not presented in blocks but rather appear continuously over training, allows experimenters to assess learning online, blurring the distinction between training and test phases that has raised a number of concerns for the deterministic versions of this paradigm (e.g., Shanks & St. John 1994).

Most of the studies conducted with the probabilistic sequence learning paradigm have adopted a procedure developed by Cleeremans and McClelland (1991), in which the structured locations are generated by following a finite-state grammar, and in which a certain proportion of random trials is interspersed within the structured trials. Figure 1 shows a grammar used by Jiménez and Méndez (1999, 2001) that presents a number of desirable properties to be highlighted below.

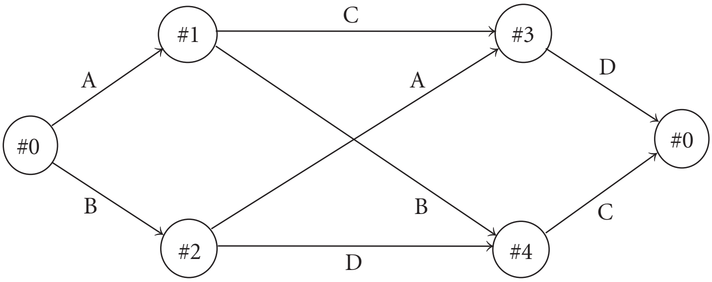


**Figure 1.** Finite state grammar used to generate the series of locations in Jiménez and Méndez (1999, 2001). See text for details.

## 2.1 A second-order conditional grammar

In these experiments, each block of trials starts with five random locations that serve as buffer trials, and then continues by generating subsequent locations as determined by the grammar. The current node of the grammar is set to be node #0 when the buffer trials have been completed, and an arc coming out of this node is selected at random. The label corresponding to this arc (i.e., either A or B) is used to determine the location of the stimulus that is assigned to the selected label in 80% of the trials. This location is replaced by a different one at chance in the remaining 20% of the trials. The current node is then updated to be the node pointed to by the selected arc (i.e., #1 or #2), and the procedure continues indefinitely by selecting another arc at random. As shown in Figure 1, the grammar is re-entrant, which means that the first and the last nodes are identical, so as to allow the generation of an indefinite number of grammatical labels. Moreover, each possible label (A, B, C, or D) appears in two different arcs pointing to different nodes in each case, so that any label predicts every other label with equal likelihood (see Jiménez & Méndez 1999, for a statistical analysis of the constraints imposed by this grammar). However, considering two consecutive labels allows the learner to discriminate between legal and illegal transitions, and thus makes this structure a probabilistic structure that is analogous to what Cohen, Ivry, and Keele (1990) called an "ambiguous" sequence. The information provided by farther elements (i.e., higher-order probabilities, see Remillard & Clark 2001) is redundant with respect to the second-order transition probabilities, and hence the grammar is specially adequate to analyze learning of second-order conditional information. Learning can be assessed by analyzing whether, with practice, participants come to respond more efficiently to a given successor (e.g., D) in terms of whether it appears following a path that it can or cannot legally follow (e.g., ACD vs. BCD). The conditional probabilities of appearance for legal successors range from .34 to .62, whereas the conditional probabilities for the illegal successors vary from .13 to .17.

Although most of the studies conducted with the paradigm have adopted this type of grammar to generate the structured events, there exist a few other studies that have followed the simpler strategy of adding a certain amount of noise to an otherwise deterministic series of locations (e.g., Cleeremans & Jiménez 1998; Schvaneveldt & Gomez 1998). As we will see below, this procedure may be specially useful to analyze the impact of intentional factors on the acquisition of deterministic and probabilistic sequences.

## 3. Intention to learn and to use what has been learned

Did María try to learn? Did she intentionally use what she had learned about the sequential structure to respond to the SRT task? In any case, could the intention to learn make any difference in sequence learning when the structure is complex and probabilistic? In the following paragraphs, I review the evidence that bears on these three issues, as it arises from studies that have used both the deterministic and the probabilistic versions of the paradigm, and which have generated the probabilistic structures by using either a grammar or a noisy series.

### 3.1 Intention to learn

Jiménez, Méndez, and Cleeremans (1996) analyzed the effects of the intention to learn on the acquisition of knowledge about a probabilistic sequence by confronting two groups of participants with 20 sessions of practice with a six-choice SRT task. The series of locations was structured according to a grammar that contained first-, second-, and third-order information. Half of the participants in this study were instructed to look for the underlying rules, and the other half were presented with the SRT task under standard incidental learning conditions. To increase the motivation to search for contingencies, all participants were paid depending on their performance. Participants assigned to the intentional condition were (1) reminded of the search instructions during each rest break, (2) informed that they would later be asked to predict each location in the context of a generation task, and (3) told that generation accuracy would be used as a factor to multiply the total earnings obtained during the previous SRT task. Despite of different instructions, both groups performed in much the same way throughout training, and they showed an equal amount of learning after a total of 20 sessions. Figure 2 shows the mean RTs of incidental and intentional learners, for different training periods, and separately for trials that were either predictable or not predictable by relying on first- and second-order information (no learning was observed concerning third-order conditionals). The results indicate that, at least in the complex and probabilistic setting of the experiment, coping with the requirements of the SRT task was just as useful for the acquistion of knowledge about the sequence as was trying to discover the underlying structure. The fact that the intentional learners produced consistently slower responses supports the claim that they used a different search strategy throughout training, and that they did not give up the search for rules or started to behave like incidental learners. Moreover, the fact that a significant

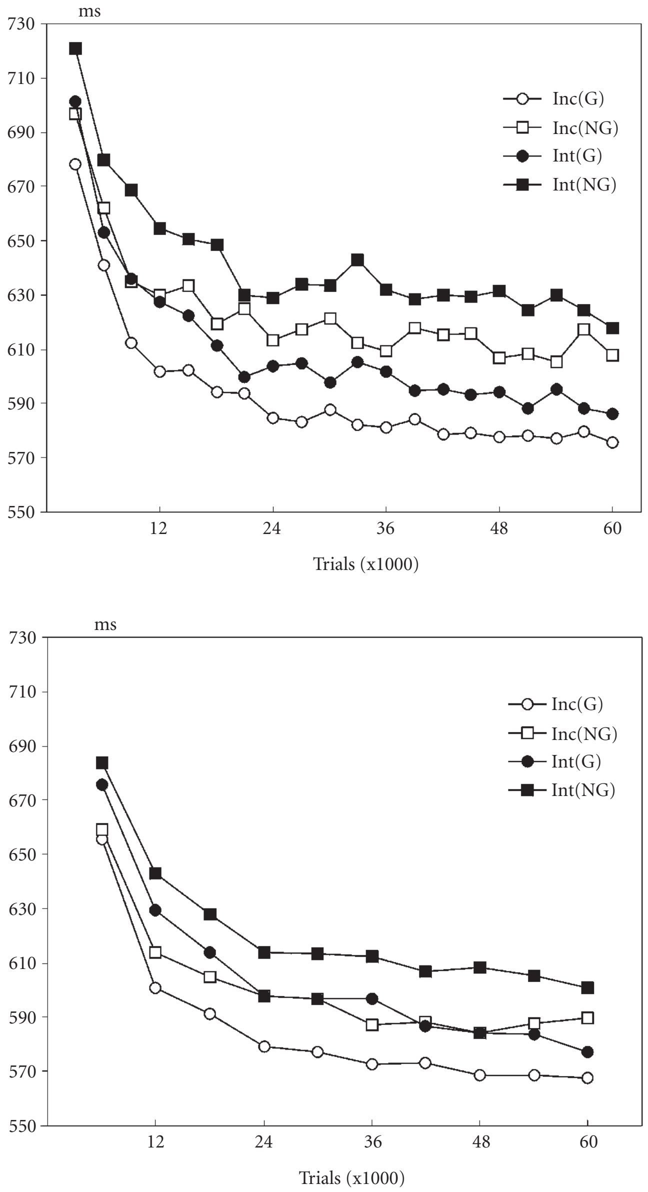


**Figure 2.** Reaction time performance averaged for each session (top panel) or for every two sessions (bottom panel) in response to grammatical (G) and non-grammatical (NG) trials, and plotted separately for incidental (Inc) and intentional (Int) conditions from Jiménez et al. (1996). Top panel represents trials in which the grammatical or non-grammatical status of a successor depended on first-order information. Bottom panel represents trials in which grammaticality depended on second-order information.

part of the knowledge obtained in both conditions was expressed exclusively in the indirect measures, and that this knowledge was not manifested in the generation task, led the authors to conclude that the learning was at least partially unconscious, and that the equal amount of learning observed under incidental and intentional conditions cannot be attributed to the fact that all learning was consciously acquired regardless of the learners' initial orientation (see Jiménez et al. 1996; and Reingold & Merikle 1988; for discussions of the assumptions underlying this conclusion). Hence, at least when the sequential structure is complex enough, participants seem to be able to learn it in much the same way, regardless of whether they are instructed to look for a sequence, or just told to respond to each stimulus in the context of the SRT task.

### 3.2 Intention to use what has been learned

Research participants learn a complex sequence in about the same way, regardless of whether they are instructed to look for contingencies or just to respond to each stimulus in the appropriate way. Now, what if participants were not only provided with search instructions, but also explicitly informed about the details of the underlying structure? Could such knowledge, if conveyed explicitly, be used immediately by participants to improve performance on the SRT task, or would the effects of the knowledge still depend on factors such as the complexity of the structure, or the presence of supplementary attentional demands?

Curran and Keele (1993) presented some results indicating that participants that were given explicit information about a deterministic sequence showed larger effects of learning than did participants who performed the SRT task under incidental conditions, but only if the task was performed under conditions of no distraction. On the contrary, if a secondary tone-counting task was subsequently added to the SRT task, the difference between intentional and incidental learners completely disappeared, and only a reduced effect of sequence learning remained. The results indicate that the use of explicit knowledge did require a continuous effort to monitor the sequence, and that the monitoring process could no longer proceed in the presence of the secondary task.

Cleeremans and Jiménez (1998), on the other hand, reported a study aimed at comparing the effects of providing explicit information, and of performing the SRT task under single- or dual-task conditions, on the expression of knowledge about either deterministic or probabilistic sequences of different complexity. We used both unique sequences (i.e., sequences in which each element was fully predictable on the basis of the previous element) and ambiguous sequences (i.e., sequences in which each element could only be predicted based on the conjunction of the two previous elements), and generated the probabilistic sequences by replacing 20% of the elements stipulated by the deterministic sequences at chance. Different groups of participants were trained with different types of sequences (unique or ambiguous), under different conditions of noise (deterministic vs. probabilistic generation), different task requirements (single vs. dual task), and different instructions (explicit vs. implicit). The results of this multifactorial study indicate that explicitly acquired information was only effective when participants responded to deterministic sequences under single-task conditions, but that it was quite ineffective, if not harmful, for any other combination of noise and distraction (see Figure 3). This pattern of results could easily be interpreted as suggesting that the information available to consciousness cannot be translated into performance automatically, but that it requires both a deliberate decision to do so (which might be suspended whenever the information is not completely reliable), and enough resources to monitor the series, retrieve the relevant explicit knowledge, and translate the knowledge into an actual anticipation of the following response. In the absence of any of these conditions, participants appear to resort to incidental processing, and consequently, the effects of learning are roughly equivalent to the effects obtained under incidental conditions. Hence, according to this interpretation, explicit orientation does not affect performance when participants respond to a probabilistic sequence, not only because they do not seem to learn more than participants who perform the task under incidental conditions, but also because any explicit knowledge that they might have acquired would be of little use when they have to respond to a noisy sequence.

### 3.3 Implicit sequence learning in the presence of explicit cues

The conclusion that the intention to learn does not affect the effects of sequence learning when the structure is noisy and probabilistic is supported by another source of evidence: Participants in the SRT task can learn a complex sequence of locations even when there exists an explicit and valid cue that allows the learner to anticipate the next stimulus location, and that arguably removes any possible motivation for the participant to engage in an explicit search for sequential contingencies (Cleeremans 1997; Jiménez & Méndez 2001). Cleeremans, for instance, showed that a complex sequence of dot locations could be learned when participants were trained in the presence of an additional cue

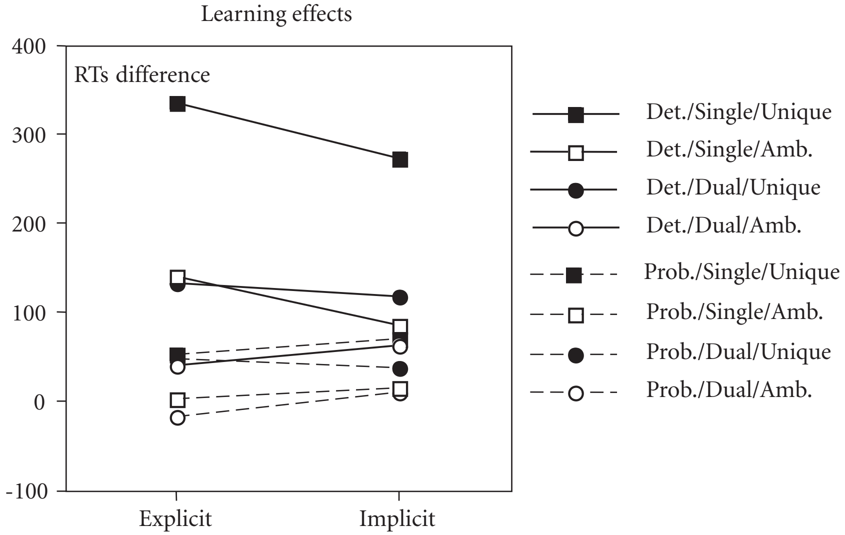


**Figure 3.** Reaction time differences between responding to predictable and non predictable trials, plotted separately for type of structure (deterministic vs. probabilistic), distraction (single vs. dual task), and type of sequence (unique vs. ambiguous) for the Explicit and Implicit training conditions from Cleeremans and Jiménez (1998).

(i.e., a cross) which, on each trial, appeared at the specific location where the next dot was going to be presented. In Experiment 2, the cross served as an exogenous and completely valid cue that allowed participants to anticipate the next target location with total accuracy. As a result, no effect of sequence learning was expressed during the training phase. However, even under such highly redundant conditions, the sequence of locations was learned, and this learning was expressed during a transfer phase in which the additional cue was removed.

Along the same lines, Jiménez and Méndez (2001) showed that whenever the information provided by the additional cue required an endogenous elaboration process, learning of the sequence of locations was not only produced, but also expressed together with knowledge about the additional explicit cue, without any interference between the two sources of knowledge. In this particular case, the location of the next stimulus was not indicated directly by the location of an additional cue, but could easily be inferred from the shape of the current stimulus. For instance, participants might be informed that the shape “x” predicted the next stimulus to appear in the leftmost location, that the shape "*" predicted the next stimulus to appear in the rightmost location, and so on. The cue-target contingencies were completely reliable during the first four sessions, and they were fulfilled in 80% of the trials during a fifth test session. An analysis of response times for the test session shows that informed participants that performed the SRT task under single-task conditions were effectively following the rules about the relationship between cues and targets (Experiment 2, see Figure 4). When the entire training period is taken into account, RT performance indicates that participants also learned the sequence of locations. Moreover, sequence learning observed in this condition was not

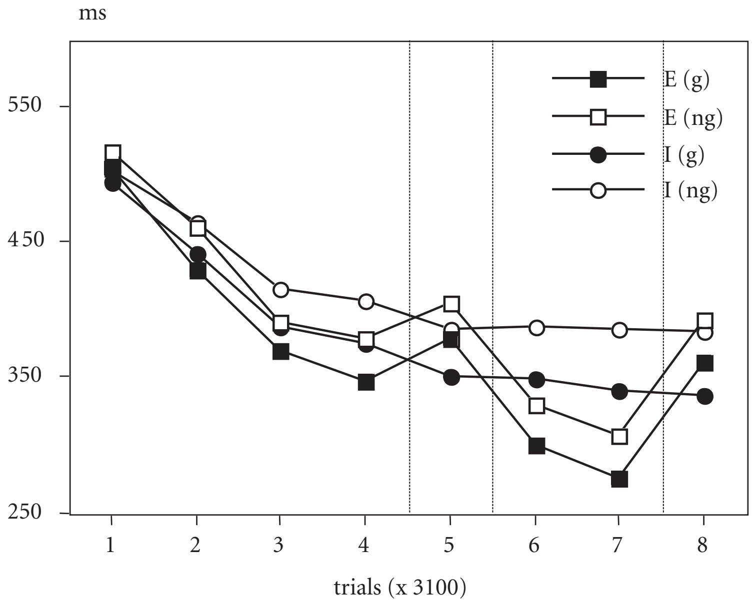


**Figure 4.** Reaction time performance for each of the experimental sessions plotted separately for grammatical (g) and non-grammatical (ng) locations, and for the conditions of explicit (E) and implicit (I) shape-learning, from the Experiment 2 of Jiménez and Méndez (2001). Note that participants in the explicit shape-learning condition produced faster responses overall, but they produced slower responses when the validity of the shapes decreased from 1 to .80, (session 5), and when the shapes were completely removed in session 8. The results indicate that participants were using explicit knowledge about the shapes to predict the next locations, but they were still able to learn about the sequence of locations to about the same extent like participants in the implicit condition.

significantly different from that observed in conditions in which the participants remained ignorant about the predictive value of the shapes, or in which they had to perform the SRT task together with a secondary task that required them to keep a running count of a pair of target shapes.Therefore, it seems reasonable to conclude that in all conditions, sequence learning proceeded in an incidental way, and was automatically translated into performance, regardless of the presence or absence of explicit contingencies, and just as a side-effect of participants' consistent attending, and responding, to each successive location.

## 4. Attentional resources and selective attention

Implicit learning has been defined as learning that occurs regardless of the learners' intention to learn, and largely in the absence of explicit knowledge of what has been acquired (Reber 1993). From this definition, it was often suggested that implicit learning ought to be accomplished through completely automatic learning mechanisms and, hence, that it ought not to be affected by attentional manipulations. It is therefore not difficult to understand why questions concerning the relation between attention and implicit learning have provoked a great deal of research, and that they still belong to the most strongly debated issues in the literature today.

The role of attention in implicit learning has been frequently assessed by relying on a dual-task procedure (e.g., Dienes, Altman, Kwan, & Goode 1995; Hayes & Broadbent 1988; Nissen & Bullemer 1987). More specifically, in sequence learning studies, the dual-task procedure has been realized by including a secondary task that typically requires participants to keep a running count of the number of trials in which an arbitrary event (e.g., a given target tone) appears in the context of the SRT task. The rationale underlying this procedure is that the secondary task is hypothesized to exhaust a proportion of the attentional resources available, and that it ought to interfere with the effects of learning if either the acquisition, the retrieval, or the use of the relevant sequence information depends on the integrity of the same attentional resources (Cleeremans & Jiménez 1998; Cohen, Ivry, & Keele 1990; Frensch, Buchner, & Lin 1994; Frensch, Lin, & Buchner 1998; Nissen & Bullemer 1987; Reed & Johnson 1994; Shanks & Johnstone 1998; Stadler 1995; Willingham, Greenberg, & Thomas 1997). In the following paragraphs, I compare the results obtained with the SRT task depending on whether the sequence is fixed or probabilistic, and I defend the thesis that implicit sequence learning can occur without recruiting specific attentional resources, but that it requires participants to pay selective attention to the predictive cues.

### 4.1 Attentional resources

The results of the studies that have employed a secondary counting task in the context of an SRT task with a deterministic sequence of locations largely indicate that performing a counting task does indeed interfere with sequence learning expressed through speeded performance (e.g., Cohen, Ivry, & Keele 1990; Nissen & Bullemer 1987). However, it is not clear at all whether this pattern of results must be taken to indicate that implicit sequence learning really depends on the availability of certain amounts of general processing resources. On the contrary, interference may arise (1) as an expression deficit that hinders the retrieval or the use of sequence knowledge, rather than its implicit acquisition (Frensch 1998; Frensch et al. 1998; but see also Shanks, this volume), (2) as interference caused by the introduction of disrupting stimuli, and not by the scarcity of attentional resources (Stadler 1995), and/or (3) as interference produced over the acquisition or the expression of *explicit* knowledge, that occurs together with implicit effects, specially when the sequence is deterministic, and when enough attentional resources are available to allow for an intentional search and monitoring of the learned regularities.

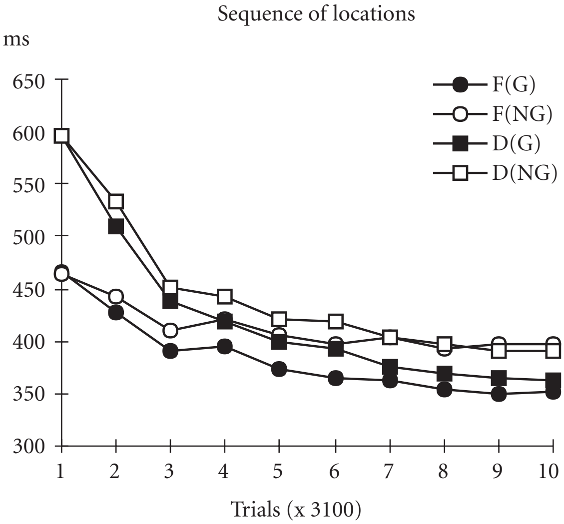
In support of the latter alternative, Jiménez and Méndez (1999, 2001) have shown repeatedly that neither the acquisition nor the expression of sequence learning are affected by the presence of a secondary task when the secondary task is conducted on the same visual stimuli on which the SRT task is carried out – thus avoiding the introduction of disrupting stimuli – and when the sequential structure is generated according to a noisy, finite-state grammar (see Figure 5, top panel). Still more to the point, Cleeremans and Jiménez (1998) also showed that transfer from single- to dual-task conditions hindered the expression of sequence learning exclusively when participants were trained with deterministic sequences, but not when they were trained with probabilistic, but otherwise similar, structures.

As a whole then, and against what has been usually observed with deterministic sequences, our studies with probabilistic structures strongly indicate that probabilistic sequence learning may not depend on the availability of a pool of general resources but, rather, that it proceeds regardless of the presence of a secondary task, provided that processing of the relevant information is granted by the requirements of the SRT task.[1] This conclusion, however, does not necessarily mean that learning can proceed independently of any form of attention, but only that participants will learn about any attribute of the stimuli to which they pay selective attention. Hence, the answer to the question whether attention is necessary for implicit sequence learning may be "no", if we refer to the dependence of learning on any set of processing resources beyond those required to perform the SRT task, but "yes", if we refer to the fact that learners ought to pay selective attention to predictive events, so that they become associated with the predicted targets. This selective meaning of attention has been relatively neglected in the empirical studies that have addressed the relation between attention and implicit learning, but it also deserves a careful analysis.

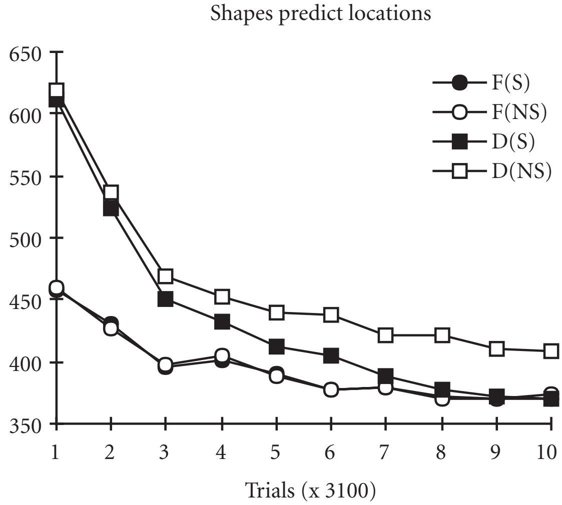
## 4.2 Selective attention

Indeed, from the general framework that I am proposing, it follows rather directly that, if learning is taken as an obligatory side-effect of processing, then selective processing of predictive features of the stimuli can be seen as the main – perhaps, even the only – pre-condition of learning. Gordon Logan and his collaborators (e.g., Logan 1998; see also Boronat & Logan 1997; Logan & Etherton 1994; Logan, Taylor, & Etherton 1996, 1999) have proposed essentially the same idea under the label of the "Attention Hypothesis", according to which attention to an event is necessary and sufficient for the event to be stored in memory, – i.e., *obligatory encoding* –, and for the representations associated with the event to be retrieved from memory – i.e., *obligatory retrieval*. If we add to this hypothesis the classical associative assumption that representations of events that are attended together tend to become associated, then this set of assumptions immediately leads us to predict that the encoding of a sequence of events can only be produced when the relevant events are selectively attended to in close succession. Therefore, even though sequence learning may not depend on the explicit intention to learn, nor on the amount of general attentional resources available, selective attention to the relevant events would be necessary for both the acquisition and the expression of this learning.

This conclusion is consistent with observed difficulties to obtain learning involving dimensions that are not explicitly relevant for the SRT task (Jiménez & Méndez 1999, 2001; Jiménez, Méndez, & Lorda 1993; Willingham, Nissen, & Bullemer 1989; but see Mayr 1996). Jiménez et al. (1993), for instance, found that participants who responded to the location of different patches of color in a four-choice SRT task did not learn a number of simple predictive relations that existed between each color and the following location, whereas they were able to learn a complex sequence involving the series of locations. Along the



Sequence of locations



Shapes predict locations

**Figure 5.** Reaction time performance for each experimental session from Experiment 1 of Jiménez and Méndez (1999). Top panel shows that, even though participants under conditions of divided attention (D) performed more slowly than participants presented with conditions of focussed attention (F), they learned the sequence of locations at about the same rate (i.e., the differences between responses to grammatical (G) and non-grammatical (NG) trials are equivalent). Bottom panel shows that participants presented with conditions of divided attention also learned about the predictive relationships between shapes and locations, as inferred from the differences between responses to signaled (S) and non-signaled (NS) trials, whereas participants in the condition of focussed attention did not learn about the shapes.

same lines, Jiménez and Méndez (1999, 2001) showed that a set of predictive relations between the shape of each stimulus and the location of the next one was not learned at all if participants were exclusively instructed to respond to the locations in the context of a single SRT task, but that the predictive contingencies could be learned implicitly if the learners were forced to pay attention to the shapes as part of a, supposedly independent, secondary task (see Figure 5, bottom panel). Interestingly, if participants were told to classify the four possible shapes ("x", "*", "?", "!") into two categories (i.e., targets and distractors), and to keep an aggregated count of the number of target shapes in each block, their performance showed that they had learned to predict which pair of locations was more likely to appear after either a target or a distractor shape, but not which specific location could be expected to follow each specific shape. Plainly, then, it seems that learning was tightly locked to the specific, task-relevant response required by the attended stimulus, rather than to the identity of the stimulus as a whole.

The results of another experiment from the same series (Jiménez & Méndez 1999; Experiment 3) supports this conclusion. In the follow-up experiment, participants were told to keep a running count of a different pair of target shapes for each successive session, and either the shapes of the stimuli or the response category (i.e., counting vs. non-counting) were systematically related to the location in which the next stimulus would appear. The results of the experiment indicated that learning was produced exclusively when the next location could be predicted on the basis of the previous response in the counting task, but not when it depended on the shape of the previous stimulus. Hence, according to these results, performing a consistent response is necessary to bring about learning, and just paying attention to a dimension (e.g, attending to shapes in order to find out whether or not to count it) is not sufficient to produce the relevant association.

Other studies have found that sequence learning is more narrowly related to the specific responses required by the orienting task than to other salient, but task-irrelevant, stimulus features (e.g., Nattkemper & Prinz 1997; Ziessler 1998; Ziessler & Nattkemper 2001). Still other results indicate that it is possible to learn about a task-irrelevant dimension (e.g., a sequence of locations in the context of a serial object discrimination task), if the procedural details make it necessary for the learners to produce any kind of response to the dimension that is not explicitly relevant. For instance, Willingham et al. (1989) found no effect of learning about a sequence of locations in a color discrimination task in which the target stimuli were presented relatively close to each other, but Mayr (1996) observed spatial learning by requiring a more difficult discrimination of the stimuli, and by increasing the distance between the locations at which the stimuli appeared. According to Mayr's interpretation of the data, the procedural changes may have forced the learners to make orienting responses and eye movements in response to each stimulus, and these minimal, but contingent, responses could have been sufficient to produce learning about the series of locations.

In sum, all of the results presented in this section can be reconciled with a response view of sequence learning, but only if we extend the notion of "response" beyond its narrow motor connotations, to encompass any kind of processing operation performed by the learner (Ziessler 1998). In this general sense, the response view converges with the "episodic-processing" account, which claims that people's encoding of events does heavily depend on how they process them, and that the effects of prior experience on future performance also depend on the similarity between the encoding domain and the transfer domain (e.g., Whittlesea 1997). Crucially, this could be true for both implicit and explicit learning and, hence, the difference between them ought not to be conceived in terms of their relative dependence on selective attention, but rather in terms of the degree of control that the learner exerts to modulate the processes of learning: In the implicit case, learning conditions are narrowly constrained by the environment, or otherwise programmed by the experimenter whereas, in the explicit case, both encoding and retrieval are, at least partially, under the learner's strategic control.

## 5. Consciousness of learning and of the learning results

So far, we have reached the following conclusions concerning María's learning experience: (1) María may have learned the probabilistic sequence of locations without trying to learn, (2) she would not have learned the sequential structure better if she was instructed to look for the regularity, or (3) if she was not instructed to perform any secondary task on the stimulus shapes. (4) If she was not instructed to selectively attend to the stimulus shapes, she would remain insensitive to the fact that the shapes bear predictive information regarding the location of the next stimulus.

Based on these conclusions, we can begin to delineate the process of sequence learning that occurs in such complex settings: Learning does not depend on the intention to learn or directly on the amount of attentional resources available to the learners, but it crucially depends on whether learners selectively attend, or respond in any way, to the relevant stimulus dimensions.

This view of implicit sequence learning is consistent with its characterization as a by-product of processing, but it tells us nothing about one of the most intriguing features of implicit learning: that it proceeds in the absence of consciousness. In what follows, I sketch an admittedly speculative view of what I see as the interactive relation between learning and consciousness. I propose a definition of learning as an adaptive process that continuously shapes both behavior and experience but, at the same time, I defend the existence, and the ecological relevance, of the phenomenon of implicit learning.

Elsewhere, I have argued that a cognitive concept of learning must be distinguished from any other phenomenon of adaptation by assuming that learning selectively occurs in experiencing creatures, and that it continuously changes not only behavior, but also conscious experience (cf. Cleeremans & Jiménez 2002). This assumption is intended to distinguish between the phenomena of cognitive adaptation that we seek to understand, and a host of partially analogous phenomena that are also adaptive, but that, crucially, lack the cognitive status of the learning phenomena we are interested in. Adaptation phenomena that fall beyond our interests may thus range from the processes of natural selection to the workings of our digestive or immune systems, or even to the dynamics of many of the so-called "artificial learning" systems. Hence, the starting point of the proposed framework is that cognitive learning must operate on a special kind of representations (i.e., *cognitive* representations ) that are, at least in principle, accessible to consciousness (cf. Searle 1990). If we do not adopt this cognitive assumption, then it may surely be easier to demonstrate the existence of learning without consciousness, but we will run the risk of confusing the cognitive concept of learning with the much broader concept of adaptation.

Now, if we accept the former definition of learning as those adaptive processes that occur in conscious systems, and that continuously change both behavior and experience, then how are we to understand the standard definition of implicit learning, which takes it to be learning that proceeds without consciousness? In trying to address this question, I have found it useful to consider five different meanings of the expression "learning without consciousness", that differ with respect to the contents of the learning situation that may be unavailable to consciousness while learning is taking place.

First, the most extreme definition of implicit learning is based on the few cases in which the term "consciousness" is used, without a qualifier, to refer to a general property of the learners, rather than to the property of particular representations held by the learners. In the most extreme case, implicit learning refers to whatever learning that happens while the learners are completely unconscious (e.g., during dreamless sleep, or under the effects of general anesthesia). Although a learning situation of this kind has interesting theoretical implications, its ecological value is severely restricted (see Andrade 1995, 2001, for reviews).

In a second sense, learning can be thought of as implicit, if it proceeds without awareness of the underlying learning *processes*. In this category, we can further distinguish two very different cases, depending on one's understanding of the term "processes". On the one hand, if one takes *processes* to refer to whatever microstructural mechanism that is responsible for the production of representational changes, then we may end up defending the position that all learning is implicit, because it is widely acknowledged that we can only gain conscious access to the results of our mental activities, but not to the actual processes that bring about mental states (e.g., Jackendoff 1987). If, on the other hand, we view the contents of consciousness as a succession of goal- and problem-states during strategic decision making, and cognitive operations as processes that bring us progressively closer to the solutions of explicit problems (e.g., Newell & Simon 1972), then we can still talk, in a very realistic sense, about conscious *processes* that, no matter how they are ultimately realized at a microstructural level, can be used as a criterion to distinguish between implicit and explicit learning. In this third sense, indeed, implicit learning is roughly equivalent to non-intentional learning (e.g., Frensch 1998) and, thus, it ought to refer to every learning phenomena that proceed regardless of the agents' strategic decisions and intentions, no matter whether they rely on the same or on different microstructural processes compared to explicit learning episodes.

Finally, there are two more uses of the concept of implicit learning, that refer to the consciousness of learning *results*, rather than to the consciousness of underlying processes. Here, again, we can think of two different conceptual scenarios depending on whether we take implicit learning to be (1) a special case of learning that causes behavioral change without producing any change whatsoever in phenomenal experience, or (2) an ubiquitous form of elementary learning that shapes both behavior and experience, but that is still "implicit" in the sense that the changes are not identified by the learners as a product of learning, that they do not produce the experience of learning, and that they do not justify the adoption of an explicit learning strategy. Under the latter approach, learning can be implicit even if its effects are not reduced to behavioral priming effects, and may include also changes in perceptual experience, in emotional reactions or in attentional functions.

There may be as many definitions of implicit learning as there are researchers in the field, but I think that a definition of this term that combines the criterion of non-intentionality with the production of results that are not consciously attributed to learning, and hence that do not trigger an intentional search, has the potential for providing an ecologically valid, and relatively non-controversial, definition of the empirical phenomenon that we investigate with standardized paradigms. By relying on this notion of implicit learning, there is no need to demonstrate the somewhat counter-intuitive claim that behavioral change can be brought about in the absence of any corresponding change in phenomenal experience, but it is possible to retain the idea that learning can occur in the absence of (1) conscious intention to learn, (2) conscious awareness of the fact that we are learning, and (3) conscious attribution of any noticed change to the effects of learning. Thus, implicit learning can be described as a by-product of processing that can be distinguished from explicit learning on two grounds: It is not caused by conscious intention to learn, and it does not initiate the adoption of an explicit search strategy over the course of learning.

From this perspective, finally, it could be argued that complex learning paradigms such as probabilistic sequence learning provide us with ideal tools to analyze implicit learning, because they allow the experimenter to produce and test learning without making explicit the learning requirements of the task. Moreover, a comparison between similar direct and indirect measures of learning can be seen as a straightforward way to ascertain whether the observed learning is implicit or explicit. The indirect measure is believed to reflect any knowledge that can be brought about by the learners, regardless of whether the knowledge is consciously identified as relevant, whereas a comparable direct measure can be taken to be more reliant on explicit metaknowledge, or on the learner's judgement about the source and relevance of the retrieved knowledge. In accordance with this view, our studies have shown that participants produce faster and more accurate responses to legal successors in a sequence of stimuli, even when they predict legal and illegal successors with equal accuracy in a direct generation task (Jiménez et al. 1996; Jiménez & Méndez 1999, 2001). Again, the pattern of results should not be taken to indicate that the behavioral effects of sequence learning can be dissociated completely from their phenomenal counterparts, but only to suggest that the knowledge acquired produces behavioral effects before these effects are consciously attributed to the results of learning.

On the other hand, this framework also allows us to assume that, with practice in an implicit learning task, the incidentally encountered regularities may give rise to a change of the way in which the task is experienced, and that this change, in turn, may be responsible for the adoption of a search strategy, and may eventually foster the development of an explicit representation of the sequence (Frensch, Haider, Rünger, Neugebauer, Voigt, and Werg, this volume). The crux of the argument, however, is not whether implicit learning paradigms produce explicit outcomes under some conditions, but rather whether there are other conditions that allow us to study implicit learning before it produces such effects.

In sum, the view of implicit learning that I am proposing here accepts that learning can produce a continuous flow of both phenomenal and behavioral changes (Perruchet & Gallego 1997), and that it is often responsible for the adoption of an explicit strategy (Frensch et al., this volume). However, these two facts should not be taken as an argument against the mere existence of implicit learning, but rather as a warning with respect to the necessity to investigate implicit learning with procedures designed to discourage the adoption, and to limit the effects, of explicit strategies. From this perspective, the paradigm of probabilistic sequence learning appears as an ideal tool to fulfil these purposes.

## 6. Concluding remarks

Throughout this chapter, I have elaborated on a number of ideas that are common-sense knowledge in the area of skill learning. I have pointed out that, if María were interested in learning a complex skill, it would not be a good strategy for her to try to memorize the complete set of rules that apply in that domain. I have claimed too that it would not be very useful for her either to try to concentrate on "turning the encoding switches on" or to "completely disregard the stimuli, and to rely passively on unconscious encoding algorithms". Instead, I have argued that María needs to pay attention to the relevant dimensions of the stimuli, and to act upon them in all the appropriate ways and in as many contexts as possible, so that her knowledge and actions become progressively attuned to the task demands. When she is trained in this way, she eventually becomes an expert on the task, and it does not make much of a difference whether she actively tries to learn, or just as actively tries to cope with the task demands.

With a general description of María's task as my starting point, I have proposed a definition of implicit learning that can be identified with most of the processes and effects that arise from her training regime. Hence, implicit learning can be seen as an ubiquitous phenomenon that is intimately related to processing and action, and that changes both the learners' experience and behavior, even though the altered contents of experience are not usually attributed by the learners to the effects of learning. This concept of implicit learning captures most of the learning processes and effects that take place whenever we are faced with a complex structure in everyday life, and it can be analyzed idealiter in the lab by creating a similarly complex learning situation, such as the probabilistic sequence learning paradigm.

By using this experimental paradigm, we have obtained results that are consistent with the following claims: (1) Learning without intention is a ubiquitous phenomenon: Conscious intention to learn improves learning only indirectly – by focusing the learners' attention on the relevant stimulus dimensions, and by guaranteeing that the appropriate processing operations are carried out – but it is neither necessary nor sufficient for learning to take place. (2) Learning, with or without intention, does continuously shape both behavior and experience, but often these changes do not include conscious awareness of the learning itself, or the corresponding attribution of the experiential changes to the effects of learning. This can be taken as an ecologically valid and relatively non-controversial definition of implicit learning. (3) Learning requires the allocation of attention to the relevant dimension by performing a task on the basis of that dimension, but no additional deployment of specific attentional resources is necessary. It is in this sense that implicit learning can be considered an automatic process and, at the same time, a by-product of attention.

## Notes

* This work was supported by Grant XUGA 21106B98 from the Consellería de Educación e Ordenación Universitaria da Xunta de Galicia (Spain) and by the DGES Grant PB97-0525 from the Ministerio de Educación y Cultura (Spain). I wish to thank Peter A. Frensch and Dennis Rünger for their help to improve the English.

1. Shanks (this volume) has argued that, under these conditions, the secondary task may become automated with practice, and hence that it could end up making litle demands on attentional resources. However, this conclusion stands in contradiction with further results reported by Jiménez and Méndez (1999, Exp. 3). In this experiment, a varied mapping was introduced between the identity of each shape and the counting responses, by asking participants to keep a running count of a different pair of target shapes for each consecutive session. Even though this manipulation should have hindered the automatization of the secondary task, sequence learning was still evident, and it was not significantly different from that obtained under single-task conditions.

## References

Andrade, J. (1995). Learning during anaesthesia: A review. *British Journal of Psychology, 86*, 479–506

Andrade, J. (2001). Learning during sedation, anesthesia, and surgery. In M. M. Ghoneim (Ed.), *Awareness during Anesthesia* (pp. 93–102). Butterworth Heinemann: Woburn, MA.

Barsalou, L. (1995). Storage side-effects: Studying processing to understand learning. In A. Ram and D. B. Leake (Eds.), *Goal-Driven Learning* (pp. 407–419). Cambridge, MA: MIT Press.

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

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

Boronat, C. B. & Logan, G. D. (1998). The role of attention in automatization: Does attention operate at encoding, or retrieval, or both? *Memory and Cognition, 25*, 36–46.

Buchner, A. & Frensch, P. A. (2000). How useful are sequence learning tasks? The theoretical status and empirical data of research paradigms. *Psychological Research, 51*, 10–18.

Buchner, A., Steffens, M. C., & Rothkegel, R. (1998). On the role of fragmentary knowledge in a sequence learning task. *Quarterly Journal of Experimental Psychology, 51A*, 251–28.

Cleeremans, A. (1997). Sequence learning in a dual-stimulus setting. *Psychological Research, 60*, 72–86.

Cleeremans, A. & Jiménez, L. (1998). Implicit sequence learning: The truth is in the details. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learning* (p. 323–364). Thousand Oaks, CA: Sage.

Cleeremans, A. & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French and A. Cleeremans (Eds.), *Implicit Learning and Consciousness: An Empirical, Computational and Philosophical Consensus in the Making?* (pp. 1–40). Hove, UK: Psychology Press.

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

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

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*, 163–191.

Cowan, N. (1995). *Attention and Memory.* New York: Oxford University Press

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

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., Altman, G. T. M., Kwan, K. & Goode , A.(1995). Unconscious knowledge of artificial grammar is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1322–1338.

Frensch, P.A. (1998). One concept, multiple meanings: On how to define implicit learning. In M. A. Stadler and P. A. Frensch, (Eds.), *Handbook of Implicit Learning* (pp. 47–104). Thousand Oaks (CA): Sage.

Frensch, P. A., Buchner, A., & Lin J., (1994). Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 567–584.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction time task. *Psychological Research, 61*, 83–98.

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

Hoffmann, J., Sebald, A., & Stöcker, C. (2001). Irrelevant response effects improve serial learning in serial reaction time tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 470–482.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge: MIT Press.

Jiménez, L. & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 236–259.

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

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 948–969.

Jiménez, L., Méndez, C., & Lorda, M. J. (1993). Aprendizaje implícito, atención y conciencia en una tarea de tiempo de reacción serial. [Implicit learning, attention and awareness in a serial, reaction-time task]. *Revista de Psicología General and Aplicada, 46*, 245–255.

Koch, I. & Hoffmann, J. (2000). The role of stimulus-based and response-based spatial information in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 863–882.

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

Logan, G. D. (1998). What is learned during automatization? II. Obligatory encoding of spatial location. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1720–1736.

Logan, G. D. & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1022–1050.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 620–638.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1999). Attention and automaticity: Toward a theoretical integration. *Psychological Researcch, 62*, 165–181.

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 350–364.

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology, 69*, 199–221.

Nattkemper, D. & Prinz, W. (1997). Stimulus and response anticipation in a serial reaction task. *Psychological Research, 60*, 98–112.

Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Englewood Clifts, NJ: Prentice Hall.

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

Perruchet, P. & Gallego, J. (1997). A subjective unit formation account of implicit learning. In D. Berry (Ed.), *How Implicit is Implicit Learning?* (pp.124–161). Oxford: Oxford University Press.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 5*, 855–863.

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

Reber, A. S. (1993). *Implicit learning and Tacit Knowledge*. London: Oxford University Press.

Reber, P. J. & Squire, L. R. (1998) Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience, 10,* 248–263.

Reed, J. & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learnt about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 585–594.

Reingold, E. M. & Merikle, P. (1988). Using direct and indirect measures to study perception without awareness. *Perception and Psychophysics, 44*, 563–575.

Remillard, G. & Clark, J. M. (2001). Implicit learning of first-, second-, and third-order transition probabilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 483–498.

Schvaneveldt, R. W. & Gomez, R. (1998). Attention and probabilistic sequence learning. *Psychological Research, 61*, 175–190.

Searle, J. R. (1990). Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences, 13*, 585–642.

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

Shanks, D. R. & Johnstone, T. (1998). Implicit knowledge in sequence learning tasks. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learning* (p. 533–572). Thousand Oaks, CA: Sage.

Shanks, D. R. & St. John, M. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367–447.

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

Whittlesea, B. W. A. (1997). The representation of general and particular knowledge. In K. Lamberts and D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 335–370). Hove, UK: Psyhcology Press.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition, 27*, 561–572.

Willingham, D. B., Greenberg, A. R., & Thomas, R. C. (1997). Response-to-stimulus interval does not affect implicit motor sequence learning, but does affect performance. *Memory and Cognition, 25*, 534–542.

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.

Willingham, D. B., Wells, L. A., Farrel, J. M., and Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory* & *Cognition, 28,* 366–375.

Ziessler, M. (1998). Response-effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 962–978.

Ziessler, M. & Nattkemper, D. (2001) Learning of event sequences is based on response-effect learning: Further evidence from a serial reaction task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 595–613.

# PART II

# Nuroscientific and computational approachese

# Neural structures that support implicit sequence learning

Eliot Hazeltine
NASA – Ames Research Center

Richard B. Ivry
University of California, Berkeley

## 1. Introduction

Motor sequence learning is a fundamental component of human behavior. Our daily lives are filled with behaviors that involve complex sequences of movements, and, when we are sufficiently experienced with these tasks, we perform them nearly effortlessly. In contrast, when a novel task – even a simple one, such as operating a new mobile phone – is attempted for the first time, it can be highly demanding. Without motor learning, our lives would be an endless series of trivial but all consuming actions.

The primary goal of this paper is to meld concepts arising from behavioral investigations of sequence learning with findings from neuroimaging and neuropsychological studies. In contrast to the extensive work on perceptual systems, motor sequence learning remains poorly understood in terms of the functional differentiation of the implicated neural substrate. Motor learning clearly involves a distributed network of neural structures, including parietal, prefrontal, premotor, and motor cortex, as well as subcortical structures such as the cerebellum to the basal ganglia. However, the functional contributions of these various structures have remained largely speculative. That is, a sizable body of empirical research has established that many regions contribute to the production of learned motor sequences, but only a few clues have been uncovered to help distinguish the computations performed by these regions.

In this paper, we will describe why we believe much can be learned from the existing neuroimaging literature about the relationship between motor learning and attention. We will then discuss how behavioral studies have contributed to the interpretation of results from neuroimaging experiments. In particular, behavioral findings have suggested that independent sequence learning systems are engaged depending on the task demands. To identify the neural substrate of these systems, we focus on the findings from neuroimaging studies, although in some cases, neurophysiological and neuropsychological studies have played a critical role in our understanding of the function of a brain region. Neuroimaging findings have also indicated that sequence learning may engage distinct systems as a function of task conditions; how to characterize these systems remains controversial. The remainder of the paper will focus on the various regions of the brain that have been shown to participate in motor sequence learning. Each region will be considered in turn, with an emphasis on the roles that attention, explicit knowledge, and nature of the sequential information play on neural activity.

## 2. Imaging and the serial reaction time (SRT) task

The advent of neuroimaging has ushered in a plethora of new experimental techniques to the field of cognitive science. Methods that allow for whole brain imaging with reasonable spatial resolution, specifically positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), have become essential tools for cognitive neuroscience over the last 20 years. There are many variations in how these technologies are employed, but in the studies described below the underlying principle is essentially the same: The scanning device is able to detect changes in bloodflow that are specific to local regions of tissue. From these changes in bloodflow, increases or decreases in neural activity are inferred.

Such techniques provide a valuable complement to traditional neuropsychological approaches to attributing function to neural structures. For instance, one reason for the difficulty in attributing specific functions to the different structures has to do with the distinction between motor learning and motor performance. Individuals with brain damage may present deficits in performing learned behaviors, but researchers must determine whether these deficits reflect problems with learning or problems in production. A person may possess intact representations of the learned behavior but be unable to express the encoded information properly due to damage to the brain structures that are necessary for producing the movements.

Imaging studies can avoid the "execution vs. learning" problem to some degree because they identify learning-related changes in a normally functioning brain. Thus, there are no disrupted systems to alter the operation of other intact systems. Moreover, imaging experiments can be designed to maximize the likelihood that any observed changes in neural signal relate to motor learning rather than motor execution. Movements are often performed with greater speed as they become well learned. In fact, the speed at which movements are performed is often used as a measure of learning. This behavioral phenomenon has important consequences for researchers using imaging techniques to study motor learning. A brain region may be sensitive to the rate at which movement is performed simply because there is more motor output within a given period of time. This type of change in activity should not be considered to necessarily reflect motor learning per se, given that the computation performed by the underlying region may remain identical as the movement becomes more practiced. Therefore, most studies have required participants to perform the movements at a constant rate regardless of the amount of practice. Under constant-rate conditions, changes in neural activation can be tightly linked to learning.

For the purposes of the present paper, we will rely primarily on studies that have used the serial reaction time (SRT) task. In the basic SRT task, participants perform a choice reaction time task, often involving visual stimuli with a spatially compatible S-R mapping. The order of the stimuli can be random or follow a fixed sequence, typically 6–12 elements long. Sequence learning is assessed by comparing reaction times of trials in which the stimuli follow a fixed sequence to trials in which the order of the stimuli is randomly determined. Because learning can include aspects of the task that are independent of the sequence per se, such as the strengthening of S-R associations, it is important that the comparison involve random and sequence trials taken from similar points in training. Therefore, blocks of random trials are usually introduced near the end of an experimental session to provide a clean measure of sequence learning.

Literally dozens of imaging studies have been conducted using variants of the SRT task. This broad empirical base allows for a somewhat detailed consideration of how the brain activation is affected by sequence learning under varying task demands. While there are clearly advantages to drawing comparisons across diverse experimental procedures, such an approach also possesses some potential pitfalls. First, the tasks present different computational demands. For example, in trial-and-error versions of the SRT task, the participants are actively testing hypotheses and using the feedback information to adjust their performance. In the basic version of the task, no feedback is provided and the participants simply are instructed that they are to perform a continuous choice reaction time task. With compatible S-R mappings (or some training with less compatible mappings), error rates are low and the measure of learning is based simply on the change in reaction time. With this approach, the experimenter can influence the degree to which participants become aware of the presence of a sequence. Indeed, in many conditions, the participants are unaware of the sequence, allowing the SRT task to be used as a model for procedural sequence learning.

A second potential pitfall of making comparisons across experiments stems from the fact that different approaches have been employed for analyzing the imaging data. Based on neurophysiological studies showing that neurons become more active as the animal learns to perform a particular movement sequence, most studies have focused on increases in activation during sequence learning, although decreases are often observed as well. In some studies (Grafton, Hazeltine, & Ivry 1995, 1998; Hazeltine, Grafton, & Ivry 1997), a correlational method was used to identify activation changes that were associated with learning (see also, Honda et al. 1998). This approach is distinct from the more traditional subtractive logic applied in imaging studies. For example, Doyon, Owen, Petrides, Sziklas, and Evans (1996) report a comparison between a block in which the stimuli followed a well-learned sequence with a subsequent block in which the successive stimuli were randomly determined (see also, Catalan, Honda, Weeks, Cohen, & Hallett 1998; Jenkins, Brooks, Nixon, Frackowiak, & Passingham 1994; Rauch et al. 1995; Sadato, Campbell, Ibanez, Deiber, & Hallett 1996). The notion that this comparison would reveal areas associated with sequence knowledge rests on the assumption these areas would immediately deactivate with the presentation of random stimuli. However, it is also possible that these areas continue to perform their sequencing-related computations, even though they are no longer appropriate. If this were so, such areas would not be detected by this comparison.

Nonetheless, there have been numerous imaging studies of motor sequencing and an integration of this work reveals some interesting points of similarity as well as raises some important issues for future work. The investigation of any psychological process benefits from the application of converging methods. Given the complexity of sequence learning, progress in identifying the psychological and neural components is particularly likely to result from comparisons across multiple studies. The manipulations employed by individual studies may affect multiple components, leading to a plurality of potential interpretations. Evaluating the patterns of results from a range of experiments should help to isolate the critical task components that affect neural activity in specific brain regions.

The imaging data is buttressed by a sizeable body of behavioral studies that have explored the representational nature of human sequence learning. Such research has provided invaluable constraints on the interpretation of imaging data. For instance, studies that have employed a transfer design (Cohen, Ivry, & Keele 1990; Hazeltine 2002; Keele, Jennings, Jones, Caulton, & Cohen 1995; Palmer & Meyer 2000; Willingham, Wells, Farrell, & Stemwedel 2000) indicate that representations formed during sequence learning do not specify anatomic units (i.e., particular finger movements) but instead involve more abstract codes that possibly include features of the environment. These findings make clear that much of the changes in neural activity associated with motor sequence learning do not stem from the encoding of relationships between low-level muscle commands (see below). Instead, these changes likely reflect the acquisition of movement endpoints (e.g., Willingham et al. 2000) or high-level action goals (e.g., Hazeltine 2002). This perspective has been slow to be absorbed by much of the neuroimaging literature; some studies have interpreted activation strictly in terms of associations between effectors.

## 3. How many forms of sequence learning?

Many studies (e.g., Cohen et al. 1990; Curran & Keele 1993; Jiménez & Méndez 1999; Schmidtke & Heuer 1997; Stadler 1995) have attempted to characterize the specific benefit of attention on sequence encoding. Since the task's introduction, it has been recognized that sequence learning is clearly affected by the availability of attentional resources (Nissen & Bullemer 1987). One major issue centers on whether there two distinct sequence learning systems, one independent of the availability of attention and the other enabled only when sufficient attentional resources are committed to the sequential task. In describing this dichotomy, we refer to the former learning system as the "non-attentional" system and the latter as the "attentional" system. In adopting this terminology, it is important to bear in mind that the non-attentional system is not unable to encode information that is attended (c.f. Jiménez & Méndez 1999). Rather, this system is, under appropriate conditions, able to form associations between task-relevant items, regardless of the attentional load.

Researchers have compared learning under single-task conditions, during which attention can be devoted to the SRT task, and dual-task conditions, during which attention must be allocated to both a secondary task and the SRT task. According to adherents of this two-system view, dual-task conditions prevent the attentional system from encoding sequential relationships in the SRT task. Thus, the imposition of an appropriate secondary task can serve to determine whether sequence learning occurs within both systems (when the secondary task is absent) or only in the non-attentional system (when the secondary task is present).

Although the two-system view has received considerable empirical support, it remains a controversial account of a large and complex body of data. The psychology of learning has emphasized the classical distinction between implicit and explicit processes. Given the dominance of this theoretical dichotomy, it is tempting to equate the attentional system with explicit learning and the non-attentional system with implicit learning. However, such a straightforward connection between the two frameworks is simplistic. Implicit learning is frequently studied under single-task conditions (e.g., Nattkemper & Prinz 1997; Reed & Johnson 1994; Wachs, Pascual-Leone, Grafman, & Hallett 1994; Willingham & Goedert-Eschmann 1999; Willingham, Nissen, & Bullemer 1989), indicating that the availability of attention is no guarantee that learning will be explicit.

Other characterizations of distinct learning systems have focused on the types of information that are encoded rather than the availability of the systems (c.f. Stadler 1995). Keele, Ivry, Mayr, Hazeltine and Heuer (under review) have proposed that one learning system is composed of modules that can encode sequence information that is restricted to a single input dimension (e.g., location or shape). Sequences can be encoded within these modules, but the encoded representations cannot include information that occurs along other dimensions. A second, independent system is accessible to inputs from multiple dimensions, and thus can form complex, multidimensional associations (as well as unidimensional associations if information is only present on a single dimension). The cost of the increased flexibility in the multidimensional system is that the system can be easily disrupted by sources of unrelated information that are attended to by the organism. In contrast, the unidimensional system, while more limited in the range of sequences it can encode, is able to identify invariance in individual input streams when multiple sources of information impinge on the organism. Explicit knowledge can only emerge from the multidimensional system.

Other researchers (e.g., Dienes & Berry 1997; Perruchet & Amorim 1992; Shanks & St. John 1994, 1996) have raised more fundamental objections to the two-system view and argued against the proposal that performance on the SRT and related tasks is mediated by multiple, distinct sequence learning systems. Differences between implicit and explicit measures may reflect the sensitivities of the various behavioral tests. According to this view, tasks that purportedly assess implicit and explicit knowledge tap the same sequence representations. When individuals are unable to report the sequence but benefit from its presence during performance of the SRT task, the representation is incomplete or in a sub-threshold state.

These theories have been constructed to account for behavioral data, but the neuroimaging literature is replete with experiments that have employed the SRT task and thus speak to the question of whether distinct learning systems encode sequence information. Indeed, several imaging studies have attempted to test the two-system theory by comparing the neural activation associated with sequence learning during single- and dual-task conditions, or by comparing activation during conditions in which there is explicit knowledge to conditions in which there is only implicit knowledge. The results, described in detail below, have supported the hypothesis that multiple, non-overlapping neural systems are involved in sequence learning, and that the degree of their involvement is highly dependent on task conditions. However, as with the behavioral findings, the data are open to multiple interpretations, and identifying the critical task components that determine the pattern neural activation has proven less than straightforward.

To address these issues, the remainder of this paper is organized in terms of the brain regions that have been implicated in motor sequence learning. Each region is discussed in terms of how attention, explicit knowledge, and the nature of the sequential information affect its activation during the performance of SRT-like tasks. Findings from neurophysiological and neuropsychological studies focusing on these regions will also be described. From these observations, we attempt to sketch some hypotheses about the functional roles of the brain structures and how they may interact to form distinct sequence learning systems.

## 4. Motor cortex

An obvious starting point for a tour of the brain structures supporting motor sequence learning is the primary motor cortex. In their seminal imaging study, Roland, Larsen, Lassen, and Skinhoj (1980) reported regional cerebral blood flow (rCBF) increases in sensorimotor cortex during movements ranging from simple isometric finger flexions to sequences involving the five digits of the hand. While the importance of motor cortex in volitional movement is well established, its role in motor learning is less clear. In some studies sequence learning is accompanied by increased rates of movement, making unclear whether increased activation reflects sequence learning per se or higher rates (e.g., Schlaug, Knorr, & Seitz 1994). However, the studies of Grafton et al. (1995) with visual spatial sequence learning and Hazeltine et al. (1997) with color sequence learning controlled for rate by employing a fixed interval between successive stimuli. That is, the fixed interstimulus interval meant that responding quickly to the stimuli did not affect the overall rate at which responses were made. Nonetheless, the researchers found clear evidence of increased motor cortex activation as sequence learning proceeded.

In these studies, motor cortex activity was consistently observed in the dual-task conditions, but not in the single task conditions. In fact, most imaging studies using single-task methods have failed to observe significant changes in motor cortex related to sequence learning (e.g., Berns, Cohen, & Mintun 1997; Doyon et al. 1996; Jenkins et al. 1994; Jueptner et al. 1997; Rauch et al. 1995). There are, however, some intriguing exceptions when the learned representations are well developed and awareness is no longer focused on the sequencing task (Jenkins et al. 1994; Passingham 1996). In support of such a view Karni, et al. (1995) found motor cortex activity to increase over many days of extensive practice on a finger sequencing task.

Pascual-Leone et al. (1994) have reported a similar pattern using transcortical magnetic stimulation (TMS) to measure changes in the size of motor fields over the course of learning. Early in training the motor fields grew larger in correspondence with behavioral indices of learning. However, the motor fields later returned to their original size as soon as subjects developed explicit knowledge of the sequence. Thus, the evidence consistently shows motor cortex involvement in sequence learning, either during dual-task conditions or when focused attention is not involved. Honda et al. (1998) report increased rCBF over contralateral motor cortex in single task SRT learning, but only during the initial stages. This area was not correlated with performance once subjects began to attend to the sequential nature of the stimuli. In fact, at this point, the prominent foci were in the right hemisphere, similar to shifts reported by Grafton et al. (1995) and Hazeltine et al. (1997).

These results have been interpreted as reflecting plasticity within the motor cortex during motor learning (Karni et al. 1995; Pascual-Leone et al. 1994).

This interpretation challenges the traditional notion that motor cortex units activate target muscle groups in a context independent manner and suggests an expanded role for motor cortex to include encoding of movement sequences. However, the behavioral phenomena pose some problems for this account.

Sequence learning in tasks such as the SRT task can occur at a relatively abstract level; sequence knowledge acquired during training with one set of effectors, such as fingers, can be transferred to other effectors, such as arm, even when subjects are unaware of the sequence or when they learn the sequence under dual-task conditions (Cohen et al. 1990; Keele et al. 1995). Such transfer across effectors suggests that an abstract sequence representation is tied to the stimulus properties or response properties (i.e., knowledge of the locations to which action is directed) rather than to motor processes. There are certainly situations in which sequential representation arises at the motor level, but such motor representation seems restricted to very short movement segments, such as the coordinated muscular events that make up a single letter stroke in handwriting (see Lindemann & Wright 1998) or between immediately adjacent strokes in typing by experts (e.g., Jordan 1995).

Beginning with the seminal neurophysiological work of Penfield (Penfield & Boldrey 1937), a high degree of effector specificity has been assumed to characterize motor cortex. Recent neuroimaging studies provide converging evidence (Grafton, Mazziotta, Woods, & Phelps 1992). It is unclear how effector-independent sequences would be encoded in motor cortex, a region presumed to involve effector-specific units. Thus, it is reasonable to question whether the activation observed in this brain region really reflects sequence encoding.

An alternative hypothesis suggests a different functional role of the motor cortex in sequence learning. Perhaps the observed increases are not the consequence of its reorganization with learning, but instead reflect an increase in input to this structure. By this view, sequential learning is restricted to regions upstream from motor cortex. For example, areas in supplementary motor cortex, having encoded the sequence, may exert a priming effect on their effector-specific targets in motor cortex. One PET study provides direct support for this hypothesis (Grafton et al. 1998). Under dual-task conditions, participants were trained on the SRT task, responding to color stimuli with their fingers. As in earlier studies, rCBF increased in contralateral motor cortex. After learning was established, they were transferred to a condition in which the responses were made with whole-arm movements, first with random stimuli and then with the original sequence. When the sequence was reinstantiated, an increase in activity was again observed in motor cortex. However, the center of activity shifted to a more dorsal position, consistent with the crude somatotopy of this region.

Studies that control movement rate seem especially sensitive to potential priming effects given that they introduce relatively long delays between consecutive responses. These conditions exist for both the Karni et al. study and the PET studies (Grafton et al. 1995; Hazeltine et al. 1997). The priming hypothesis is related to the distinction between performance and learning. Does an increase in metabolic activity reflect local reorganization or changes in upstream processes? A performance-based interpretation is consistent with several neuroimaging studies that have failed to find learning-related changes in motor cortex (Friston, Frith, Passingham, Liddle, & Frackowiak 1992; Jenkins et al. 1994; Rauch et al. 1995; Rauch et al. 1997). Moreover, activity in motor cortex is comparable when subjects perform simple and complex movements (Roland et al. 1980; Shibasaki et al. 1993). Finally, when subjects are asked to imagine producing movement sequences, motor cortex activity remains at baseline levels, suggesting that this region does not play a role in the storage of sequential knowledge (Decety et al. 1994; Roland et al. 1980).

## 5. Premotor cortex and SMA

Premotor cortex, including both the medial supplementary motor area (SMA) and lateral regions (referred to here as PMC), has consistently been linked to sequential behavior. Numerous PET studies have shown increased rCBF during the production and acquisition of movement sequences, even when the number of movements is equated across conditions (Grafton et al. 1995; Hazeltine et al. 1997; Honda et al. 1998; Jenkins et al. 1994; Rauch et al. 1995; Sadato et al. 1996; Shibasaki et al. 1993). Unlike the results for motor cortex, these areas are also activated when subjects are asked to imagine producing movement sequences (Decety et al. 1994; Roland et al. 1980).

These imaging data are supported by findings from lesion studies in both humans and monkeys. Halsband et al. (1993) report that patients with premotor lesions are impaired in the production of rhythmic, sequential movements, with the deficit most marked when the lesions encompassed SMA. Similarly, premotor lesions severely disrupted the ability of monkeys to relearn movement sequences (Passingham 1993). Again, the effect was most evident when the lesions were made in the medial portion of premotor cortex.

Neurophysiological studies provide perhaps the most compelling evidence that sequential knowledge is encoded in premotor cortex. Tanji and Shima (1994; also Mushiake et al. 1991) trained monkeys to perform 3-element sequences consisting of push, pull, and turn gestures. Neurons in SMA responded selectively in advance of one of the gestures, but only when the gesture was embedded in a particular transition (e.g., a "pull" neuron would respond in the sequences push-pull-turn and turn-push-pull, but not for the sequences push-turn-pull and turn-pull-push). Mushiake et al. (1991) also found that cells in PMC were selectively activated prior to the onset of movement sequences rather than being linked to particular elements within the sequence.

This is not to say that the role of the two structures is identical. A variety of hypotheses have been proposed to differentiate the respective contributions of PMC and SMA, independent from sequence learning. One hypothesis centers on an external-internal distinction, with the PMC prominent in the control of externally-driven movements and the SMA associated with internally-generated movements (Goldberg 1985; Halsband et al. 1993). External movements are those cued by events such as the appearance of a visual stimulus to be touched or a cue indicating that a movement should be initiated. Internal movements are those initiated and guided without the assistance of environmental cues. Mushiake et al. (1991) classified PMC and SMA neurons on the basis of whether their activity was related to visually-guided or internally-generated movement. While the populations overlapped, PMC neurons were twice as likely to be active during visually-guided movement compared to internally-generated movement; the reverse was observed for SMA neurons.

An alternative way to characterize the differences between SMA and PMC focuses on computational requirements that coexist during the performance of most movements. SMA may play a more prominent role in the representation and generation of sequential actions (see Tanji & Shima 1994). In addition to the single-cell research cited previously, lesions of the SMA in humans impair performance on sequencing tasks (Halsband et al. 1993; Watson, Fleet, Gonzalez-Rothi, & Heilman 1986) and, in the monkey, produce a greater impairment on sequence tasks than do lesions of PMC (Passingham 1993). In contrast, neurons in PMC underlie associations between stimuli and movements. In this view, PMC may not be involved in sequence representation per se, but rather in the formation of links between external events and appropriate actions (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti 1992; Rizzolatti & Gentilucci 1988). A sequence representation composed of codes that refer to external events could account for the patterns of behavior in SRT tasks in which performance benefits are observed in transfer conditions in which the effectors are changed but the response endpoints or feedback remains the same (Hazeltine 2002; Keele et al. 1995; Willingham et al. 2000).

An important point here is that these links can be relatively arbitrary. While in many situations, stimulus-action associations are direct – for example, when we reach to catch a thrown ball – many actions involve relatively arbitrary associations. For example, we learn to stop at a red light or press the remote to change channels on the television. Wise, di Pellegrino, and Boussaoud (1996) suggest that PMC is especially critical when the appropriate action does not conform to the most direct stimulus-response mapping. As reviewed in Passingham (1993), lesions of PMC in the monkey and human lead to severe impairments on tasks in which the subjects must learn arbitrary associations.

One way to integrate the internal vs. external perspective with the sequential vs. arbitrary approach is to consider how the cues guiding sequential actions can vary with learning. Jenkins et al. (1994) trained subjects to produce a sequence of eight elements. Feedback signals were provided after each response, enabling the subjects to learn the sequence through trial and error. After mastering one sequence, the subjects repeated the process with another sequence. In this way, PET scans could be obtained during the performance of a well-learned sequence and during acquisition of a new one, without confounding scan order with skill level. Activation was greater in PMC when subjects were learning a new sequence (see also, Jueptner et al. 1997), and shifted to SMA when they performed the well-learned one. This can be interpreted as a change from performance being guided by the external feedback signals to one in which the subjects have internally encoded the sequence.

It is interesting to note that across many imaging studies of sequence learning, there has been a consistent dissociation between SMA and PMC. With one exception (Honda et al. 1998), task-related neural activity has been observed in one area or the other, but not both. In the SRT PET studies (Grafton et al. 1995, 1998; Hazeltine et al. 1997), learning related changes were obtained in SMA under dual-task conditions, whereas under single-task conditions, the changes were found in PMC.

An account of these differences can be formulated by considering the types of representations and computations performed within these secondary motor areas. In the dual-task condition, response sequences may be organized in relation to each other. That is, the production of one response element in the sequence primes subsequent responses, and it is this implicit knowledge of the sequence that is reflected in the SMA activation (and interconnected parietal cortical regions). This priming effect also spills over into motor cortex as the sequence-specific SMA and parietal representations activate particular effectors.

The level of representation in SMA is conceptualized as relatively abstract in two distinct senses. First, the evidence from single-neuron recording suggests that SMA is involved in coding groups of responses, not single responses, and hence also would not be coding for specific movements. Second, as described in the previous section, the behavioral evidence suggests that the sequence code can be transferred among different movement effectors, and hence more likely represents response goals. This difference in specificity between SMA and motor cortex is made clear in one PET study (Grafton, Mazziotta, Presty et al. 1992). Similar loci of activation were found in SMA for tracking movements performed with either the finger or tongue. In contrast, effector specificity was evident in motor cortex; these conditions led to a significant shift in the center of activation. Similarly, in the SRT transfer study of Grafton et al. (1998), the SMA focus of activation remains high across scans during which the sequence is produced with finger or arm movements.

Under single-task conditions, responses are guided by the retrieval of successive stimulus-response associations. Learning-related changes in rCBF in this case occur in PMC as this region provides the essential mapping. According to this framework, PMC would also be recruited during dual-task performance when the signals for both tasks are predictive of one another. However, in PET dual-task studies to date, an increase in PMC activity was not observed across the sequence blocks because the retrieval of stimulus-response associations alternated between the sequential visual events and the random tones. Because PMC operates on both types of inputs, the random tones precluded the formation of a sequential representation. Consistent with this proposal, Rauch et al. (1995) observed increased rCBF in premotor cortex during single-task learning when the subjects were unaware of the sequence, although the center of activity was considerably inferior to the premotor foci reported in the other PET studies. PMC but not SMA activity is also correlated with sequence complexity when subjects produce well-learned movement patterns (Catalan et al. 1998; Sadato et al. 1996).

One exception to this pattern is found in the results of Honda et al. (1998). Using a correlational analysis similar to Grafton et al. (1995), parallel increases in PMC and SMA were observed during single task SRT learning. Based on their continuous assessment of awareness, these changes were initially evident during the early stages of sequence knowledge, the point where subjects indicated they believed the stimuli were non-random, but could not report the sequence. Focusing the analysis on this phase may provide a window in which metabolic changes can be detected in both learning systems. Interestingly, as in the Grafton studies, the SMA activity was centered in the contralateral hemisphere whereas the PMC activity was lateralized to the ipsilateral hemisphere. The fact that these studies consistently required participants to respond with their right hands complicates theorization about the significance of these results.

## 6. Prefrontal cortex

Activation in the lateral prefrontal cortex (PFC) has been reported in most of the published SRT PET studies under single-task conditions (Doyon et al. 1996; Grafton et al. 1995; Hazeltine et al. 1997; Honda et al. 1998; Rauch et al. 1995). Moreover, a learning related prefrontal increase in rCBF was reported by Berns et al. (1997), using a different motor sequence task. In their task, the sequence did not follow a fixed pattern, but instead the successive responses were determined by a set of implicitly-learned rules applied in a probablistic manner. As reaction times decreased, activity increased in the right prefrontal region.

There are considerable differences in the exact locus of the learning-related changes in PFC across these studies. In the Grafton, Berns, Hazeltine, and Honda studies, increases in rCBF were found in the right PFC. This laterality pattern is especially intriguing given that in all four studies the responses were made exclusively with the fingers of the ipsilateral right hand. Doyon et al. (1996) report greater activity in the right PFC when subjects preformed a sequence they could explicitly report compared to the earlier performance of that sequence when awareness was less developed. When the explicit condition was compared to a condition in which a new sequence was introduced, however, PFC activity was observed in the left hemisphere. This finding is similar to the left PFC activity reported by Rauch et al. (1995), who used a similar comparison (explicit-random). It is noteworthy that, unlike in the other experiments, the movements were bimanual in the Rauch et al. study.

In addition to the laterality issue, the exact loci within prefrontal cortex varies considerably across studies. Hazeltine et al. (1997) hypothesized the existence of dimension specific regions within lateral PFC. The linking of spatial stimuli to responses was attributed to a relatively dorsal region (area 46) of PFC, and the corresponding operation for color stimuli was attributed to a more ventral region (area 45). However, this picture is not well supported across the set of sequence learning imaging studies. The area 45 activation in Rauch et al. and Doyon et al. during spatial sequence learning are quite close to the focus observed by Hazeltine et al. during color sequence learning. The activation reported by Berns et al., is also centered in a similar region. Moreover, similar foci were identified by Grafton et al. and Honda et al., but the former used a spatial sequence and the latter a digit sequence. Thus, there is consistent activation of area 45/46 during single-task sequence learning, but little indication exists for a systematic effect of stimulus type on the locus of the activity.

These lateral prefrontal regions have been hypothesized to form a critical component of a working memory system. According to this framework, transient representations provide links between perceptual-based knowledge distributed across posterior regions and the task-relevant goals. For example, in the delayed-matching-to-sample task, the animal must remember a cued location for a forthcoming response. Thus, the working memory function of lateral PFC is unlikely to be restricted to sequence learning. Indeed this area shows activation across a wide range of tasks (e.g., McCarthy et al. 1994; Smith et al. 1995).

Recent studies of response competition have also emphasized the importance of lateral PFC. When task conditions suggest inappropriate responses, either through expectations (Garavan, Ross, & Stein 1999; Konishi, Nakajima, Uchida, Sekihara, & Miyashita 1998), irrelevant stimuli (Casey et al. 2000; Hazeltine, Poldrack, & Gabrieli 2000), or contradictory feedback (Fink et al. 1999), activity in the ventral portion of the right lateral prefrontal cortex near area 45 is frequently observed. Such findings suggest an alternative account for the similar activity observed during single-task sequence learning. Perhaps as subjects are better able to anticipate subsequent stimuli, they exert more inhibitory control over the responses that are not likely to be immediately produced. For example, once learning that response 2 follows response 3, the subject may inhibit responses 1 and 4 to facilitate the response selection processes and improve performance.

A general feature across the SRT imaging studies is that sequence related prefrontal activation appears to be restricted to learning that occurs under single-task conditions, that is, when there are no intervening random events between successive elements of the sequence. Grafton et al. (1995) did observe some activation associated with learning during dual-task conditions in the prefrontal cortex, but the focus was located in a more anterior and medial portion of the frontal lobe than the foci reported in single-task studies (e.g., Berns et al. 1997; Doyon et al. 1996; Eliassen, Souza, & Sanes 2001; Grafton et al. 1995; Hazeltine et al. 1997; Honda et al. 1998; Jenkins et al. 1994; Jueptner et al. 1997; Toni, Krams, Turner, & Passingham 1998).

The reports of prefrontal activation share other features as well. In all five PET SRT studies, as well as Berns et al. (1997), activity in prefrontal cortex remains high when learned responses are made following the presentation of visual stimuli. Sequence-specific cells have also been recorded in the lateral PFC of the monkey during the production of a spatially cued sequence (Barone & Joseph 1989). In contrast, no change in activity is observed in this area when humans produce well-learned sequences from memory (Catalan et al. 1998; Deiber et al. 1991; Jenkins et al. 1994; Sadato et al. 1996; Shibasaki et al. 1993). Indeed, the lack of activity is particularly striking given that in the Shibasaki et al. and Jenkins et al. studies, comparisons were made to a resting condition. Here, the production of sequential movements was associated with activity in SMA. This dissociation of prefrontal and SMA activity fits well with the external-internal account of the differential contributions of lateral and medial motor pathways.

One issue here is how internally generated motor sequences are learned. Jenkins et al. (see also Ghiladri et al. 2000, 1994) used tones to indicate whether responses were correct in a trial and error learning procedure. During the learning phase, lateral PFC was active compared to rest. However, when subjects performed a well-learned sequence, the activation in the PFC returned to resting levels. What changes in the course of learning dual-task sequences or internally generated movements is the utility of external information. The PFC may serve as a conduit for perceptually-based feedback necessary for sequence acquisition, even for internally generated movements. Once the sequence is sufficiently encoded and feedback becomes less relevant, this region is no longer required. Thus, it may be that the PFC is necessary early in learning, when the sequence representation is insufficient to drive behavior without the external stimuli that indicate the appropriate responses (see Toni et al. 1998).

Pascual-Leone et al. (1996) found further support for a dissociation between dorsolateral prefrontal cortex (DLPFC) and SMA during sequence learning using transcranial magnetic stimulation. They applied transcranial magnetic stimulation continuously during SRT training. The magnetic coil was centered over SMA, ipsilateral DLPFC, or contralateral DLPFC. Magnetic stimulation had minimal effect on overall reaction time for all groups. However, compared to a control group that did not receive magnetic stimulation, stimulation over contralateral DLPFC significantly reduced sequence learning. The effect was specific to this area as shown by the fact that stimulation to SMA and ipsilateral DLPFC did not affect learning. Given that sequence learning occurred under single-task conditions in this experiment, we would expect the DLPFC to be essential: Participants responded to the series of visuospatial signals and there were no intervening random events. Note that the TMS findings are at odds with the patient and imaging studies in one respect; the latter suggests a prominent role of right PFC even for right-hand movements.

An interesting prediction derives from this explanation. If this experiment were repeated under dual-task conditions, stimulation over SMA should disrupt learning, whereas stimulation over DLPFC would have no effect. In effect, the random tones may act much like transcranial magnetic stimulation by disrupting the representation of sequential information in DLPFC. Under such conditions, sequence learning is restricted to the non-attentional learning system involving SMA. Now, magnetic stimulation of SMA should disable this system.

A final issue regarding the role of the frontal lobes in sequence learning focuses on awareness. Frontal lobe function, including working memory, has have frequently been linked to consciousness, and during single-task learning, in which prefrontal activity is seen, subjects frequently become explicitly aware of the sequence (Cohen et al. 1990; Frensch, Buchner, & Lin 1994; Willingham et al. 1989). Grafton et al. (1995) found that subjects who became aware in the single-task condition showed significantly greater activation in PFC during sequence learning than those who did not become aware. Similarly, prefrontal activation has been correlated with explicit sequence knowledge in the other SRT PET studies (Doyon et al. 1996; Honda et al. 1998; Rauch et al. 1995).

Other evidence from the imaging literature suggests that lateral PFC activity may be associated with factors unrelated to the establishment of sequence awareness. Rauch et al. did not find significant differences in this area during the performance of explicitly and implicitly learned sequences. Similarly, learning related increases in PFC was observed for the unaware subjects in the Hazeltine et al. (1997) study and the Grafton et al. (1995) study, though statistically reliable only in the former case. While awareness has proven to be a thorny problem in the study of implicit motor learning (see Perruchet & Amorim 1992; Shanks & St. John 1994), Berns et al. (1997) report that their subjects showed essentially no evidence for awareness in their probablistic sequence learning study. Nonetheless, performance was significantly correlated with increases in lateral PFC. In sum, there are some reports of activation in the PFC without the development of explicit knowledge, but few reports of explicit knowledge without PFC activation. Thus, it appears that PFC is a component of a learning system that is engaged during single-task sequence learning from which explicit knowledge of the sequence can emerge.

## 7. Temporal and occipital cortex

In addition to the action-related anterior regions, learning-related changes in activity are often reported in the occipital and temporal lobes during sequence learning. Such foci might seem surprising given that these areas are usually considered to be involved in perception rather than action. However, during sequence learning in the SRT, stimuli as well as responses can be anticipated as performance improves. Thus, the most likely interpretation of activation in these regions is that it reflects stimulus expectations or classical perceptual learning. To date, none of the imaging studies using dual-task conditions have reported activity in either the temporal or occipital lobes, suggesting that, as with the PFC, the availability of attention may be critical for these structures to participate in the performance of learned sequences.

For example, area 19 consistently shows learning-related activation (Grafton et al. 1995; Hazeltine et al. 1997; Rauch et al. 1995), frequently bilaterally under single task conditions. Temporal lobe loci were also observed in these studies. The exact locations are not consistent across the studies. Hazeltine et al. attributed these differences to stimulus-specific characteristics and the recruitment of distinct neural processes required for analyzing the spatial and color information used to indicate responses. The loci, however, do not match up with those identified in imaging studies designed to map human visual cortex. The area 19 focus showing increased rCBF with learning in the color-cued task of Hazeltine et al. is more lateral and inferior than the color center identified in PET (Lueck et al. 1989; Zeki et al. 1991) and fMRI studies (Clark et al. 1997).

That the occipital and temporal regions appear to be active only under single-task conditions suggests that they are functional linked with the PMC and lateral PFC. Although activity is not always observed within all three regions in a given experiment (e.g., Doyon et al. 1996; Rauch et al. 1995), the pattern across studies is consistent with the structures forming a sequence learning system that is engaged when attention is available. The PMC and PFC, which presumably encode the associations among the sequence elements, may facilitate perceptual priming, via either direct or indirect projections from the PFC to posterior structures (Barbas 1988; Barbas & Pandya 1989). This idea borrows from current theories concerning how lateral PFC modulates activity in temporal and extrastriate cortex in visual memory (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen 1991; Desimone, 1996; Desimone & Duncan 1995). According to this framework, activity in the premotor cortex could be due to additional priming from PFC or from inputs originating in posterior cortex. This account indicates that learning here can be stimulus-based, with the system anticipating “what” perceptual characteristics will appear next.

To date, all of the PET SRT studies have relied on stimulus sequences that are restricted to a single input channel (e.g., color or location). However, some behavioral evidence suggests that without attention, sequence learning may not be able to form associations across distinct channels of information (e.g., Hazeltine, Ivry, & Chan, in preparation; Schmidtke & Heuer 1997). Therefore, one might suppose that a system comprised by PFC, PMC, and regions with the temporal and occipital lobes is necessary to learn cross-modal sequences (Keele et al., under review). Mesulam (1998) associates consciousness with cross-modal regions of the brain, pointing out that we do not consciously perceive single, unbound features. In this regard, it is interesting to note that this learning system, and the PFC in particular, are often associated with explicit knowledge of the sequence (e.g., Doyon et al. 1996; Ghiladri et al. 2000; Grafton et al. 1995; Hazeltine et al. 1997; Honda et al. 1998; Jenkins et al. 1994).

Indeed, evidence from non-learning tasks is consistent with the proposal that these structures are well-suited for integrating information across separate dimensions. This property is prevalent in theories of processing within the ventral stream of posterior cortex (Goodale, Milner, Jakobson, & Darey 1991; Ungerleider & Mishkin 1982). For example, temporal lobe areas are critical in face recognition and object recognition processes, with computational models emphasizing that this area can sustain the requisite multidimensional, integrative representations (Desimone & Duncan 1995). Stimulus-based priming might occur in modality-specific posterior regions, or it might be instantiated in polysensory regions of the superior temporal cortex (e.g., Watanabe & Iwai 1991) or PFC (see Fuster, Bodner, & Kroger 2000; Mesulam 1998).

## 8. Parietal cortex

Regions within parietal cortex are activated during SRT sequence learning as frequently as any other region of the brain. When learning occurs under dual-task conditions, the activation consistently increases in the anterior parietal lobe. A left hemisphere focus at the border of areas 40 and 7 was identified in the study using spatial stimuli (Grafton et al. 1995); numerous bilateral foci in more inferior regions of area 40 were seen with the color stimuli (Grafton et al. 1995; Hazeltine et al. 1997). Parietal foci are also observed to show learning-related changes during studies of sequence learning under single-task conditions. In contrast to the more anterior area 40 activation observed in the dual-task studies, parietal foci under single-task conditions are generally limited to area 39, especially near the border of area 19.

A potential role for these parietal areas is that they encode action-based representations for spatially directed movements. Indeed, some theorists of motor sequence learning propose that sequences are encoded in terms of spatial coordinates (e.g., Hikosaka et al. 1999; Willingham et al. 2000). This proposed role in learning is related to more general accounts of parietal function: this region has been viewed as a critical component of a system for generating the spatial characteristics of an action (Crammond 1996). Thus, whereas the temporal-PFC-PMC network may serve to anticipate the properties of successive stimuli, the parietal region appears to provide the spatial codes for goal-directed actions (Andersen 1994).

This idea is related to the "what-where" distinction of the functions of posterior cortex. In the initial formulation of Ungerlieder and Mishkin (1982), the dorsal "where" system was essential for computing an object's position in exocentric space. More recently, "where" has been extended to include "how", to capture the importance of spatial information for the performance of goal-directed movements (Goodale et al. 1991; Jeannerod et al. 1994; Milner & Goodale 1995). While the temporal and occipital regions that constitute the ventral "what" system may process input properties relating to the identity of identity of the upcoming stimuli, the parietal regions may encode the action-related information. Although controversial (see Carey 2001), some studies (e.g., Glover & Dixon 2001; Haffenden & Goodale 1998) have indicated that visual tasks that presumably engage the "how" system are less sensitive to visual illusions induced by irrelevant stimulus information than tasks that invoke the "what" system. This difference in sensitivity parallels the conjecture that there is one learning system that is disrupted by a secondary task and a second learning system that is less susceptible to irrelevant information but more limited in terms of flexibility (Keele et al., under review; Schmidtke & Heuer 1997).

The role of parietal cortex in goal-directed movements is supported by a wide range of evidence. This area is consistently activated in PET studies during the production of voluntary movements in extrapersonal space (Deiber et al. 1991; Harrington et al. 2000; Jenkins et al. 1994; Roland et al. 1980). Grafton et al. (1992) report bilateral activation of dorsal parietal cortex during visually guided finger movements and this activation increases when the spatial complexity of the task is increased. Increases in rCBF in this area are also obtained during imagined movements (Crammond 1996; Decety, Kawashima, Gulyas, & Roland 1992). The apraxia literature has also pointed towards the importance of parietal cortex in the long-term representation of spatially directed action.

Of note here is that apraxic patients with posterior lesions not only have difficulty in producing coordinated gestures but are also impaired in perceiving them (Heilman, Rothi, & Valenstein 1982). Such a deficit suggests a common representation subserving perception and action.

The well-documented involvement of the parietal lobe in spatial transformations and its near-ubiquity in SRT learning suggest that its role may relate to encoding the movements that constitute the sequence. Behavior SRT studies have provided considerable evidence that under dual-task conditions, sequence learning is largely based on representations specifying particular actions. Hazeltine (2002) found that sequence learning under dual-task conditions was preserved when stimulus order was altered but response order remained the same (see also Nattkemper & Prinz 1997; Willingham et al. 2000). Thus, it appears that sequence encoding must occur on representations that are not bound to particular stimulus properties.

However, while the evidence suggests that sequence learning is response-based, this is not to say that learning under dual-task conditions is restricted to low-level motor codes, or even locations in egocentric space (Grafton et al. 1998; Hazeltine 2002). As pointed out above, the learned sequence representation appears to include abstract information about the goals or intentions associated with the component movements. “Mirror” neurons, which respond whenever the animal performs or observes others perform particular actions, might be well-suited for providing goal-based representations that can guide behavior regardless of how the actions are actually implemented. Mirror neurons were originally identified in the premotor cortex (di Pellegrino et al. 1992); more recently, similar profiles are observed in parietal neurons (Gallese, Fadiga, Fogassi, & Rizzolatti, in press). These regions may serve to represent the constituent actions contained in the learned sequence. In this view, circuits involving the SMA, along with the basal ganglia (see below), serve to link representations encoded in the parietal cortex. A similar arrangement may exist with the PFC and premotor cortex, with the former structure providing the associations that bind actions represented by the latter structure. However, learning within PFC-premotor circuit may be easily disrupted by secondary tasks that demand executive control and working memory processes. Encoding within the parietal lobe, in contrast, appears to be robust across a range of experimental conditions.

## 9. Basal ganglia

The basal ganglia seem ideally suited for performing the integrative operations required for sensorimotor learning. The striatum, the principal input region of the basal ganglia, is massively innervated by axons originating in many cortical regions, especially the parietal and frontal lobes. The basal ganglia output projects to thalamic nuclei that in turn innervate cortical regions including motor, premotor, and prefrontal regions of the frontal lobe. Moreover, the dopaminergic pathways within the basal ganglia have been hypothesized to be a critical component of a behavioral reinforcement system. The capability to selectively reinforce particular actions over others is a necessary component of learning.

The PET SRT studies have consistently found increased blood flow in the basal ganglia during dual- and single-task learning. There are two notable differences between these two conditions. First, within the striatum, the activation during dual-task learning (Grafton et al. 1995; Hazeltine et al. 1997) is considerably inferior to that observed during single-task learning (Doyon et al. 1996; Grafton et al. 1995; Rauch et al. 1995). Second, dual-task learning is associated with activation of the left side of the basal ganglia whereas single-task learning is associated with activation on the right side, matching the laterality effects observed in the cortex.

Patients with basal ganglia pathology are impaired on a range of motor tasks, especially those that entail sequential movements (Agostino, Berardelli, Formica, Accornero, & Manfredi 1992; Harrington & Haaland 1991; Roy, Saint-Cyr, Taylor, & Lang 1993). While these deficits may reflect a fundamental problem in motor control, it has also been proposed that the basal ganglia are essential for the acquisition of novel movement patterns. Indeed, some theorists have proposed a generalized role for the basal ganglia in procedural learning (Mishkin, Malamut, & Bachevalier 1984), encompassing both the acquisition of movement sequences and implicit cognitive routines (Knowlton, Mangels, & Squire 1996; Saint-Cyr, Taylor, & Lang 1988). Lesion studies in animals also support the hypothesis that the basal ganglia are essential for the performance of movement sequences (e.g., Berridge & Whishaw 1992).

Patients with degenerative disorders of the basal ganglia, have been tested on the SRT task in a number of studies, always under single-task conditions. In general, the results indicate a learning deficit, even when performance impairments are taken into consideration. Patients with both Huntington's disease (Willingham & Koroshetz 1993) and Parkinson's disease have been found to show reduced learning, although the magnitude of the deficit varies across studies (Ferraro, Balota, & Connor 1993; Jackson, Jackson, Harrison, Henderson, & Kennard 1995; Pascual-Leone et al. 1993). Shin and Ivry (in preparation) introduced a variant of the SRT task in which the timing between the successive visual stimuli was manipulated to form a secondary, temporal sequence. Three different stimulus-response intervals were employed, with the length of this sequence identical to that used for the visuospatial sequence. Patients with Parkinson's disease were able to learn both the spatial and temporal sequence (when probed separately), but unlike control participants, they failed to integrate the two streams of information into a multidimensional sequence.

These findings suggest two possible roles for the basal ganglia in sequence encoding. First, learned associations may be formed within the basal ganglia. Sequence-specific cells have been identified in the striatum of the monkey (Kermadi, Jurquet, Arzi, & Joseph 1993). Computational models have emphasized that the dopaminergic pathways of the basal ganglia provide a critical reinforcement signal for the development of task-relevant associations (Berns & Sejnowski 1998; Hikosaka 1993; Houk, Adams, & Barto 1995; Schultz et al. 1995).

Second, a failure of sequence learning may be related to a problem in set shifting rather than learning per se (Cools 1980; Hayes, Davidson, Keele, & Rafal 1998; Robertson & Flowers 1990). Performance of learned movement sequences may require a shifting operation to schedule the successive elements or groups of elements in the appropriate order. Single-cell recordings in the basal ganglia indicate that this structure may play a role in scheduling successive actions. Chevalier and Deniau (1990) examined the relationship of activity in the superior colliculus and the inhibitory afferent signal these neurons receive from the basal ganglia. Discharge patterns in the colliculus were correlated with eye movements, and this activity was predictive of eye movement parameters well in advance of movement onset. However, the onset of movement occurred after the basal ganglia input was itself inhibited. Thus, within this system, the basal ganglia operate as a gating mechanism, in which movement plans represented in the colliculus are released through disinhibition (see also, Berns & Sejnowski 1996).

Brotchie, Iansek, and Horne (1991) have offered a related proposal based on the interaction between the basal ganglia and supplementary motor area during the generation of sequential arm movements and have outlining how shifting may be part of a system for motor priming. When making a series of wrist movements, an increase in the inhibitory output from the basal ganglia to the thalamus is observed near the end of each submovement. Interestingly, these bursts are only evident during the production of learned movements, again underscoring the role of the basal ganglia in learning. However, it is unclear if the basal ganglia have learned the movement sequence per se, or whether they are providing a shifting cue to a sequence represented in other, cortical structures. The inhibitory process may only be triggered when an input signal is provided to the basal ganglia indicating the next element in a learned sequence. In the Brotchie et al. task, it is possible that the SMA provides a representation of the learned movement sequence (see Romo & Schultz 1992), and the basal ganglia provides the signals to allow the animal to shift from one element to the next. That is, increasing activation within SMA, held in check by the basal ganglia, could serve as the basis for motor priming. The failure of the Parkinson patients to integrate temporal and spatial information may reflect a similar problem in scheduling a series of responses (Shin & Ivry, in preparation). The fact that the patients exhibited learning of the spatial sequence would be consistent with the idea that the sequence representation itself is not dependent on the basal ganglia.

The basal ganglia may interact in a similar manner with the PFC and SMA. This proposal is plausible given that similar connections exist between these latter structures and the basal ganglia (e.g., Alexander, Crutcher, & DeLong 1990). It is noteworthy that the dual-task studies activate both the SMA and putamen, whereas the single-task studies activate PFC and more superior regions of the striatum. This pattern converges with anatomical data showing that the PFC has relatively more connections to the caudate than putamen, and the reverse holds for SMA (Bates & Goldman-Rakic 1993). Thus, in a generalized way, the basal ganglia may implement a switching operation to move from one sequence chunk to the next with different regions recruited for the two sequence learning systems.

Clues about the computational role of the basal ganglia are provided by the observation that this structure is activated across a range of SRT tasks involving both implicit and explicit knowledge, and both single- and dual-task conditions. The widespread observations of activation here suggest that basal ganglia may perform an operation that is tapped by both attentional and non-attentional learning systems. One possibility is that the basal ganglia serve to reinforce associations formed between sequence elements as training progresses. A related hypothesis focuses on the structure's role in controlling internally generated responses (e.g., Goldberg 1985; Romo & Schultz 1992). Prior to learning, movements are triggered by the appearance of unanticipated, external visual cues. At this point, neither the SMA nor PFC has extracted a representation of the sequence of responses or of stimuli. Such representations emerge over the course of learning, entailing a shift from externally guided to internally guided control processes. With basal ganglia dysfunction, one would expect to see a dominance of externally guided processes as well as a loss of benefit from predictable conditions. Evidence for both of these behavioral phenomena have been reported in studies of Parkinson's disease (see Georgiou et al. 1994; Jahanshahi, Brown, & Marsden 1992; Majsak, Kaminski, Gentile, & Flanagan 1998). As noted above, the apparent impairment in SRT learning in Parkinson patients may reflect the fact that the keypresses are primarily being evoked in response to the visual signals rather than by anticipatory priming on either the response or stimulus side. Priming would be impaired due to slowness in switching operations, not by lack of sequence learning per se. In support of such argument, Parkinson patients perform comparably to control subjects when tracking a random visual cue, but, unlike controls, they show little improvement when the cue follows a predictable course (Gabrieli, Stebbins, Singh, Willingham, & Goetz 1997; Henderson & Goodrich 1993).

## 10. Cerebellum

The cerebellum has traditionally been viewed as central in the production and acquisition of skilled movement. Patients with cerebellar lesions have difficulty in terminating a pointing movement at a desired location and are especially impaired in movements that require the coordination across different muscle groups. Many experiments have demonstrated that the acquisition of new motor skills is disrupted (e.g., Ghez 1991; Ito 1984; Sanes, Dimitrov, & Hallett 1990; Thach, Goodkin, & Keating 1992). Studies show that adaptation to new sensorimotor mappings is either absent or slower to emerge following cerebellar damage (e.g., Martin, Keating, Goodkin, Bastian, & Thach 1996; Ojakangas & Ebner 1994; Raymond, Lisberger, & Mauk 1996; Thompson 1990).

Some PET studies of motor learning are consistent with the view of the cerebellum as part of a motor learning system. However, there is a striking difference in the pattern of activation general observed here: Unlike the other areas described above, improved performance is typically associated with a *decrease* in cerebellar activation. Jenkins et al. (1994) found that cerebellar activation was greater when subjects were engaged in explicitly acquiring a new sequence compared to when they were performing a well-learned sequence. Friston, Frith, Passingham, Liddle, and Frackowiak (1992) had subjects alternate between periods of sequential finger movements and rest. Whereas the magnitude of motor cortex activity was constant across the movement phases, the difference between movement and rest within the cerebellum diminished as subjects became more practiced. A decrease in cerebellar activation has also been observed in a non-motor learning task (Raichle et al. 1994).

The SRT studies present a murkier picture. Consistent with the other imaging studies, Hazeltine et al. (1997) observed decreases in cerebellar activation over the learning blocks in both their single- and dual-task conditions, although the foci were quite distant from one another. However, other studies have either failed to find significant changes in the cerebellum (Grafton et al. 1995) or found cerebellar activation to be greater during the production of learned sequences compared to random blocks (Doyon et al. 1996; Rauch et al. 1995).

In contrast to the imaging results, neuropsychological studies have found consistent and striking learning deficits in patients with bilateral lesions due to cerebellar atrophy (Doyon et al. 1998; Pascual-Leone et al. 1993) and patients with unilateral focal lesions (Gomez-Beldarrain, Garcia-Monco, Rubio, & Pascual-Leone 1998; Molinari et al. 1997). For example, Molinari et al. (1997, Experiment 2) found minimal reduction in RT during sequence training and no increase with the reintroduction of a random block. Interestingly, the deficit was comparable for both the ipsilesional and contralesional hands (but see Gomez-Beldarrain et al. 1998). Further evidence that the problem cannot be attributed to the patients' motor impairments comes from experiments in which the subjects simply watch the visual displays. Here, too, the patients are unable to learn the sequence (Molinari et al. 1997; Pascual-Leone et al. 1993).

All of the neuropsychological findings have emerged from studies using single-task conditions, indicating the cerebellum is involved in the attentional learning system, perhaps through its connections with premotor and prefrontal cortex (Goldberg 1985; Middleton & Strick 1994). There are some reports that individuals with cerebellar damage are impaired on tests of explicit knowledge of sequence learning (Pascual-Leone et al. 1993). Thus, the contribution of the cerebellum to learning on this task may not be restricted to fine-tuning motor commands but may also involve encoding more abstract representations. PET studies on a wide range of motor and non-motor tasks have frequently reported correlated patterns of activation in prefrontal cortex and cerebellum (see Fiez et al. 1996). How might the cerebellum contribute to this learning circuit? One conjecture is that the cerebellum is essential when the sequence is sufficiently complex that subjects must monitor their current place within the sequence and retain placeholders for a series of forthcoming responses (Inhoff, Diener, Rafal, & Ivry 1989; Pascual-Leone et al. 1993). The cerebellum might be well suited to perform this placeholding function given its central role in the representation of temporal information (see Ivry 1996).

Along these lines, Shin and Ivry (in preparation) observed that patients with cerebellar lesions not only fail to learn the spatial sequence in their SRT task, but also showed no evidence of incidental temporal sequence learning. Perhaps the timing deficit contributes to the spatial sequence learning deficit: The mechanisms required for building associations between successive events may not only require that the events are contiguous, but also that the temporal relationships remain relatively constant. Non-temporal associations may become difficult to form when the representation of the temporal relationships between the events is noisy.

## 11. Summary

Neural activity related to sequence learning has been observed in regions spanning nearly the entire brain, although in any given study, the extent of the activation is more limited. We summarize the findings reviewed in this paper in Table 1. As a whole, the data suggest that distinct sets of neural regions are recruited for sequence encoding depending on the task conditions. Regions within the prefrontal and premotor cortex become activated during sequence learning under single-task conditions when attention can be devoted without interruption to the SRT task. However, there is considerable evidence that prefrontal activation can be observed in the absence of explicit knowledge. Nonetheless, the prefrontal and premotor cortices may belong to a learning system that operates only when attention tracks the sequential information in an uninterrupted fashion. These regions do not appear to be involved with sequence learning when attention must be allocated to a secondary task. Under such dual-task conditions, the SMA and parietal regions show learning-related activation, although they have also been observed in some studies using single-task conditions. The SMA and parietal regions, along with the basal ganglia, likely represent major components of an implicit learning system that is capable of encoding a series of responses under a broad range of task demands.

Despite the many imaging studies employing variants of the SRT task, characterizing the specific computational role of these regions remains controversial. Activation can be interpreted as reflecting many processes, including perceptual or motor priming, the representation of individual actions, and, of course, sequence encoding. This paper has offered some preliminary hypotheses about the cognitive operations performed by the neural structures identified across a range of sequence learning studies. Most of these proposals have relied considerably on data from behavioral SRT experiments as well as imag-

**Table 1.** Summary of the findings from the imaging studies of sequence learning. Xs indicate that increased activity was identified in the comparison indicated in the third column. Abbreviations: MC = Motor cortex, PM = Premotor cortex, SMA = supplementary motor area, PFC = prefrontal cortex, ACC = Anterior cingulate cortex, P = Parietal, T = Temporal, BG = Basal ganglia, C = Cerebellum.

| | Aware | Task/Comparison | MC | PM | SMA | PFC | ACC | P | T | BG | C |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dual-task SRT Tasks** | | | | | | | | | | | |
| Grafton et al. (1995) | No | correlational + | X | | X | X | | X | | X | |
| Grafton et al. (1998) | No | correlational + | X | | X | | X | X | | | |
| | | Retrieval - Random | X | | | | X | X | | | |
| Hazeltine et al. (1997) | No | correlational + | X | | X | | | X | | X | |
| **Single-task SRT tasks** | | | | | | | | | | | |
| Berns et al. (1997) | No | correlational + | | | | X | | X | | | |
| | | correlational - | | X | | | X | | | X | |
| Doyon et al. (1996) | Mix | Learned – Random | | | | | X | X | | X | X |
| | | Explicit – Learned | | | | X | | | | | |
| | Yes | Explicit – New | | | | X | | | | | X |
| | | Learned – New | | | | | | | | | X |
| Eliassen et al. (2001) | Yes | Learned – Random | | | | X | | X | | | X |
| | Yes | Late learning – Early | | | | X | | X | | | |
| Grafton et al. (1995) | Mix | correlational + | | X | | X | | X | X | X | |
| Hazeltine et al. (1997) | Mix | correlational + | | X | | X | X | X | X | | |
| Honda et al. (1998) | No | Implicit, correlational | X | | | | | | | | |
| | Yes | Developing Explicit, corr. | | X | X | | | X | | | |
| | Yes | Post-Explicit, correlational | X | | X | | | | | | |
| | Yes | Explicit accuracy,corr. | | X | X | X | | X | | | |
| Rauch et al. (1995) | Yes | Explicit – Random | | | NA | X | | X | | | X |
| | | Explicit – Implicit | | | | | | X | | | |
| | | Implicit – Random | | X | | | | X | | X | |
| Toni et al. (1998) | Yes | correlational + (40 mins) | | | X | | | X | | | |
| | | correlational – (40 mins) | X | X | | X | X | X | | X | X |
| **Trial-and-Error Explicit Learning** | | | | | | | | | | | |
| Ghiladri et al. (2000) | Yes | New – Prelearned | | X | | X | X | X | | | X |
| Jenkins et al. (1994) | Yes | Prelearned - New | | | X | | X | X | X | | |
| | Yes | New - Prelearned | | X | | X | X | X | | | X |
| Jueptner et al. (1997a) | Yes | New - Prelearned | | X | | X | X | X | | X | X |
| | Yes | Attended - Prelearned | | | | X | X | | | | |
| | Yes | New - Attended | | X | | X | X | X | | X | X |
| **Sequence Performance** | | | | | | | | | | | |
| Catalan et al. (1998) | Yes | Complexity w/ Explicit | X | X | | | | X | | | X |
| Harrington et al. (2000) | Yes | Sequence – Repetition | | X | | | | X | | | |
| Karni et al. (1995) | Yes | Explicit (MC only) | X | NA | NA | NA | NA | NA | NA | NA | NA |
| Sadato et al. (1996) | Yes | Complexity w/ Explicit | | X | | | | X | | | X |

ing studies of seemingly unrelated tasks. In this way, the theorizing about the neural substrate of sequence learning is connected to diverse topics in psychology and neuroscience. Progress here will both require and produce advances in our understanding of memory, attention, motor control and consciousness.

## References

Agostino, R., Berardelli, A., Formica, A., Accornero, N., & Manfredi, M. (1992). Sequential arm movements in patients with Parkinson's disease, Huntington's disease and dystonia. *Brain*, *115*, 1481–1495.

Alexander, G. E., Crutcher, M. D., & DeLong, M. R. (1990). Basal ganglia thalamo-cortical circuits: Parallel substrates for motor control, oculomotor, "prefrontal" and "limbic" functions. *Progress in Brain Research*, *85*, 119–146.

Barbas, H. (1988). Anatomic organization of basoventral and mediodorsal visual recipient prefrontal regions in the rhesus monkey. *Journal of Comparative Neurology*, *276*, 313–342.

Barbas, H. & Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *286*, 353–375.

Barone, P. & Joseph, J. P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research*, *78*, 447–464.

Bates, J. F. & Goldman-Rakic, P. S. (1993). Prefrontal connections of medial motor areas in the rhesus monkey. *Journal of Comparative Neurology*, *336*, 211–228.

Berns, G. S., Cohen, J. D., & Mintun, M. A. (1997). Brain regions responsive to novelty in the absence of awareness. *Science*, *276*, 1272–1275.

Berns, G. S. & Sejnowski, T. J. (1996). How the basal ganglia make decisions. In A. R. Damasio (Ed.), *Neurobiology of Decision-Making* (pp. 101–113). New York: Springer-Verlag.

Berns, G. S. & Sejnowski, T. J. (1998). A computational model of how the basal ganglia produce sequences. *Journal of Cognitive Neuroscience*, *10*, 108–121.

Berridge, K. C. & Whishaw, I. Q. (1992). Cortex, striatum, and cerebellum: Control of serial order in a grooming task. *Experimental Brain Research*, *90*, 275–290.

Brotchie, P., Iansek, R., & Horne, M. K. (1991). Motor function of the monkey globus pallidus: 2. Cognitive aspects of movement and phasic neuronal activity. *Brain*, *114*, 1685–1702.

Carey, D. P. (2001). Do action systems resist visual illusions? *Trends in Cognitive Science, 5*, 109–113.

Casey, B. J., Thomas, K. M., Welsh, T. F., Badgaiyan, R. D., Eccard, C. H., Jennings, J. R., & Crone, E. A. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences, USA*, *97*, 8728–8733.

Catalan, M. J., Honda, M., Weeks, R. A., Cohen, L. G., & Hallett, M. (1998). The functional neuroanatomy of simple and complex sequential finger movements: A PET study. *Brain*, *121*, 253–264.

Chevalier, G. & Deniau, J. M. (1990). Disinhibition as a basic process in the expression of striatal functions. *Trends in Neuroscience*, *13*, 277–280.

Clark, V. P., Parasuraman, R., Keil, K., Kulansky, R., Fannon, S., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1997). Selective attention to face identity and color studied with fMRI. *Human Brain Mapping, 5*, 293–297.

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

Cools, A. R. (1980). Role of neostriatal dopaminergic activity in sequencing and selecting behavioral strategies: Facilitation of processes involved in selecting the best strategy in a stressful situation. *Behavioral Brain Research, 1*, 361–368.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience*, *11*, 2383–2402.

Crammond, D. J. (1996). Motor imagery: Never in your wildest dreams. *Trends in Neuroscience*, *20*, 54–57.

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

Decety, J., Kawashima, R., Gulyas, B., & Roland, P. (1992). Preparation for reaching: a PET study of the participating structures in the human brain. *NeuroReport, 3*, 761–764.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J. C., & Fazio, F. (1994). Mapping motor representations with positron emission tomography. *Nature*, *371*, 600–602.

Deiber, M. P., Passingham, R. E., Colebatch, J. G., Friston, K. J., Nixon, P. D., & Frackowiak, R. S. J. (1991). Cortical areas and the selection of movement: A study with positron emission tomography. *Experimental Brain Research*, *84*, 393–402.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, USA*, *93*, 13494–13499.

Desimone, R. & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*, 176–180.

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

Doyon, J., LaForce, R., Bouchard, G., Gaudreau, D., Roy, J., Poirier, M., Bedard, P. J., Bedard, F., & Bouchard, J.-P. (1998). Role of the striatum, cerebellum, and frontal lobes in the automatization of a repeated visuomotor sequence of movements. *Neuropsychologia*, *36*, 625–641.

Doyon, J., Owen, A. M., Petrides, M., Sziklas, V., & Evans, A. C. (1996). Functional anatomy of visuomotor skill learning in human subjects examined with Positron Emission Tomography. *European Journal of Neuroscience, 9*, 637–648.

Eliassen, J. C., Souza, T., & Sanes, J. N. (2001). Human brain activation accompanying explicitly directed movement sequence learning. *Experimental Brain Research*, *141*, 269–280.

Ferraro, F. R., Balota, D. A., & Connor, L. T. (1993). Implicit memory and the formation of new associations in nondemented Parkinson's disease individuals and individuals with senile dementia of the Alzheimer type: A serial reaction time (SRT) investigation. *Brain and Cognition*, *21*, 163–180.

Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *Journal of Neuroscience*, *16*, 808–822.

Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Driver, J., Frackowiak, R. S. J., & Dolan, R. J. (1999). The neural consequences of conflict between intention and the senses. *Brain*, *122*, 497–512.

Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transactions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 567–584.

Friston, K. J., Frith, C. D., Passingham, R. E., Liddle, P. F., & Frackowiak, R. S. J. (1992). Motor practice and neurophysiological adaptation in the cerebellum: A positron tomography study. *Proceedings of the Royal Society of London. B: Biology*, *248*, 223–228.

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*, 347–351.

Gabrieli, J. D. E., Stebbins, G. T., Singh, J., Willingham, D. B., & Goetz, C. G. (1997). Intact mirror-tracing and impaired rotary-pursuit skill learning in patients with Huntington's disease: Evidence for dissociable memory systems in skill learning. *Neuropsychology*, *11*, 272–281.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (in press). Action representation and the inferior parietal lobule. In W. Prinz and B. Hommel (Eds.), *Attention and Performance* (Vol. XIX). Oxford: Oxford University Press.

Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitoty control: An event-related functional MRI study. *Proceedings of the National Academy of Sciences, USA*, *96*, 8301–8306.

Georgiou, N., Bradshaw, J. L., Iansek, R., Phillips, J. G., Mattingley, J. B., & Bradshaw, J. A. (1994). Reduction in external cues and movement sequencing in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *57*, 368–370.

Ghez, C. (1991). The Cerebellum: Chapter 41. In E. Kandel, J. H. Schwarz, and T. M. Jessell (Eds.), *Principles of Neural Science: 3rd Edition* (pp. 626–646). New York: Elsevier.

Ghiladri, M.-F., Ghez, C., Dhawan, V., Moeller, J., Mentis, M., Nakamura, T., Antonini, A., & Eidelberg, D. (2000). Patterns of regional brain activation associated wiith different forms of motor learning. *Brain Research*, *871*, 127–145.

Glover, S. R. & Dixon, P. (2001). Dynamic illusion effects in a reaching task: Evidence for separate visual representations in the planning and control of reaching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 560–572.

Goldberg, G. (1985). Supplementary motor area structure and function: Review and hypothesis. *Behavioral and Brain Sciences, 8*, 567–616.

Gomez-Beldarrain, M., Garcia-Monco, J. C., Rubio, B., & Pascual-Leone, A. (1998). Effect of focal cerebellar lesions on procedural learning in the serial reaction time task. *Experimental Brain Research*, *120*, 25–30.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Darey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, *349*, 154–156.

Goschke, T. (1998). Implicit learning of perceptual and motor sequences: Evidence for independent learning systems. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learning*. (p. 401–444). Thousand Oaks, CA: Sage Publications.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience, 7*, 497–510.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1998). Abstract and effector-specific representations of motor sequences identified with PET. *Journal of Neurophysiology*, *18*, 9420–9428.

Grafton, S. T., Mazziotta, J. C., Presty, S., Friston, K. J., Frackowiak, R. S. J., & Phelps, M. E. (1992). Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *Journal of Neuroscience*, *12*, 2542–2548.

Grafton, S. T., Mazziotta, J. C., Woods, R. P., & Phelps, M. E. (1992). Human functional anatomy of visually guided finger movements. *Brain*, *115*, 565–587.

Haffenden, A. M. & Goodale, M. A. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neuroscience*, *10*, 122–136.

Halsband, U., Ito, M., Tanji, J., & Freund, H.-J. (1993). The role of premotor cortex and the supplementary motor area in the temporal control of movement in man. *Brain*, *16*, 243–266.

Harrington, D. I. & Haaland, K. Y. (1991). Sequencing in Parkinson's disease: Abnormalities in programming and controlling movement. *Brain*, *114*, 99–115.

Harrington, D. I., Rao, S. C., Haaland, K. Y., Bobholz, J. A., Mayer, A. R., Binder, J. R., & Cox, R. W. (2000). Specialized neural systems underlying representations of sequential movements. *Journal of Cognitive Neuroscience*, *12*, 56–77.

Hayes, A., Davidson, M., Keele, S. W., & Rafal, R. D. (1998). Toward a functional analysis of the basal ganglia. *Journal of Cognitive Neuroscience*, *10*, 178–198.

Hazeltine, E. (2002). The representational nature of sequence learning: Evidence for goal-based codes. In W. Prinz and B. Hommel (Eds.), *Attention and Performance* (Vol. XIX, pp. 673–689). Oxford: University.

Hazeltine, E., Grafton, S. T., & Ivry, R. B. (1997). Attention and stimulus characteristics determine the locus of motor sequence encoding: A PET study. *Brain*, *120*, 123–140.

Hazeltine, E., Ivry, R. B., & Chan, D. (in preparation). Attention and the modularity of implicit learning.

Hazeltine, E., Poldrack, R., & Gabrieli, J. D. E. (2000). Neural activation during response competition. *Journal of Cognitive Neuroscience*, *12*(Supplement 2), 118–129.

Heilman, K. M., Rothi, L. J., & Valenstein, E. (1982). Two forms of ideomotor apraxia. *Neurology*, *32*, 342–346.

Henderson, L. & Goodrich, S. J. (1993). Simple reaction time and predictive tracking in Parkinson's Disease: Do they converge on a single, fixed impairment of preparation? *Journal of Motor Behavior*, *25*, 89–96.

Hikosaka, O. (1993). Role of the basal ganglia in motor learning: A hypothesis. In L. R. Squire, M. E. Raichle, D. I. Perrett, and M. Fukuda (Eds.), *Brain Mechanisms of Perception and Memory: From Neuron to Behavior* (pp. 497–513). New York: Oxford University Press.

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., & Doya, K. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neuroscience*, *22*, 464–471.

Honda, M., Dieber, M.-P., Ibanez, V., Pascual-Leone, A., Zhuang, P., & Hallett, M. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning: A PET study. *Brain*, *121*, 2159–2173.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, and D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 249–270). Cambridge, MA: MIT Press.

Inhoff, A. W., Diener, H. C., Rafal, R. D., & Ivry, R. B. (1989). The role of cerebellar structures in the execution of serial movements. *Brain*, *112*, 565–581.

Ito, M. (1984). *The Cerebellum and Neural Control*. New York: Raven Press.

Ivry, R. B. (1996). The representation of temporal information in perception and motor control. *Current Opinion in Neurobiology, 6*, 851–857.

Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard, C. R. (1995). Serial reaction time learning and Parkinson's disease: Evidence for a procedural learning deficit. *Neuropsychologia*, *33*, 577–593.

Jahanshahi, M., Brown, R. G., & Marsden, C. D. (1992). Simple and choice reaction time and the use of advance information for motor preparation in Parkinson's disease. *Brain*, *115*, 529–564.

Jeannerod, M., Decety, J., Perani, D., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J. C., & Fazio, F. (1994). Mapping motor representations with positron emission tomography. *Nature*, *371*, 600–602.

Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S. J., & Passingham, R. E. (1994). Motor sequence learning: A study with positron emission tomography. *Journal of Neuroscience*, *14*, 3775–3790.

Jiménez, L. & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 236–259.

Jordan, M. I. (1995). The organization of action sequences: Evidence from a learning task. *Journal of Motor Behavior*, *27*, 179–192.

Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., & Passingham, R. E. (1997). Anatomy of motor learning. I. Frontal cortex and attention to action. *Journal of Neurophysiology*, *77*, 1313–1324.

Karni, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature*, *377*, 155–158.

Keele, S. W., Ivry, R. B., Mayr, U., Hazeltine, E., & Heuer, H. (under review). The cognitive and neural architecture of sequence representation.

Keele, S. W., Jennings, P., Jones, S., Caulton, D., & Cohen, A. (1995). On the modularity of sequence representation. *Journal of Motor Behavior*, *27*, 17–30.

Kermadi, I., Jurquet, Y., Arzi, M., & Joseph, J. P. (1993). Neural activity in the caudate nucleus of monkeys during spatial sequencing. *Experimental Brain Research*, *94*, 352–356.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.

Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., & Miyashita, Y. (1998). No-go dominant brain activation in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *European Journal of Neuroscience*, *10*, 1209–1213.

Lindemann, P. G. & Wright, C. E. (1998). Skill acquisition and plans for actions: Learning to write with your other hand. In S. Sternberg and D. Scarborough (Eds.), *Invitation to Cognitive Science* (Vol. 4, pp. 523–584). Cambridge, MA: MIT Press.

Lueck, C. J., Zeki, S., Friston, K. J., Deiber, M. P., Cope, P., Cunningham, V. J., Lammertsma, A. A., Kennard, C., & Frackowiak, R. S. J. (1989). The colour centre in the cerebral cortex of man. *Nature*, *340*, 386–389.

Majsak, M. J., Kaminski, T., Gentile, A. M., & Flanagan, J. R. (1998). The reaching movements of patients with Parkinson's disease under self-determined maximal speed and visually cued conditions. *Brain*, *121*, 755–766.

Martin, T. A., Keating, J. G., Goodkin, H. P., Bastian, A. J., & Thach, W. T. (1996). Throwing while looking through prisms. I. Focal olivocerebellar lesions impair adaptation. *Brain*, *119*, 1183–1198.

McCarthy, G., Blamire, A. M., Puce, A., Nobe, A. C., Bloch, G., Hyder, F., Goldman-Rakic, P. S., & Shulman, R. G. (1994). Functional magnetic resonance imaging of human prefrontal cortex activation during a spatial working memory task. *Proceedings of the National Academy of Sciences, USA*, *91*, 8690–8694.

Mesulam, M.-M. (1998). From sensation to cognition. *Brain*, *121*, 1013–1052.

Middleton, F. A. & Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, *266*, 4528–4561.

Milner, A. D. & Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: University Press.

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

Molinari, M., Leggio, M. G., Solida, A., Ciorra, R., Misciagna, S., Silveri, M. C., & Petrosini, L. (1997). Cerebellum and procedural learning: Evidence from focal cerebellar lesions. *Brain*, *120*, 1753–1762.

Mushiake, H., Inase, M., & Tanji, J. (1991). Neuronal activity in the primate premotor, supplementary, and precentral motor cortex during during visually guided and internally determined sequential movements. *Journal of Neurophysiology*, *66*, 705–718.

Nattkemper, D. & Prinz, W. (1997). Stimulus and response anticipation in a serial reaction time task. *Psychological Research*, *60*, 98–112.

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

Ojakangas, C. L. & Ebner, T. J. (1994). Purkinje cell complex spike activity during voluntary motor learning: Relationship to kinematics. *Journal of Neurophysiology*, *72*, 2617–2630.

Palmer, C. & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science*, *11*, 63–68.

Pascual-Leone, A., Grafman, J., Clark, K., Stewart, M., Massaquoi, S., Lou, J.-S., & Hallett, M. (1993). Procedural learning in Parkinson's disease and cerebellar degeneration. *Annals of Neurology*, *34*, 594–602.

Pascual-Leone, A., Grafman, J., & Hallett, M. (1994). Modulation of cortical motor output maps during development of implicit and explicit knowledge. *Science*, *263*, 1287–1289.

Pascual-Leone, A., Wassermann, E. M., Grafman, J., & Hallett, M. (1996). The role of dorsolateral prefrontal cortex in implicit procedural learning. *Experimental Brain Research*, *107*, 479–485.

Passingham, R. E. (1993). *The Frontal Lobes and Voluntary Action*. Oxford, England: Oxford University Press.

Passingham, R. E. (1996). Attention to action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*, 1473–1479.

Penfield, W. & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*, 389–443.

Perruchet, P. & Amorim, M. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 785–800.

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. K., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex, 4*, 8–26.

Rauch, S. L., Savage, C. R., Brown, H. D., Curran, T., Alpert, N. M., Kendrick, A., Fischman, A. J., & Kosslyn, S. M. (1995). A PET investigation of implicit and explicit sequence learning. *Human Brain Mapping, 3*, 271–286.

Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., Bush, G., Breiter, H. C., & Rosen, B. R. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping, 5*, 124–132.

Raymond, J. L., Lisberger, S. G., & Mauk, M. D. (1996). The cerebellum: A neuronal learning machine? *Science*, *272*, 1126–1131.

Reed, J. & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 585–594.

Rizzolatti, G. & Gentilucci, M. (1988). Motor and visual-motor functions of the premotor cortex. In P. Rakic and W. Singer (Eds.), *Neurobiology of Neocortex* (pp. 269–284). New York: Wiley & Sons.

Robertson, C. & Flowers, K. A. (1990). Motor set in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *53*, 583–592.

Roland, P. E., Larsen, B., Lassen, N. A., & Skinhoj, E. (1980). Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology*, *43*, 118–136.

Romo, R. & Schultz, W. (1992). Role of the primate basal ganglia and frontal cortex in the internal generation of movements: III. Neuronal activity in the supplementary motor area. *Experimental Brain Research*, *91*, 396–407.

Roy, E. A., Saint-Cyr, J., Taylor, A., & Lang, A. (1993). Movement sequencing disorders in Parkinson's disease. *International Journal of Neuroscience*, *73*, 183–194.

Sadato, N., Campbell, G., Ibanez, V., Deiber, M. P., & Hallett, M. (1996). Complexity affects regional cerebral blood flow change during sequential finger movements. *Journal of Neuroscience*, *16*, 2691–2700.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain*, *111*, 941–959.

Sanes, J. N., Dimitrov, B., & Hallett, M. (1990). Motor learning in patients with cerebellar dysfunction. *Brain*, *113*, 103–120.

Schlaug, G., Knorr, U., & Seitz, R. J. (1994). Inter-subject variability of cerebral activations in acquiring a motor skill: A study with positron emission tomography. *Experimental Brain Research*, *98*, 523–534.

Schmidtke, V. & Heuer, H. (1997). Task integration as a factor in secondary-task effects on sequence learning. *Psychological Research*, *60*, 53–71.

Schultz, W., Romo, R., Ljungberg, T., Mirenowicz, J., Hollerman, J. R., & Dickinson, A. (1995). Reward-related signals carried by dopamine neurons. In J. C. Houk, J. L. Davis, and D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 233–248). Cambridge, MA: MIT Press.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.

Shanks, D. R. & St. John, M. F. (1996). Implicit learning: What does it all mean? *Behavioral and Brain Sciences*, *19*, 557–558.

Shibasaki, H., Sadato, N., Lyshkow, H., Yonekura, Y., Honda, M., Nagamine, T., Suwazonol, S., Magata, Y., Ikeda, A., Miyazaki, M., Fukuyama, H., Asato, R., & Konishi, J. (1993). Both primary motor cortex and supplementary motor area play an important role in complex finger movement. *Brain*, *116*, 1387–1398.

Shin, J. C. & Ivry, R. B. (in preparation). Performance of patients with Parkinson's disease or cerebellar on spatial and temporal sequence learning.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Satoshi, M. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience, 7*, 337–356.

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

Tanji, J. & Shima, K. (1994). Role of supplementary motor area cells in planning several movements ahead. *Nature*, *371*, 413–416.

Thach, W. T., Goodkin, H. P., & Keating, J. G. (1992). The cerebellum and the adaptive coordination of movement. *Annual Review of Neuroscience*, *15*, 403–442.

Thompson, R. F. (1990). Neural mechanisms of classical conditioning in mammals. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *329*, 161–170.

Toni, I., Krams, M., Turner, R., & Passingham, R. E. (1998). The time course of changes during motor sequence learning: A whole-brain fMRI study. *NeuroImage, 8*, 50–61.

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In D. J. Engle, M. A. Goodale, and R. J. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Wachs, J., Pascual-Leone, A., Grafman, J., & Hallett, M. (1994). Intermanual transfer of implicit knowledge of sequential finger movements. *Neurology*, *44*(Supplement 2), A329.

Watanabe, J. & Iwai, E. (1991). Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. *Brain Research Bulletin*, *26*, 583–592.

Watson, R. T., Fleet, W. S., Gonzalez-Rothi, L., & Heilman, K. M. (1986). Apraxia and the supplementary motor area. *Archives of Neurology*, *43*, 787–792.

Willingham, D. B. & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. *Psychological Science*, *10*, 531–534.

Willingham, D. B. & Koroshetz, W. J. (1993). Evidence of dissociable motor skills in Huntington's disease patients. *Psychology*, *21*, 173–182.

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.

Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory and Cognition*, *28*, 366–375.

Wise, S. P., Di Pellegrino, G., & Boussaoud, D. (1996). The premotor cortex and nonstandard sensorimotor mapping. *Canadian Journal of Physiology and Pharmacology*, *74*, 469–482.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, *11*, 641–649.

# The cognitive neuroscience of implicit category learning

F. Gregory Ashby & Michael B. Casale*
University of California, Santa Barbara

There is much recent interest in the question of whether people have available a single category learning system or a number of qualitatively different systems. Most proponents of multiple systems have hypothesized an explicit, rule-based system and some type of implicit system. Although there has been general agreement about the nature of the explicit system, there has been disagreement about the exact nature of the implicit system. This chapter explores the question of whether there is implicit category learning, and if there is, what form it might take. First, we examine what the word "implicit" means in the categorization literature. Next, we review some of the evidence that supports the notion that people have available one or more implicit categorization systems. Finally, we consider the nature of implicit categorization by focusing on three alternatives: an exemplar memory-based system, a procedural memory system, and an implicit system that uses the perceptual representation memory system.

## 1. The cognitive neuroscience of implicit category learning

Categorization is the act of responding differently to objects and events in the environment that belong to separate classes or categories. It is a critical process that every organism must perform in at least a rudimentary form because it allows them to respond differently, for example, to nutrients and poisons, and to predators and prey.

Much of the recent categorization literature has focused on the question of whether people have available a single category learning system or a number of qualitatively different systems. For example, although the early literature was dominated by theories postulating a single system, a number of recent theories have proposed multiple category learning systems (Ashby, Alfonso-Reese, Turken, & Waldron 1998; Brooks 1978; Erickson & Kruschke 1998; Pickering 1997). Interestingly, many of these papers have hypothesized at least two similar systems: 1) an explicit, rule-based system that is tied to language function and conscious awareness, and 2) an implicit system that may not have access to conscious awareness. For example, Ashby et al. (1998) proposed a formal neuropsychological theory of multiple category learning systems called COVIS (COmpetition between Verbal and Implicit Systems), which assumes separate explicit (rule-based) and implicit (procedural learning-based) systems.

There is still much disagreement however. First, the proposal that there are multiple category learning systems is disputed. In particular, Nosofsky and his colleagues have argued that single system models can account for many of the phenomena that have been used to support the notion of multiple systems (Nosofsky & Johansen 2000; Nosofsky & Kruschke 2002; Nosofsky & Zaki 1998). Second, even among those researchers postulating separate explicit and implicit systems, there is disagreement about the nature of the implicit system. As mentioned above, Ashby et al. (1998) proposed a procedural-memory based implicit system (see also Ashby & Waldron 1999; Ashby, Waldron, Lee, & Berkman 2001). In contrast, several researchers have proposed that the implicit system is exemplar-memory based (Erickson & Kruschke 1998; Pickering 1997), and there have also been proposals that the perceptual representation memory system participates in implicit category learning (Ashby & Ell 2001; Knowlton, Squire et al. 1996; Reber, Stark, & Squire 1998).

This chapter explores the question of whether there is implicit category learning. First, we examine what is meant by explicit and implicit categorization. These are important questions because both terms are used somewhat differently in the categorization literature than in the memory literature. Next, we briefly review evidence supporting the notion that people have available one or more implicit categorization systems. Finally, we focus on two putative implicit category learning systems, one that uses procedural memory and one that uses the perceptual representation memory system.

## 2. What are explicit and implicit categorization?

### 2.1 Explicit categorization

Categorization processes are said to be explicit if they are accessible to conscious awareness. This would include traditional declarative memory processes that might be invoked when participants try to memorize responses associated with the various stimuli. However, it could also include simple rule-based strategies such as, "the stimulus belongs to category A if it is red, and it belongs to category B if it is blue."

One danger with equating explicit processing with conscious awareness is that this shifts the debate from how to define 'explicit' to how to define 'conscious awareness'. Ashby et al. (1998) suggested that one pragmatic solution to this problem is to adopt the criterion that category learning is explicit if the subject can verbally describe the categorization rule that he or she used. This definition works well in most cases, but it seems unlikely that verbalizability should be a requirement for explicit reasoning. For example, the insight displayed by Köhler's (1925) famous apes seems an obvious example of explicit reasoning in the absence of language. For now we will use the criterion of verbalizability for explicit category learning but ultimately, a theoretically motivated criterion for conscious awareness is needed.

One way to develop a theory of conscious awareness is by exploiting the relationship between awareness and working memory. For example, the contents of working memory are clearly accessible to conscious awareness. In fact, because of its close association to executive attention, a strong argument can be made that the contents of working memory *define* our conscious awareness. When we say that we are consciously aware of some object or event, we mean that our executive attention has been directed to that stimulus. Its representation in our working memory gives it a moment-to-moment permanence. Working memory makes it possible to link events in the immediate past with those in the present, and it allows us to anticipate events in the near future. All of these are defining properties of conscious awareness.

The association between working memory and the prefrontal cortex makes it possible to formulate cognitive neuroscience models of conscious awareness. The most influential such model was developed by Francis Crick and Christof Koch (Crick & Koch 1990, 1995, 1998). The Crick-Koch hypothesis states that one can have conscious awareness only of activity in brain areas that project directly to the prefrontal cortex.[1] For example, consider two brain areas X and Y. Suppose area X projects directly to the prefrontal cortex, but area Y projects only to area X (i.e., and not directly to the prefrontal cortex). If working memory and conscious awareness reside in prefrontal cortex, then we can be consciously aware of activity in area X because it can be loaded directly into working memory. On the other hand, if activity in area Y is transformed by area X before reaching prefrontal cortex and conscious awareness, then there is no way to be aware of activity in area Y – only of the transformed activity that leaves area X.

Primary visual cortex (Area V1) does not project directly to the prefrontal cortex, so the Crick-Koch hypothesis asserts that we cannot be consciously aware of activity in V1. Crick and Koch (1995, 1998) described evidence in support of this prediction. Of course, many other brain regions also do not project directly to the prefrontal cortex. For example, the basal ganglia do not project directly to the prefrontal cortex (i.e., they first project through the thalamus), so the Crick-Koch hypothesis predicts that we are not aware of activity within the basal ganglia. Memory theorists believe that the basal ganglia mediate procedural memories (Jahanshahi, Brown, & Marsden 1992; Mishkin, Malamut, & Bachevalier 1984; Saint-Cyr, Taylor, & Lang 1988; Willingham, Nissen, & Bullemer 1989), so the Crick-Koch hypothesis provides an explanation of why we don't seem to be aware of procedural (e.g., motor) learning.

In summary, although the Crick-Koch hypothesis offers a promising start, a complete theory of conscious awareness does not yet exist. Therefore, in this chapter we will adopt the operational definition that a categorization process is explicit if it can be described verbally.

## 2.2 Implicit categorization

During the past 10 years, about 120 articles have appeared in the psychological literature that discuss implicit category learning or implicit categorization, whereas the decade of the 1980s saw only about 20 such articles. This recent interest in implicit category learning has profoundly affected the categorization literature, and has formed bridges to the memory literature, where of course, the study of implicit processes have a long and rich history. Even so, a memory researcher interested in implicit categorization may be confused by how the term "implicit" is used in the categorization literature.

Many memory theorists adopt the strong criterion that a memory is implicit only if there is no conscious awareness of its details *and* there is no knowledge that a memory has even been stored (e.g., Schacter 1987). In a typical categorization task, these criteria are impossible to meet because trial-by-trial feedback is routinely provided. When an observer receives feedback that a response is correct, then this alone makes it obvious that learning has occurred, even if there is no internal access to the system that is mediating this learning. Thus, in category learning, a weaker criterion for implicit learning is typically used in which the observer is required only to have no conscious access to the nature of the learning, even though he or she would be expected to know that some learning has occurred.

The stronger criterion for implicit processing that has been adopted in much of the memory literature could be applied in unsupervised category learning tasks, in which no trial-by-trial feedback of any kind is provided. In the typical unsupervised task, observers are told the number of contrasting categories and are asked to assign stimuli to these categories, but are never told whether a particular response is correct or incorrect. Free sorting is a similar, but more unstructured task in which participants are not given feedback about the accuracy of their responses, nor are they even told the number of contrasting categories (e.g., Ashby & Maddox 1998). Thus, in both unsupervised and free sorting tasks there is no feedback that observers can use to infer that learning has occurred. As a result, these tasks are ideal for using the stricter criterion to test for implicit learning. Even so, to date the only learning that has been demonstrated in such tasks is explicit (Ashby, Queller, & Berretty 1999; Medin, Wattenmaker, & Hampson 1997).

## 3. Evidence for separate explicit and implicit category learning systems

### 3.1 Three different category learning tasks

Much of the data that has been used to argue for multiple category learning systems came from the observation that changing the nature of the contrasting categories that subjects were asked to learn sometimes qualitatively changed learning behavior. Ashby and Ell (2001) identified three different types of category structures that are often associated with such qualitative differences in performance. To anticipate our later discussion, in the next section we will argue that these three tasks load primarily on three different memory systems.

*Rule-based tasks* are those in which subjects can learn the category structures via some explicit reasoning process. In the most common applications, only one stimulus dimension is relevant, and the subject's task is to discover this relevant dimension and then to map the different dimensional values to the relevant categories. Figure 1 shows the stimuli and category structure of a recent rule-based task that used 8 exemplars per category (Waldron & Ashby

2001). The categorization stimuli were colored geometric figures presented on a colored background. The stimuli varied on four binary-valued dimensions: background color (blue or yellow; here denoted as light and dark gray, respectively), embedded symbol color (green or red; here denoted as black and white, respectively), symbol number (1 or 2), and symbol shape (square or circle). This yields a total of 16 possible stimuli. To create rule-based category structures, one dimension is selected arbitrarily to be relevant. The two values on that dimension are then assigned to the two contrasting categories.

An important property of rule-based category learning tasks is that the optimal rule is often easy to describe verbally (Ashby et al. 1998). As a result, subjects can learn the category structures via an explicit process of hypothesis testing (Bruner, Goodnow, & Austin 1956) or theory construction and testing (Murphy & Medin 1985). Virtually all standard neuropsychological categorization tasks are of this type – including the well known Wisconsin Card Sorting Test (Heaton 1981). Rule-based tasks, which have a long history in cognitive psychology, have been favored by proponents of the so-called classical theory of categorization, which assumes that category learning is the process of discovering the set of necessary and sufficient conditions that determine category membership (Smith & Medin 1981).

In the Figure 1 example, the explicit rule that perfectly separates the stimuli into the two categories is unidimensional. Although the optimal rule in rule-based tasks is often unidimensional, this is not a requirement. For example, a task is rule-based if the optimal rule is a conjunction of the form:

Respond A if the background is blue and the embedded symbol is round; otherwise respond B.

The critical criterion is that this rule is easy to describe verbally, and to learn through an explicit reasoning process. Note that according to this criterion, there is no limit on the complexity of the optimal rule in rule-based tasks. However, as the complexity of the optimal rule increases, its salience decreases and it becomes less likely that observers will learn the associated categories through an explicit reasoning process. In fact, Alfonso-Reese (1997) found that even simple conjunction rules have far lower salience than unidimensional rules. This does not mean that people can not learn conjunction rules. Only that they are unlikely to experiment with such rules unless feedback compels them in this direction. This discussion should make it clear that the boundary on what constitutes a rule-based task is fuzzy. Tasks in which the optimal rule is unidimensional are unambiguously rule-based (at least with separable stimulus dimensions), and tasks in which the optimal rule is significantly more complex than a conjunction rule are almost never rule-based. In

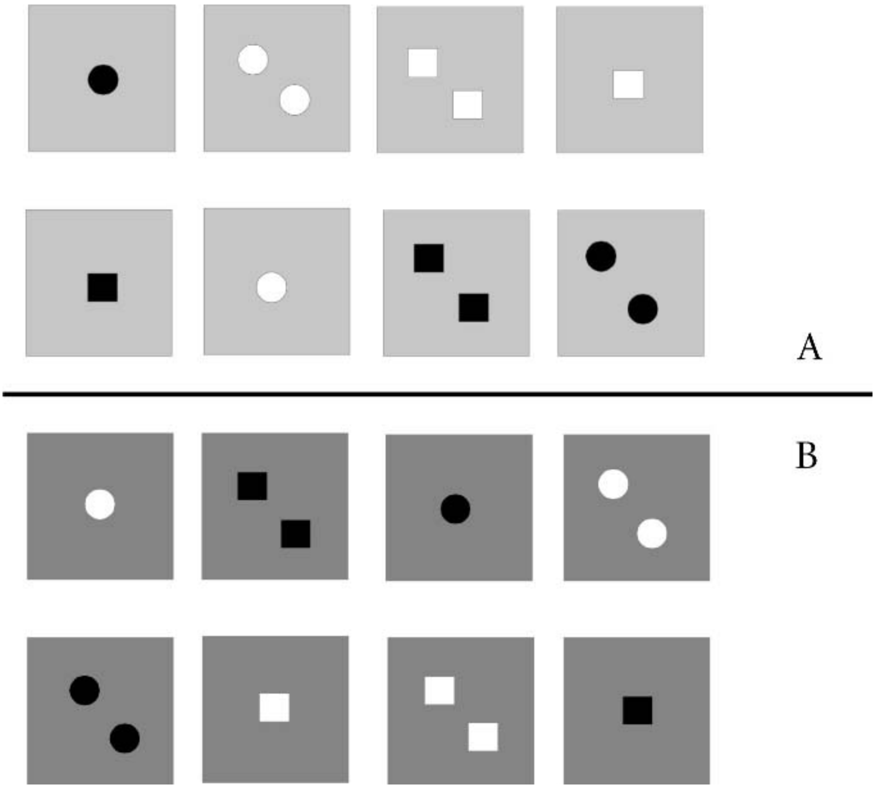


**Figure 1.** Category structure of a rule-based category learning task. The optimal rule is: Respond A if the background color is blue (depicted as light gray), and respond B if the background color is yellow (depicted as dark gray).

between, the classification is not so clear-cut. For this reason, the rule-based tasks we discuss in this chapter will all have a unidimensional optimal rule.

*Information-integration tasks* are those in which accuracy is maximized only if information from two or more stimulus components (or dimensions) is integrated at some pre-decisional stage (Ashby & Gott 1988). Perceptual integration could take many forms – from treating the stimulus as a Gestalt to computing a weighted linear combination of the dimensional values. However, a conjunction rule is a rule-based task rather than an information-integration task because separate decisions are first made about each dimension (e.g., small or large) and then the outcome of these decisions is combined (integration is not pre-decisional). In many cases, the optimal rule in information-integration tasks is difficult or impossible to describe verbally (Ashby et al. 1998). The neu-

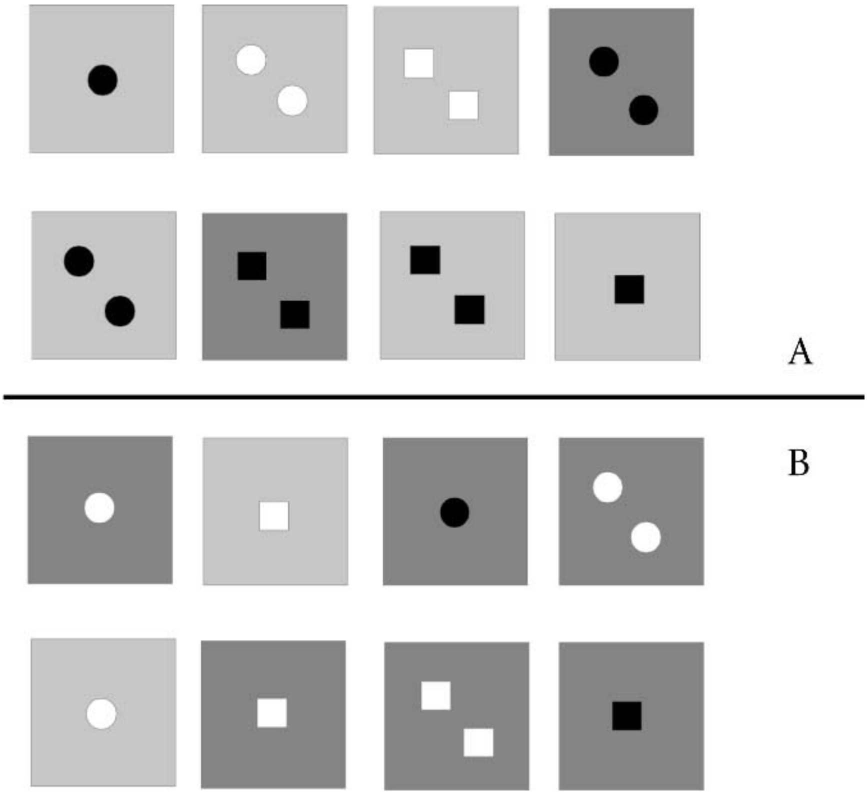


**Figure 2.** Category structure of an information integration category learning task with only a few exemplars in each category.

ropsychological data reviewed below suggests that performance in such tasks is qualitatively different depending on the size of the categories – in particular, when a category contains only a few highly distinct exemplars, memorization is feasible. However, when the relevant categories contain many exemplars (e.g., hundreds), memorization is less efficient.

Figure 2 shows the stimuli and category structure of a recent information-integration task that used only 8 exemplars per category (Waldron & Ashby 2001). The categorization stimuli are the same as in Figure 1. To create information-integration category structures, one dimension is arbitrarily selected to be irrelevant. For example, in Figure 2, the irrelevant dimension is symbol shape. Next, one level on each relevant dimension is arbitrarily assigned a value of +1 and the other level is assigned a value of 0. In Figure 2, a background color of blue (denoted as light gray), a symbol color of green

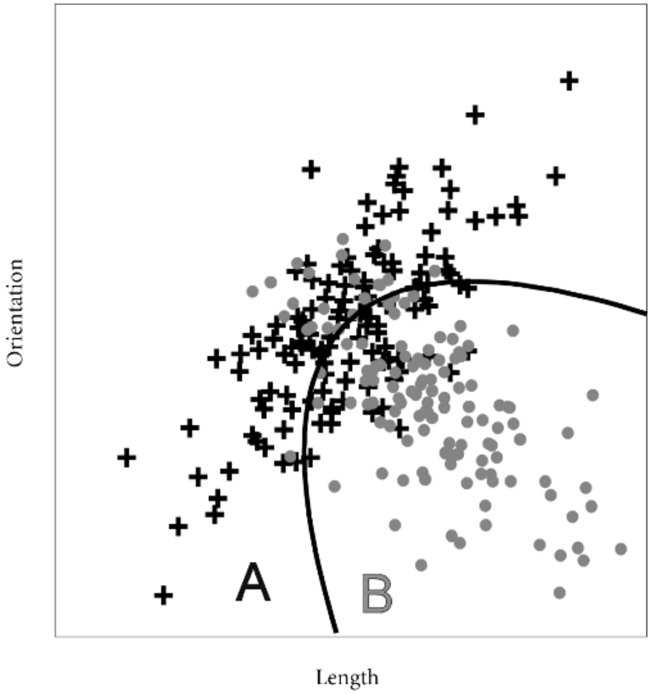


**Figure 3.** Category structure of an information integration category learning task with many exemplars per category. Each stimulus is a line that varies across trials in length and orientation. Every black plus depicts the length and orientation of a line in Category A and every gray dot depicts the length and orientation of a line in Category B. The quadratic curve is the boundary that maximizes accuracy.

(denoted as black), and a symbol number of 2 are all assigned a value of +1. Finally, the category assignments are determined by the following rule:

The stimulus belongs to category A if the sum of values on the relevant dimensions > 1.5; Otherwise it belongs to category B.

This rule is readily learned by healthy young adults, but even after achieving perfect performance, they can virtually never accurately describe the rule they used.[2]

Figure 3 is an abstract representation of the category structure of an information-integration task in which there are hundreds of exemplars in each category (developed by Ashby & Gott 1988). In this experiment, each stimulus is a line that varies across trials in length and orientation. Each cross in Figure 3 denotes the length and orientation of an exemplar in Category A and each dot denotes the length and orientation of an exemplar in Category B. The categories overlap, so perfect accuracy is impossible in this example. Even so, the quadratic curve is the boundary that maximizes response accuracy – that is, accuracy is maximized if subjects respond B to any stimulus falling inside the quadratic region (in the lower right quadrant), and A to any stimulus falling outside of this region. Note that such a rule is impossible to describe verbally. Many experiments have shown that, given enough practice, the performance of subjects in this task is well described by a quadratic decision boundary (e.g., Ashby & Maddox 1992; Maddox & Ashby 1993).

Information-integration tasks with few exemplars per category have been the favorites of exemplar theorists, who argue that categorization requires accessing the memory representations of every previously seen exemplar from each relevant category (e.g., Estes 1986, 1994; Medin & Schaffer 1978; Nosofsky 1986; Smith & Minda 2000). In contrast, decision bound theorists, who argue that category learning is a process of associating category labels with regions of perceptual space, have traditionally used information-integration tasks with many exemplars per category (e.g., Ashby & Maddox 1992; Maddox & Ashby 1993).

*Prototype distortion tasks* are a third type of category learning task in which each category is created by first defining a category prototype and then creating the category members by randomly distorting these prototypes. In the most popular version of the prototype distortion task, the category exemplars are random dot patterns (Posner & Keele 1968). An example is shown in Figure 4. In a typical application, many stimuli are created by randomly placing a number of dots on the display. One of these dot patterns is then chosen as the prototype for category A. The others become stimuli not belonging to category A. The other exemplars in category A are then created by randomly perturbing the position of each dot in the category A prototype. A consequence of this process that will prove important in our later discussions is that the stimuli that are not in category A have no coherent structure. For this reason, participants are often instructed to respond "yes" or "no" depending on whether the presented stimulus is a member of category A, rather than "A" or "B" as in the tasks illustrated in Figures 1–3. As the name suggests, prototype distortion tasks have been commonly used by prototype theorists, who argue that categorization is the act of comparing the presented stimulus to the prototype of each contrasting category (Homa, Sterling, & Trepel 1981; Posner & Keele 1968; Minda & Smith 2001).

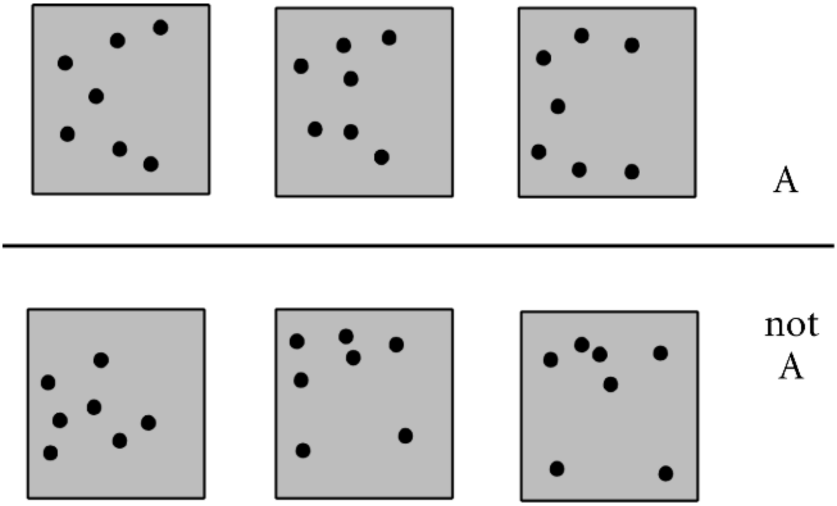


**Figure 4.** Some exemplars from a prototype distortion category learning task with random dot patterns.

## 3.2 Category learning dissociations

We have now observed a number of different dissociations between performance in rule-based and information-integration category learning tasks. Collectively, these provide strong evidence that learning in these two types of tasks is mediated by separate systems. A number of these results show that the nature and timing of trial-by-trial feedback about response accuracy is critical with information-integration categories but not with rule-based categories. First, in the absence of any trial-by-trial feedback about response accuracy, people can learn some rule-based categories, but there is no evidence that they can learn information-integration categories (Ashby, Queller, & Berretty 1999). Second, even when feedback is provided on every trial, information-integration category learning is impaired if the feedback signal is delayed by as little as five seconds after the response. In contrast, such delays have no effect on rule-based category learning (Maddox, Ashby, & Bohil 2002). Third, similar results are obtained when observational learning is compared to traditional feedback learning. Ashby, Maddox, and Bohil (2002) trained subjects on rule-based and information-integration categories using an observational training paradigm in which subjects are informed before stimulus presentation of what category the ensuing stimulus is from. Following stimulus presentation, subjects then pressed the appropriate response key. Traditional feedback training was as effective as observational training with rule-based categories, but with information-integration categories, feedback training was significantly more effective than observational training.

Another qualitative difference between these two tasks is that information-integration category learning is more closely tied to motor outputs than rule-based category learning. Ashby, Ell, and Waldron (2002) had subjects learn either rule-based or information integration categories using traditional feedback training. Next, some subjects continued as before, some switched their hands on the response keys, and for some the location of the response keys was switched (so the Category A key was assigned to Category B and vice versa). For those subjects learning rule-based categories, there was no difference among any of these transfer instructions, thereby suggesting that abstract category labels are learned in rule-based categorization. In contrast, for those subjects learning information-integration categories, switching hands on the response keys caused no interference, but switching the locations of the response keys caused a significant decrease in accuracy. Thus, it appears that response locations are learned in information-integration categorization, but not specific motor programs.

One criticism of all these results is that information-integration tasks are usually more difficult than rule-based tasks, in the sense that information integration tasks usually require more training to reach the same level of expertise. Because of this difficulty difference, one concern is that, collectively, these studies might show only that there are many ways to disrupt learning of difficult tasks compared to simpler tasks. However, several results argue strongly against this hypothesis. First, Waldron and Ashby (2001) had subjects learn rule-based and information-integration categories (shown in Figures 1 and 2, respectively) under typical single-task conditions and when simultaneously performing a secondary task known to activate frontal cortical structures (i.e., a numerical Stroop task). If task difficulty was the relevant variable, then the dual task should interfere more strongly with the difficult information-integration task than with the simpler rule-based task (since it is harder to do two difficult things at once than two simple things). However, in contrast to this prediction, the dual task interfered much more strongly with the ability of subjects to learn the rule-based task than the information-integration task.

Second, Ashby, Noble et al. (2002) found that the same group of Parkinson's disease patients were much more impaired at rule-based category learning (the Figure 1 task) than at information integration category learning (the Figure 2 task). If a single system mediates learning in these two types of categorization tasks, and if Parkinson's disease damages this system, then we would expect the more serious deficits to occur in the more difficult information integration tasks.

These dissociations strongly argue that people learn rule-based and information-integration categories using separate systems. For example, consider just the single Waldron and Ashby (2001) dual-task experiment. Arguably the most successful existing single-process model of category learning is Kruschke's (1992) ALCOVE model. Ashby and Ell (2002a) showed that the only versions of ALCOVE that can fit the Waldron and Ashby data make the strong prediction that after reaching criterion accuracy on the simple (unidimensional) rule-based structures, participants would have no idea that only one dimension was relevant in the dual-task conditions. Ashby and Ell reported empirical evidence that strongly disconfirmed this prediction of ALCOVE. Thus, the best available single-system model fails to account even for the one dissociation reported by Waldron and Ashby (2001).

In addition to dissociations in experiments with healthy young adults, a number of related dissociations have been reported with neuropsychological patient groups. In particular, Ashby and Ell (2001) reviewed the current neuropsychological category learning data and found evidence of a different set of dissociations across these three categorization tasks. Presently, there is extensive category learning data on only a few neuropsychological populations. The best data come from four different groups: 1) patients with frontal lobe lesions, 2) patients with medial temporal lobe amnesia, and two types of patients suffering from a disease of the basal ganglia – either 3) Parkinson's or 4) Huntington's disease. Table 1 summarizes the performance of these groups on the three different types of category learning tasks.

Note first that Table 1 indicates a double dissociation between frontal lobe patients and medial temporal lobe amnesiacs on rule-based tasks and information-integration tasks with few exemplars per category. Specifically, frontal patients are impaired on rule-based tasks (e.g., the Wisconsin Card Sorting Test; Kolb & Whishaw 1990) but medial temporal lobe amnesiacs are normal (e.g., Janowsky, Kritchevsky, & Squire 1989; Leng & Parkin 1988). At the same time, the available data on information-integration tasks with few exemplars per category indicates that frontal patients are normal (Knowlton, Mangels, & Squire 1996), but medial temporal lobe amnesiacs are impaired (i.e., they show a late-training deficit – that is, they learn normally during the first 50 trials or so, but thereafter show impaired learning relative to age-matched controls; Knowlton, Squire, & Gluck 1994). Therefore, the neuropsychological data also support the hypothesis that at least two systems partic-

**Table 1.** Performance of various neuropsychological populations on three types of category learning tasks.

| Neuropsychological Group | | Task | | | |
|---|---|---|---|---|---|
| | | Rule-Based | Information-Integration | | Prototype Distortion |
| | | | Many Exemplars | Few Exemplars | |
| Frontal Lobe Lesions | | Impaired | ? | Normal | ? |
| Basal Ganglia Disease | Parkinson's Disease | Impaired | Impaired | Impaired | ? |
| | Huntington's Disease | Impaired | Impaired | Impaired | ? |
| Medial Temporal Lobe Amnesia | | Normal | Normal | Late Training Deficit | Normal |

ipate in category learning. Of course, until more data are collected on the information-integration tasks, this conclusion must be considered tentative.

Table 1 can also be used to construct first hypotheses about which neural structures mediate learning in the various category learning tasks. For example, patients with frontal or basal ganglia dysfunction are impaired in rule-based tasks (e.g., Brown & Marsden 1988; Cools et al. 1984; Kolb & Whishaw 1990; Robinson, Heaton, Lehman, & Stilson 1980), but patients with medial temporal lobe damage are normal in this type of category learning task (e.g., Janowsky et al. 1989; Leng & Parkin 1988). Thus, an obvious first hypothesis is that the prefrontal cortex and the basal ganglia participate in this type of learning, but the medial temporal lobes do not. Converging evidence for the hypothesis that these are important structures in rule-based category learning comes from several sources. First, an fMRI study of a rule-based task similar to the Wisconsin Card Sorting Test showed activation in the right dorsal-lateral prefrontal cortex, the anterior cingulate, and the head of the right caudate nucleus (among other regions) (Rao et al. 1997). Similar results were recently obtained in an fMRI study of the Wisconsin Card Sorting Test (Monchi et al. 2001). Second, many studies have implicated these structures as key components of executive attention (Posner & Petersen 1990) and working memory (e.g., Fuster 1989; Goldman-Rakic 1987, 1995), both of which are likely to be critically important to the explicit processes of rule formation and testing that are assumed to mediate rule-based category learning. Third, a recent neuroimaging study identified the (dorsal) anterior cingulate as the site of hypothesis generation in a rule-based category-learning task (Elliott & Dolan 1998). Fourth, lesion studies in rats implicate the dorsal caudate nucleus in rule switching (Winocur & Eskes 1998).

Next, note that in information integration tasks with large categories, only patients with basal ganglia dysfunction are known to be impaired (Filoteo, Maddox, & Davis 2001a; Maddox & Filoteo 2001). In particular, medial temporal lobe patients are normal (Filoteo, Maddox, & Davis 2001b). So a first hypothesis should be that the basal ganglia are critical in this task, but the medial temporal lobes are not. If the number of exemplars per category is reduced in this task to a small number (e.g., 4 to 8), then medial temporal lobe amnesiacs show late training deficits – that is, they learn normally during the first 50 trials or so, but thereafter show impaired learning relative to age-matched controls (Knowlton, Squire, & Gluck 1994). An obvious possibility in this case, is that normal observers begin memorizing responses to at least a few of the more distinctive stimuli – a strategy that is not available to the medial temporal lobe amnesiacs, and which is either not helpful or impossible when the categories contain many exemplars. Since patients with basal ganglia dysfunction are also impaired with small categories requiring information-integration (Knowlton, Mangels et al. 1996; Knowlton, Squire et al. 1996), a first hypothesis should be that learning in such tasks depends on the basal ganglia and on medial temporal lobe structures.

Finally, to our knowledge, of the patient groups identified in Table 1, only amnesiacs have been run in prototype distortion tasks. Several studies have reported that this patient group shows normal learning in prototype distortion tasks, which suggests that learning in this task does not depend on an intact medial temporal lobe (Knowlton & Squire 1993; Kolodny 1994). Ashby and Ell (2001) suggested that under certain conditions, learning in prototype distortion tasks might depend, in part, on the perceptual representation memory system – through a perceptual learning process. In the random dot pattern experiments, this seems plausible because all category A exemplars are created by randomly perturbing the positions of the dots that form the category A prototype (see Figure 4). Thus, if there are cells in visual cortex that respond strongly to the category A prototype, they are also likely to respond to the other category A exemplars, and perceptual learning will increase their response. If this occurs, the observer could perform well in this task by responding "yes" to any stimulus that elicits a strong feeling of visual familiarity. Recent fMRI studies of subjects in prototype distortion tasks show learning related changes in visual cortex (Reber et al. 1998), and are thus consistent with this hypothesis. Before drawing any strong conclusions however, it is vital to obtain category learning data on prototype distortion tasks from patients with basal ganglia disease or frontal lobe lesions.

In artificial grammar learning, subjects must decide whether or not a letter string has a familiar (artificial) grammatical structure (e.g., Reber 1989). Although seemingly very different from prototype distortion, it has also been proposed that artificial grammar learning depends on the perceptual representation memory system (Knowlton, Squire et al. 1992). Indirect support for this hypothesis comes from a number of studies showing that amnesiacs and basal ganglia disease patients exhibit normal artificial grammar learning (Knowlton, Squire et al. 1996; Knowlton, Ramus, & Squire 1992; Meulemans, Peigneux, & Van der Linden 1998). Future research should explore the possible connections between prototype distortion category learning and artificial grammar learning.

## 4. Are there multiple implicit category learning systems?

The results reviewed above suggest that there may be multiple qualitatively different implicit category learning systems. Two obvious possibilities are a procedural-learning based system that is mediated, in part, by the basal ganglia, and a perceptual representation system that relies on perceptual learning in visual cortex. The next two sections consider these possibilities in some detail. A third possibility that should also be considered, however, is whether there is an exemplar memory-based system.

In cognitive psychology, one of the most popular and influential theories of category learning is exemplar theory (Brooks 1978; Estes 1986; Medin & Schaffer 1978; Nosofsky 1986), which assumes that categorization decisions are made by accessing memory representations of all previously seen exemplars. Exemplar theorists are careful not to assume that this process of accessing memory representations is explicit, but most exemplar theorists have not taken a strong stand about the neural basis by which these memory representations are encoded. A natural candidate is the hippocampus and other medial temporal lobe structures (e.g., Pickering 1997). However, this is problematic because these brain areas are thought to mediate (the consolidation of) episodic memory, which is considered to be explicit (Fuster 1989; Knowlton & Squire 1993; Reber & Squire 1994). Certainly people are not consciously aware of recalling all previously seen exemplars when making categorization decisions.

There are situations in which episodic memory may contribute to category learning. In particular, with categories that contain a few highly distinct exemplars, people may memorize responses to at least some category members. Then, when a particularly distinct exemplar is presented, subjects may use episodic memory to recall the correct response. As mentioned previously, this might be the cause of the late-training deficit that has been reported when medial temporal lobe amnesiacs learn information-integration categories. Note, however, that the possible use of episodic memory to recall the response associated with the single current stimulus is very different from the processes hypothesized by exemplar theory. According to exemplar theory, the memory representations of all previously seen exemplars are accessed on every trial. Although they both seem to involve a similar type of memory trace, psychologically these two possibilities are very different. Recalling the response to a distinct stimulus is an explicit process, whereas accessing all previously seen exemplars almost necessarily must be implicit (since subjects report no awareness of such massive activation). On the other hand, there is evidence that at least some of the success of exemplar theory is due to the ability of exemplar models to mimic this explicit recall process. For example, Smith and Minda (2000) found that the best fits of a powerful exemplar model to category learning data collected using a popular information integration category structure (with a few highly distinct stimuli in each category) occurred when the response probabilities were determined almost completely by the presented stimulus. The representations of other category members were also activated, but the model parameters were such that these were so dissimilar to the presented stimulus that they had virtually no effect on the predictions of the model.

In summary, there is some evidence that an explicit, episodic memory-based process may contribute to category learning in some situations (e.g., when categories contain a few highly distinct exemplars). There is also theoretical reason to expect that an implicit exemplar memory-based system may contribute to category learning. However, the only attempts that have been made to describe the neurobiological basis of such a system have focused on the hippocampus and related structures that are thought to mediate explicit, episodic memories (Gluck, Oliver, & Myers 1996; Pickering 1997). Thus, currently, an unresolved, but extremely important question is whether there exists some implicit, exemplar-memory based categorization system.

## 5. A procedural learning-based categorization system

Figure 5 shows the circuit of a putative procedural memory-based category learning system (proposed by Ashby et al. 1998; Ashby & Waldron 1999). The

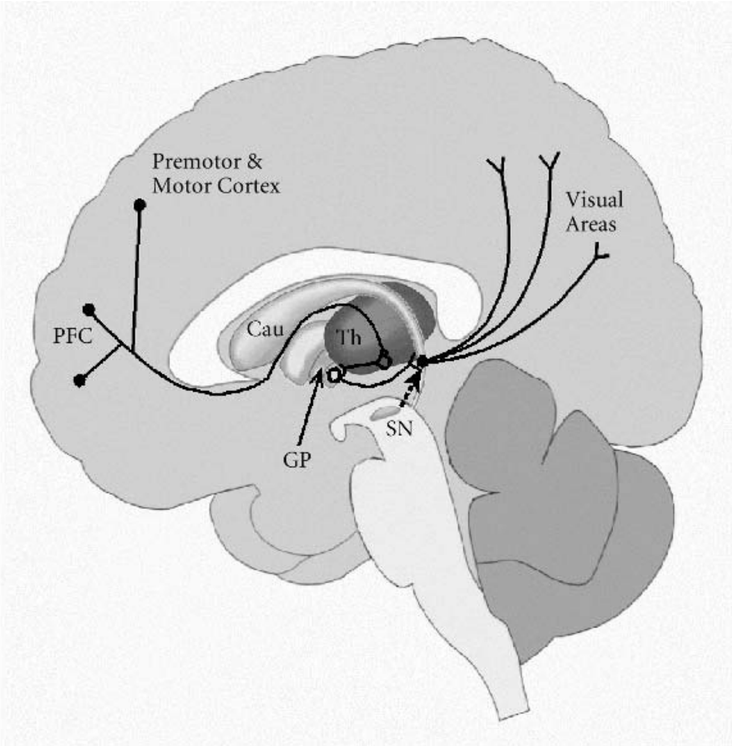


**Figure 5.** A procedural-memory-based category learning system. Excitatory projections end in solid circles, inhibitory projections end in open circles, and dopaminergic projections are dashed. PFC = prefrontal cortex, Cau = caudate nucleus, GP = globus pallidus, and Th = Thalamus.

key structure in this model is the caudate nucleus, a major input structure within the basal ganglia. In primates, all of extrastriate visual cortex projects directly to the tail of the caudate nucleus, with about 10,000 visual cortical cells converging on each caudate cell (Wilson 1995). Cells in the tail of the caudate (i.e., medium spiny cells) then project to prefrontal and premotor cortex (via the globus pallidus and thalamus; e.g., Alexander, DeLong, & Strick 1986). The model assumes that, through a procedural learning process, each caudate unit learns to associate a category label, or perhaps an abstract motor program, with a large group of visual cortical cells (i.e., all that project to it).

Perhaps the best evidence for a basal ganglia contribution to category learning comes from a long series of lesion studies in rats and monkeys that show that the tail of the caudate nucleus is both necessary *and* sufficient for visual discrimination learning. Many studies have shown that lesions of the tail of the caudate nucleus impair the ability of animals to learn visual discriminations that require one response to one stimulus and a different response to some other stimulus (e.g., McDonald & White 1993, 1994; Packard, Hirsch, & White 1989; Packard & McGaugh 1992). For example, in one study, rats with lesions in the tail of the caudate could not learn to discriminate between safe and unsafe platforms in the Morris water maze when the safe platform was marked with horizontal lines and the unsafe platform was marked with vertical lines (Packard & McGaugh 1992). The same animals learned normally, however, when the cues signaling which platform was safe were spatial. Since the visual cortex is intact in these animals, it is unlikely that their difficulty is in perceiving the stimuli. Rather, it appears that their difficulty is in learning to associate an appropriate response with each stimulus alternative, and in fact, many researchers have hypothesized that this is the primary role of the neostriatum (e.g., Rolls 1994; Wickens 1993). Technically, such studies are categorization tasks with one exemplar per category. It is difficult to imagine how adding more exemplars to each category could alleviate the deficits caused by caudate lesions, and it is for this reason that the caudate lesion studies support the hypothesis that the caudate contributes to normal category learning.

The sufficiency of the caudate nucleus for visual discrimination learning was shown in a series of studies by Gaffan and colleagues that lesioned all pathways out of visual cortex except into the tail of the caudate (e.g., projections into prefrontal cortex were lesioned by Eacott & Gaffan 1991, and Gaffan & Eacott 1995; projections to the hippocampus and amygdala were lesioned by Gaffan & Harrison 1987). None of these lesions affected visual discrimination learning.

The procedural learning that has been hypothesized to occur in the caudate nucleus is thought to be facilitated by a dopamine mediated reward signal from the substantia nigra (pars compacta) (e.g., Wickens 1993). There is a large literature linking dopamine and reward, and many researchers have argued that a primary function of dopamine is to serve as the reward signal in reward-mediated learning (e.g., Beninger 1983; Miller, Sanghera, & German 1981; Montague, Dayan, & Sejnowski 1996; White 1989; Wickens 1993). For example, it has been shown that rewards, and events that signal reward, elicit release of dopamine from several brainstem sites (for reviews, see, e.g, Bozarth 1994; Pfaus & Phillips 1991; Phillips, Blaha, Pfaus, & Blackburn 1992), and it is well known that dopamine antagonists (i.e., neuroleptics) disrupt the reward signal and render reinforcement ineffective (e.g., Ataly & Wise 1983).

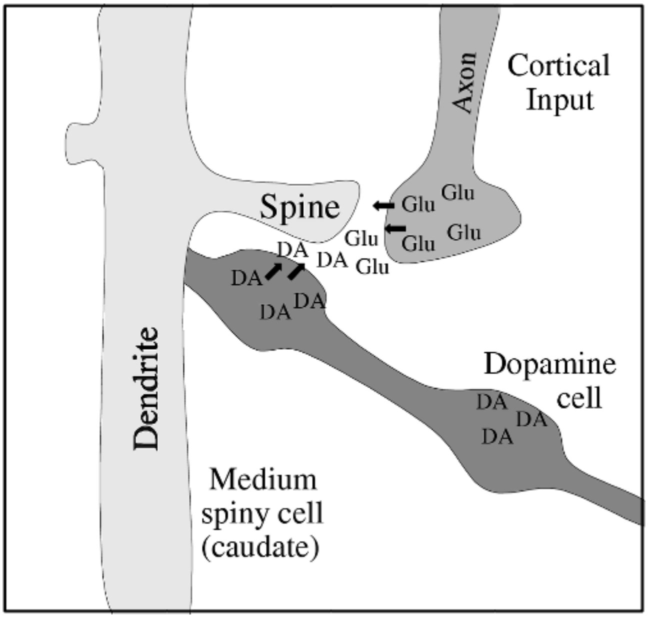


**Figure 6.** A closer view of a cortical-striatal synapse. Here, a cortical cell terminal releases glutamate (Glu) onto the dentritic spine of a medium spiny cell of the caudate nucleus. Dopamine cells of the substantia nigra also project onto medium spiny cells and upon presentation of reward, release dopamine (DA) into the same synapse.

Fairly specific neurobiological models of this learning process have been developed (e.g., Wickens 1993). Figure 6 shows a close-up view of a synapse between the axon of a pyramidal cell originating in visual cortex and the dendrite of a medium spiny cell in the caudate nucleus. Note that glutamate projections from visual cortex and dopamine projections from the substantia nigra both synapse on the dendritic spines of caudate medium spiny cells (DiFiglia, Pasik, & Pasik 1978; Freund, Powell, & Smith 1984; Smiley et al. 1994). A cortical signal causes an influx of free $Ca^{2+}$ into the spines (through NMDA receptors). Because of its strong positive charge, free $Ca^{2+}$ is buffered very quickly within the intracellular medium. The main effect of $Ca^{2+}$ entering the cell is to activate Ca-dependent protein kinases, which then perform a number of cellular functions, including strengthening (long term potentiation – LTP) and weakening (long term depression – LTD) the synapse (e.g., Cooper, Bloom, & Roth 1991; Lynch et al. 1983; Wickens 1993). Because the spines are somewhat separated from the bulk of the intracellular medium, free $Ca^{2+}$ persists for several seconds after entering the cell (Gamble & Koch 1987; MacDermott et al. 1986).

Under ideal conditions, the dopamine-mediated reward signal will arrive during this time, and there is substantial evidence that it will interact with the glutamate signal. The most popular model of this interaction assumes that after dopamine binds to the $D_1$ receptor and activates its associated G protein, a sequence of chemical reactions result that ultimately inhibit the deactivation of the Ca-dependent protein kinases that are activated after glutamate binds to the NMDA receptor (Nairn, Hemmings, Walaas, & Greengard 1998; Pessin et al. 1994; Wickens 1990, 1993). The effect of this inhibition is that dopamine locks the glutamate second messenger in the "on" position, thereby potentiating the learning effect. Thus, the presence of dopamine strengthens the synapses that were active on a trial when reward was delivered (e.g., Huang & Kandel 1995).

The model described in Figures 5 and 6 easily accounts for all of the dissociations between rule-based and information integration category learning tasks that were described above. First, because the dopamine mediated reward signal is thought to be necessary for learning (e.g., LTP) to occur in the caudate nucleus, the absence of such a reward signal should greatly interfere with this form of implicit category learning. For this reason, the model predicts that learning in information integration tasks should be impaired (relative, say, to learning in rule-based tasks) during unsupervised categorization, or when the category label is shown before stimulus presentation (rather than after the response). In addition, as mentioned above, the timing of the reward signal relative to the response is critical for this type of learning. In reward-mediated learning, it is essential to strengthen those (and only those) synapses that actively participated in the response that elicited the reward. Because there is necessarily some delay between response and reward delivery, this means, therefore, that some trace must be maintained that signals which synapses were recently active. In the case of the medium spiny cells in the caudate nucleus, the morphology of the dendritic spines allows this trace to exist for several seconds after the response is initiated (Gamble & Koch 1987; MacDermott et al. 1986). If the reward is delayed by more than this amount, then the ensuing dopamine release will strengthen inappropriate synapses and learning will be adversely affected.

The model described in this section does not make strong predictions about the effects of switching hands or response locations after learning is complete. This is because there are projections from the caudate nucleus to all frontal areas, including prefrontal, premotor, and motor cortices (via the globus pallidus and the thalamus; e.g. Alexander et al. 1986). Even so, the neostriatum (i.e., the caudate and putamen) has been strongly implicated in procedural motor learning (Jahanshahi et al. 1992; Mishkin et al. 1984; Saint-Cyr et al. 1988; Willingham et al. 1989), so it is not unexpected that an implicit category learning system situated in the tail of the caudate nucleus would engage in response learning more strongly than, say, a rule-based system that is largely mediated within prefrontal cortex.

Finally, the model described here is also consistent with the dual-task study of Waldron and Ashby (2001). The numerical Stroop task that was used as the dual task in this study was selected specifically because it is known to activate frontal cortical areas. As such, it was predicted to interfere more strongly with the frontal-based explicit reasoning system than with the caudate-based implicit system.

In addition to accounting for these dissociations, the model described in Figures 5 and 6 also accounts for the dissociations that have been reported for various neuropsychological patient groups (i.e., summarized in Table 1). First, the model predicts category learning deficits in information-integration tasks in patients with Parkinson's or Huntington's disease because both of these populations suffer from caudate dysfunction. It also explains why frontal patients and medial temporal lobe amnesiacs are relatively normal in these tasks – that is, because neither prefrontal cortex nor medial temporal lobe structures play a prominent role in the Figure 5 model.

Before closing this section, it should be noted that the model shown in Figure 5 is strictly a model of *visual* category learning. However, it is feasible that a similar system exists in the other modalities, since they almost all also project directly to the basal ganglia, and then indirectly to frontal cortical areas (again via the globus pallidus and the thalamus; e.g., Chudler, Sugiyama, & Dong 1995). The main difference is in where within the basal ganglia they initially project. For example, auditory cortex projects directly to the body of the caudate (i.e., rather than to the tail; Arnalud, Jeantet, Arsaut, & Demotes-Mainard 1996).

## 6. A possible perceptual representation category learning system

No one has yet proposed a detailed category learning model that uses the perceptual representation memory system. However, based on work in the memory literature, it seems likely that such a category learning system, if it exists, would be based in sensory cortex (Curran & Schacter 1996; Schacter 1994) and would involve some form of perceptual learning. As mentioned above, it has been suggested that such a system might play a prominent role in prototype distortion tasks (Ashby & Ell 2001).

Before investigating this possibility further, it is worth noting that even if the perceptual representation memory system did contribute to learning in prototype distortion tasks, it is not clear that prototype abstraction would meet the standard criteria of a separate system (Ashby & Ell 2002b). When the stimuli are visual in nature, then any category learning system must receive input from the visual system. If some category learning system X depends on input from the brain region mediating prototype abstraction, then system X and the prototype abstraction system would not be mediated by separate neural pathways – a condition often considered necessary for separate systems (e.g., Ashby & Ell 2002b). For example, under this scenario, a double dissociation between system X and the prototype system should be impossible. Damage to the neural structures downstream from visual cortex that mediate system X should induce deficits in category learning tasks mediated by system X, but not necessarily in prototype abstraction tasks. On the other hand, damage to visual cortex should impair all types of visual category learning. Thus, if prototype abstraction is mediated within visual cortex, then any group impaired in prototype abstraction should also be impaired on all other category learning tasks. In addition, it should be extremely difficult, or impossible, to find neuropsychological patient groups that are impaired in prototype abstraction, but not in other types of category learning. The available neuropsychological data supports this prediction, but as Table 1 indicates, only very limited tests of this prediction are currently possible.

Although the term "perceptual learning" is often broadly defined (e.g., Kellman 2002), in this chapter we use the term to refer specifically to learning related changes in sensory cortex. Perceptual learning of this type is thought to occur any time repeated presentations of the same stimulus occur during some relatively brief time interval (Dosher & Lu 1999). Unlike the reward-mediated learning that is thought to occur in the basal ganglia, no reward seems necessary for perceptual learning (e.g., Kellman 2002). In fact, a response does not even seem to be required (e.g., Posner & Keele 1968; Homa & Cultice 1984). Presumably then, perceptual learning is mediated by a form of LTP that is quite close to classical Hebbian learning. In other words, rather than the three-factor learning rule described in the previous section in which learning occurs only in the presence of presynaptic activation, postsynaptic activation, and reward, apparently with perceptual learning, only pre- and postsynaptic activation are necessary.

In the visual cortex, LTP has been shown to occur at synapses between cortical pyramidal cells. Like the LTP that occurs in the procedural learning system, LTP in the perceptual representation system requires presynaptic activation from cortical cells releasing glutamate. However, this system does not require activation of dopamine receptors for LTP to occur. In the procedural learning system, the activation of dopamine receptors eventually potentiates the learning-related effects of the protein kinases that are thought to be activated by the glutamate signal (Wickens 1993). The most widely known mechanism of cortical LTP also requires activation of NMDA channels. As in medium spiny cells, activation of NMDA receptors in cortex leads to an increase in intracellular $Ca^{2+}$, and subsequently to an increase in a protein kinase that has been shown to mediate LTP (i.e., calcium dependent protein kinase II). Unlike medium spiny cells in the caudate nucleus, however, this process apparently does not require dopamine (i.e., the cortical protein kinase undergoes autophosphorylation) (Malenka & Nicoll 1999).

Many different types of categorization experiments have been reported in the literature. Ashby and Maddox (1998) distinguished between what they called (A, B) tasks and (A, not A) tasks. In an (A, B) task, subjects are presented a series of exemplars that are each from some category A or from a contrasting category B. The task of the subject is to respond with the correct category label on each trial (i.e., “A” or “B”). In an (A, not A) task, there is a single central category A and subjects are presented with a series of stimuli that each are either an exemplar from category A or a stimulus that does not belong to category A. The subject’s task is to respond “Yes” or “No” depending on whether the presented stimulus was or was not a member of category A. Historically, prototype distortion tasks have been run both in (A, B) form and in (A, not A) form. An important difference is that in an (A, B) task, the stimuli associated with both responses each have a coherent structure – that is, they each have a central prototypical member around which the other category members cluster (and likelihood tends to decrease monotonically with psychological distance from the prototype). In an (A, not A) task, this is true of the stimuli associated with the “A” (or “Yes”) response, but not of the stimuli associated with the “not A” (or “No”) response. The “not A” stimuli have no central member, no coherent structure, and over a reasonably large region of stimulus space, any given pattern is as likely to be associated with this response as any other pattern (with the exception of course, of the part of the space in which the category A exemplars are clustered).

This digression is important because if the perceptual representation system contributes to category learning, then it likely will have very different effects in (A, not A) and (A, B) tasks. Consider first an (A, not A) task. The category A prototype will induce a graded pattern of activation throughout visual cortex. One particular cell (or small group of cells) will fire most rapidly to the presentation of this pattern. Call this cell A. In other words, cell A will fire to a particular range of visually similar patterns that includes the category A prototype. A low level distortion of the category A prototype will be visually similar to the prototype and therefore will also likely cause cell A to fire. Thus cell A will repeatedly fire throughout training on the category A exemplars. As a result, perceptual learning will cause the magnitude of the cell A response to increase throughout training. In contrast, the stimuli associated with the "not A" response will be visually dissimilar to the category A prototype and therefore will be unlikely to cause cell A to fire. During the transfer or testing phase of the experiment, the subject can use the increased sensitivity of cell A to respond accurately. In particular, stimuli from category A are likely to lead to an enhanced visual response compared to stimuli that do not belong to category A. From the subject's perspective, this enhanced visual response might be interpreted as an increased visual familiarity. Thus, to respond with above chance accuracy, subjects need only respond "A" or "Yes" to any stimulus that elicits a feeling of familiarity.

Next, consider an (A, B) task. In this case there will be some cell A maximally tuned to the category A prototype, but there will be some other cell B that is tuned to the category B prototype. During training, every presented stimulus is a distortion of either the category A or category B prototype, so it is likely that either cell A or B will fire on many trials. The actual number will depend on how much the prototypes are distorted to create the two categories. During the testing phase, all stimuli are again from either category A or B, and so stimuli from both categories will be equally likely to elicit an enhanced visual response (assuming the same level of distortion was used to create both categories). As a result, almost everything will feel familiar to the subject, so this feeling of familiarity will not help subjects decide whether to respond "A" or "B".

The conclusion therefore, is that if the perceptual representation system involves two-factor Hebbian learning, then that system could greatly assist in learning in (A, not A) tasks, but it would be of little help in (A, B) tasks. This is not to say that learning in (A, B) prototype distortion tasks is impossible, only that other learning systems must be used. Kolodny (1994) reported that amnesiacs learn normally in (A, B) prototype distortion tasks [actually in (A, B, C) tasks], so it seems unlikely that people memorize the category label associated with each prototype. One obvious possibility is that they instead use a procedural-memory based system of the type described in the previous section. If so, then several strong, yet untested predictions follow. First, patients with basal ganglia disease (e.g., Parkinson's or Huntington's disease) should be normal in (A, not A) prototype distortion tasks, but impaired in

(A, B) tasks. Second, because feedback is much more important to procedural learning than perceptual learning, unsupervised prototype distortion category learning should be better in (A, not A) tasks than in (A, B) tasks. Homa and Cultice (1984) showed that unsupervised learning is possible in (A, B) tasks if the category members are all low-level distortions of the prototypes, but to our knowledge, no one has systematically compared unsupervised learning in (A, not A) and (A, B) tasks.

Because there is so little available data, the predictions and inferences drawn in this section are highly speculative. Therefore, much more work needs to be done before we will have a clear understanding of the role played by the perceptual representation memory system in category learning.

## 7. Summary and conclusions

The issue of whether human category learning is mediated by one or several category learning systems is a question of intense current debate. Although this issue is still unresolved, recent cognitive, neuropsychological, and neuroimaging data support the weaker hypothesis that different memory systems may participate in different types of category learning tasks. This chapter focused on two memory systems that may contribute to implicit category learning – procedural memory and the perceptual representation memory system.

The recent surge of interest in implicit category learning has a number of practical benefits. First, it immediately ties the categorization literature to the large and well established memory literature. Second, it organizes new research efforts, and it encourages collecting data of a qualitatively different nature than have been collected in the past. Third, it encourages a more critical examination of categorization theories than has been common in the past – largely because it adds constraints on both psychological process and neural structure that historically have not received much attention in the categorization literature. Thus, no matter how it is eventually resolved, the field will benefit from the current interest in implicit categorization.

## Notes

* This research was supported in part by National Science Foundation Grant BCS99-75037. We thank Luis Jiménez and Eliot Hazeltine for their helpful comments. Correspondence

concerning this chapter should be addressed to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106 (e-mail: ashby@psych.ucsb.edu).

**1.** Crick and Koch (1998) did not take the strong position that working memory is necessary for conscious awareness. Even so, they did argue that some short-term memory store is required. However, they left open the possibility that an extremely transient iconic memory might be sufficient.

**2.** Note that there is an explicit rule that also yields perfect accuracy, but it involves three "ands" and two "ors". Despite running many subjects through the Figure 2 categories, we have never had a subject describe this explicit rule at the end of training, even though almost all subjects eventually learn these categories perfectly. For this reason, the Figure 2 task is better described as an information-integration task, rather than as a rule-based task.

## References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9*, 357–381.

Alfonso-Reese, L. A. (1997). Dynamics of Category Learning. Unpublished doctoral dissertation, University of California, Santa Barbara.

Arnalud, E., Jeantet, Y., Arsaut, J., & Demotes-Mainard, J. (1996). Involvement of the caudal striatum in auditory processing: c-fos response to cortical application of picrotoxin and to auditory stimulation. *Brain Research: Molecular Brain Research, 41*, 27–35.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481.

Ashby, F. G., & Ell, S. W. (2001). The neurobiological basis of category learning. *Trends in Cognitive Science, 5*, 204–210.

Ashby, F. G., & Ell, S. W. (2002a). Single versus multiple systems of category learning: Reply to Nosofsky and Kruschke (2002). *Psychonomic Bulletin & Review, 9*, 175–180.

Ashby, F. G., & Ell, S. W. (2002b). Single versus multiple systems of learning and memory. In J. Wixted and H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology: Vol. 4 Methodology in Experimental Psychology* (3rd ed., pp. 655–691). New York: Wiley.

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2002). Abstract category labels are learned in rule-based categorization, but response positions are learned in information-integration categorization. Manuscript submitted for publication.

Ashby, F. G. & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 33–53.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception & Performance, 18*, 50–71.

Ashby, F. G. & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Handbook of perception & cognition: Judgment, Decision Making, and Measurement* (Vol. 3). New York: Academic Press.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30*, 666–677.

Ashby, F. G., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2002). Category learning deficits in Parkinson's disease. Manuscript submitted for publication.

Ashby, F. G., Queller, S., & Berretty, P. T. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*, 1178–1199.

Ashby, F. G. & Waldron, E. M. (1999). The nature of implicit categorization. *Psychonomic Bulletin & Review, 6*, 363–378.

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, *130,* 77–96.

Atalay, J. & Wise, R. A. (1983). Time course of pimozide effects on brain stimulation reward. *Pharmacology, Biochemistry and Behavior, 18,* 655–658.

Beninger, R. J. (1983). The role of dopamine in locomotor activity and learning. *Brain Research, 287,* 173–196.

Bozarth, M. A. (1994). Opiate reinforcement processes: Re-assembling multiple mechanisms. *Addiction, 89,* 1425–1434.

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

Brown, R. G. & Marsden, C. D. (1988). Internal versus external cures and the control of attention in Parkinson's disease. *Brain, 111*, 323–345.

Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A Study of Thinking.* New York: Wiley.

Chudler, E. H., Sugiyama, K., & Dong, W. K. (1995). Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Research, 674*, 33–45.

Cools, A. R., van den Bercken, J. H. L., Horstink, M. W. I., van Spaendonck, K. P. M., & Berger, H. J. C. (1984). Cognitive and motor shifting aptitude disorder in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 47*, 443–453.

Cooper, J. R., Bloom, F. E., & Roth, R. H. (1991). *The Biochemical Basis of Neuropharmacology (Sixth Edition)*. New York: Oxford.

Crick, F. & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in Neuroscience, 2*, 2263–2275.

Crick, F. & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375*, 121–123.

Crick, F. & Koch, C. (1998). Cousciousness and neuroscience. *Cerebral Cortex, 8*, 97–107.

Curran, T. & Schacter, D. L. (1996). Memory: Cognitive neuropsychological aspects. In T. E. Feinberg and M. J. Farah (Eds.), *Behavioral Neurology and Neuropsychology* (pp. 463–471). New York: McGraw-Hill.

Difiglia, M., Pasik, T., & Pasik, P. (1978). A Golgi study of afferent fibers in the neostriatum of monkeys. *Brain Research, 152,* 341–347.

Dosher, B. A. & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*, 3197–3221.

Eacott, M. J. & Gaffan, D. (1991). The role of monkey inferior parietal cortex in visual discrimination of identity and orientation of shapes. *Behavioural Brain Research, 46,* 95–98.

Elliott, R. & Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *Neuroimage, 8*, 17–29.

Erickson, M. A. & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*, 107–140.

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

Estes, W. K. (1994). *Classification and Cognition*. Oxford: Oxford University Press.

Filoteo, J. V., Maddox, W. T., & Davis, J. (2001a). A possible role of the striatum in linear and nonlinear categorization rule learning: Evidence from patients with Huntington's disease. *Behavioral Neuroscience*, *115*, 786–798.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001b). Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society, 7*, 1–19.

Freund, T. F., Powell, J. F., & Smith, A. D. (1984). Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spine. *Neuroscience, 13,* 1189–1215.

Fuster, J. M. (1989). *The Prefrontal Cortex* (2nd Edition). New York: Raven Press.

Gaffan, D. & Eacott, M. J. (1995) Visual learning for an auditory secondary reinforcer by macaques is intact after uncinate fascicle section: indirect evidence for the involvement of the corpus striatum. *European Journal of Neuroscience*, *7,* 1866–1871.

Gaffan, D. & Harrison, S. (1987). Amygdalectomy and disconnection in visual learning for auditory secondary reinforcement by monkeys. *Journal of Neuroscience ,7,* 2285–2292.

Gamble, E. & Koch, C. (1987). The dynamics of free calcium in dendritic spines in response to repetitive synaptic input. *Science, 236,* 1311–1315.

Gluck, M. A., Oliver, L. M., & Myers, C. E. (1996). Late-training amnesic deficits in probabilistic category learning: A neurocomputational analysis. *Learning and Memory, 3*, 326–340.

Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In Plum, F. and Mountcastle, V. (Eds.), *Handbook of Physiology* (pp. 373–417). American Physiological Society.

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron, 14*, 477–485.

Heaton, R. K. (1981). A manual for the Wisconsin Card Sorting Test. Odessa, FL: Psychological Assessment Resources.

Homa, D. & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 83–94.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 418–439.

Huang, Y. Y. & Kandel, E. R. (1995). D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CA1 region of the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 2446–2450.

Jahanshahi, M., Brown, R. G., & Marsden, C. (1992). The effect of withdrawal of dopaminergic medication on simple and choice reaction time and the use of advance information in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*, 1168–1176.

Janowsky, J. S., Kritchevsky, A. P., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103,* 548–560.

Kellman, P. J. (2002). Perceptual learning. In R. Gallistel and H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology: Vol. 3 Learning, Motivation, and Emotion* (3rd ed., pp. 259–299). New York: Wiley.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science, 3*, 172–179.

Knowlton, B. J. & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1*, 106–120.

Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., Swenson, M., & Butters, N. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology*, *10*, 538–548.

Köhler, W. (1925). *The Mentality of Apes*. New York: Harcourt, Brace & Co.

Kolb, B., & Whishaw, I. Q. (1990). *Fundamentals of Human Neuropsychology* (3rd Ed.). New York: W. H. Freeman & Company.

Kolodny, J. A. (1994). Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychological Science, 5*, 164–169.

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

Leng, N. R. & Parkin, A. J. (1988). Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology, 27*, 359–362.

Lynch, G., Larson, J., Kelso, S., Barrionuevo, G., & Schottler, F. (1983). Intracellular injections of EGTA block induction of hippocampal long-term potentiation. *Nature, 305,* 719–721.

MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J., & Barker, J. L. (1986). NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spinal cord neurones. *Nature, 321,* 519–522.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics, 53*, 49–70.

Maddox, W. T. & Ashby, F. G., & Bohil, C. J. (2002). Delayed feedback effects on rule-based and information-integration category learning. Manuscript under review.

Maddox, W. T., & Filoteo, J. V. (2001). Striatal contribution to category learning: Quantitative modeling of simple linear and complex non-linear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society, 7*, 710–727.

Malenka, R. C., & Nicoll, R. A. (1999). Long-term potentiation – a decade of progress? *Science, 285,* 1870–1874.

McDonald, R. J. & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience, 107*, 3–22.

McDonald, R. J. & White, N. M. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral and Neural Biology, 61,* 260–270.

Medin, D. L. & Schaffer, M. M. (1978) Context theory of classification learning. *Psychological Review, 85*, 207–238.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1997). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242–279.

Meulemans, T., Peigneux, P., & Van der Linden, M. (1998). Preserved artificial grammar learning in Parkinson's disease. *Brain & Cognition, 37,* 109–112.

Miller, J. D, Sanghera, M. K., & German, D. C. (1981). Mesencephalic dopaminergic unit activity in the behaviorally conditioned rat. *Life Sciences, 29,* 1255–1263.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 3,* 775–799.

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

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *21*, 7733–7741.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16,* 1936–1947.

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

Nairn, A. C., Hemmings, H. C., Walaas, S. I., & Greengard, P. (1988). DARPP-32 and phosphatase inhibitor-1, two structurally related inhibitors of protein phosphatase-1, are both present in striatonigral neurons. *Journal of Neurochemistry, 50,* 257–262.

Nosofsky, R. M. (1986) Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M. & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7*, 375–402.

Nosofsky, R. M. & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review, 9*, 169–174.

Nosofsky, R. M. & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247–255.

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience, 9*, 1465–1472.

Packard, M. G. & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience*, *106*, 439–446.

Pessin, M. S., Snyder, G. L., Halpain, S., Giraut, J.-A., Aperia, A., & Greengard, P. (1994). DARPP-32/protein phosphatase-1/Na+/K+ ATPase system: A mechanism for bidirectional control of cell function. In K. Fuxe, L F. Agnat, B. Bjelke, and D. Ottoson (Eds.), *Trophic Regulation of the Basal Ganglia* (pp. 43–57). New York: Elsevier Science.

Pfaus, J. G. & Phillips, A. G. (1991). Role of dopamine in anticipatory and consummatory aspects of sexual behavior in the male rat. *Behavioral Neuroscience, 105*, 727–743.

Phillips, A. G., Blaha, C. D., Pfaus, J. G., & Blackburn, J. R. (1992). Neurobiological correlates of positive emotional states: Dopamine, anticipation and reward. In: Ken T. Strongman, Ed. *International Review of Studies on Emotion*, Vol. 2.. New York, NY: John Wiley & Sons.

Pickering, A. D. (1997). New approaches to the study of amnesic patients: What can a neurofunctional philosophy and neural network methods offer? *Memory, 5*, 255–300.

Posner, M. I. & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.

Posner, M. I. & Petersen, S. E. (1990). Attention systems in the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J., Cunningham, J. M., Cox, R. W., Stein, E. A., & Binder, J. R. (1997). Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport, 27*, 1987–1993.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *18*, 219–235.

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory, 1*, 217–229.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory, 5*, 420–428.

Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology, 48*, 605–614.

Rolls, E. T. (1994). Neurophysiology and cognitive functions of the striatum. *Revue Neurologique, 150,* 648–660.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain*, *111*, 941–959.

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

Schacter, D. L. (1994) Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter and E. Tulving (Eds.), *Memory Systems* (pp. 233–268). Cambridge: MIT Press.

Smiley, J. F., Levey, A. I., Ciliax, B. J., & Goldman-Rakic, P. S. (1994). D1 dopamine receptor immunoreactivity in human and monkey cerebral cortex: predominant and extrasynaptic localization in dendritic spines. *Proceedings of the National Academy of Sciences of the United States of America, 9,* 5720–5724.

Smith, D. J. & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 3–27.

Smith, E. E. & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Waldron, E. M. & Ashby, F. G. (2001). The effects of concurrent task interference on category learning. *Psychonomic Bulletin & Review, 8*, 168–176.

White, N. M. (1989). A functional hypothesis concerning the striatal matrix and patches: Mediation of S-R memory and reward. *Life Sciences, 45,* 1943–1957.

Wickens, J. (1990). Striatal dopamine in motor activation and reward-mediated learning: Steps towards a unifying model. *Journal of Neural Transmission: General Section, 80,* 9–31.

Wickens, J. (1993). *A Theory of the Striatum*. New York: Pergamon Press.

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.

Wilson, C. J. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In J. C. Houk, J. L. Davis, and D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 29–50). Cambridge: Bradford.

Winocur, G. & Eskes, G. (1998). Prefrontal cortex and caudate nucleus in conditional associative learning: Dissociated effects of selective brain lesions in rats. *Behavioral Neuroscience, 112*, 89–101.

# Structure and function in sequence learning

## Evidence from experimental, neuropsychological and simulation studies

Peter F. Dominey*
Institut de Sciences Cognitives, CNRS, BRON, France

## 1. Introduction

In this chapter I will identify three different aspects or dimensions of sequential structure, with the objective of determining to what extent these dimensions rely on common or dissociated neurophysiological and computational processes, and to what extent attentional processing is required.

Serial structure or order is defined by the relation between an element or set of elements, and its successor. This dimension can be characterized in terms of length and complexity. Length is the number of elements in the sequence. Complexity refers to the maximum number of elements that must be remembered in order to know the correct successor. Consider, for example, the sequence *A*-*B*-*C*-D-*A*-*B*-*C*-E-*A*-*B*-*C*-F. In order to correctly produce "E," the system must remember the four previous elements that define the context for E, thus the complexity of this sequence is four. Temporal structure is defined in terms of the durations of elements (and the possible pauses that separate them), and intuitively corresponds to the familiar notion of rhythm. Thus, two sequences may have identical serial structure and different temporal structure, or the opposite. Abstract structure is defined in terms of generative rules that describe relations between repeating elements within a sequence. Thus, the two sequences A-B-C-B-A-C and D-E-F-E-D-F have different serial structure, but are both generated from the same abstract structure 123–213, and are thus said to be isomorphic. While perhaps not exhaustive, these three dimensions at least partially span the space of possible behavioral sequence structure.

The following sections will address these three dimensions of sequential structure from a multidisciplinary approach that includes behavioral, neuropsychological (Parkinson's disease, schizophrenia, agrammatic aphasia) and neural network simulation perspectives. The net outcome will be the proposition of a framework in which the processing of these different dimensions of sequential structure is realized by dissociable neurophysiological mechanisms with distinct attentional requirements. In this framework serial and temporal structure can be accommodated by a neural system based on corticocortical connections, and connections from cortex to the basal ganglia (the frontostriatal system) that operates with minimal attentional requirements. In contrast, abstract structure is processed by brain systems that are common to those required for particular aspects of syntax processing, and require elevated levels of attentional processing.

## 2. Behavioral studies of learning

This section will review behavioral studies of sequence learning, leading to the following two conclusions. First, that serial and temporal structure appear to be behaviorally linked or correlated, in that modification of one necessarily influences the other. This will have important implications in subsequent sections that address the neurophysiological and computational basis of these functions. The second conclusion will be that while serial and temporal structure can be acquired without overt attention to these dimensions, robust acquisition of abstract structure appears to require this overt attention. Again, this will have important implications in the subsequent sections.

### 2.1 Interactions between serial and temporal structure

In the serial reaction time (SRT) task, stimuli are presented in a repeating sequence, and the reaction times (RTs) for these sequential stimuli become significantly reduced with respect to RTs for the same stimuli presented in random order. The original studies by Nissen and Bullemer (1987), and numerous studies that followed used the SRT task to study the interactions between serial structure and attentional processing through the use of an imposed "dual" task, that typically requires subjects to discriminate tones that accompany each sequence element as high vs. low pitched, and to maintain a running total of the high pitched tones. Since its introduction, this dual task SRT learning protocol has provided a method for the dissociation of different forms of sequence learning and their interaction with sequence complexity and attention (e.g., Cohen et al. 1990; Willingham et al. 1989; Curran & Keele 1993; Jiménez, Méndez, & Cleeremans 1996). In this line of research, different investigators studied the degree of transfer of sequence knowledge between single and dual task conditions (and vice versa) in an effort to determine if dissociable attentional and non-attentional mechanisms could be isolated. Observations of impairments both in the acquisition and the expression of serial structure in dual task conditions were interpreted as evidence that an attentional form of sequence learning that yielded superior performance in single task conditions was blocked in dual task conditions due to the attentional load (Curran & Keele 1993).

Stadler (1995) noted the possibility of an alternative explanation for the dual task transfer effects. He suggested that the dual task condition disrupts sequence learning by preventing consistent temporal organization of the sequence due to the temporal delays introduced during the response-stimulus interval (RSI) by the dual task processing. In a new experiment he dissociated the effects of this temporal disorder from those due to attentional load by testing SRT performance in four conditions: 1) standard SRT task, 2) dual task, thus introducing both attentional load and temporal disruption, 3) requiring subjects to retain in memory a list of letters to be reported after the experiment, in order to introduce an attentional load without temporal disruption, and 4) randomly introducing RSI pauses before half of the trials, in order to introduce a temporal disruption without attentional load. Stadler demonstrated that, in fact, the dual task performance impairment (condition 2) more closely resembled the impairment from temporal disruption of the sequence (condition 4) than the impairment from attentional load of a non-temporal dual task (condition 3). This indicates that the dual task impairments are linked less to attentional deprivations, and more to perturbation of the temporal context in which the sequence is being learned.

Stadler thus suggested a reconsideration of the proposition that attentional learning acquired in single task conditions could not be expressed in dual task conditions. He suggested instead, that while the training occurred in one temporal context, the testing occurred in a different temporal context due to the temporally irregular grouping imposed by the dual task conditions. The fact that the testing occurred in a different context could be the explanation for the performance decrement, which could allow the possibility that there was, in fact, only a single learning mechanism at work, rather than two.

A related effect of the temporal structure on sequence learning has been reported by Frensch and Miner (1994). They observed that sequential learning is reduced when RSIs of 500 ms are uniformly increased to 1500 ms in an SRT

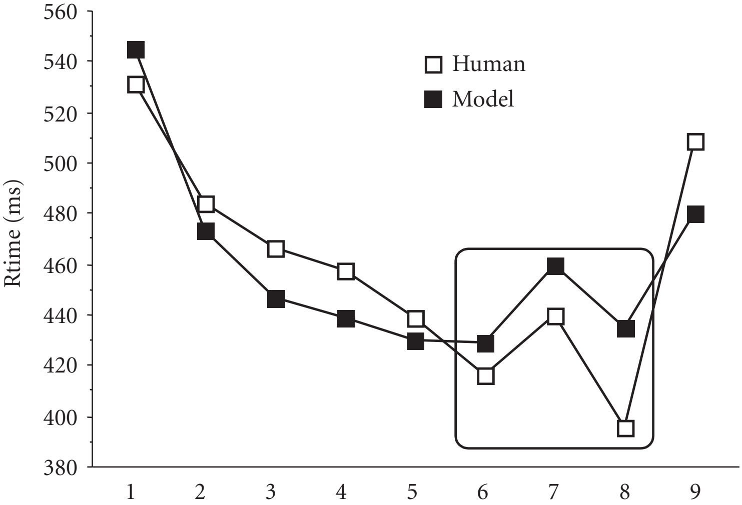


**Figure 1.** Serial and temporal structure learning: Human performance in a modified SRT task that examines temporal structure learning, and comparison with simulation results (Section 3). Six successive blocks of 80 trails using the sequence B-C-B-D-C-A-D-A-C-D, and one of the two temporal structures. Transfer to the different temporal structure occurs in block 7. Block 8 is the same as blocks 1–6, and Block 9 is random. RTs are expressed as means for the 80 RTs in each block in ms, with the simulation equivalent to ms as calculated by the linear regression (Section 3). Learning is measured as the difference in RAND-Seq RTs. RTs in block 7 (Same Serial structure, Different Temporal Structure) are significantly increased from those in blocks 6 and 8, indicating the sensitivity to temporal structure. (From Dominey 1998, Experiment 4).

task. They interpreted this learning reduction in terms of a reduced level of activation of the sequence elements (as induced by the increased RSI) in long-term memory. This observation contributes to the position that not only the serial structure but also the temporal structure is an important parameter in sequence learning. Interestingly, Willingham, Greenberg and Thomas (1997) propose RT changes induced by RSI timing modifications are attributed to effects on performance, rather than on learning. This remains consistent with the idea that the global organization of a sensorimotor sequence is disrupted by random RSI changes.

If the use of random RSIs perturbs learning, will a structured and coherent change of RSIs have the same effect? In order to further investigate the interaction between serial and temporal structure, Dominey (1998a) trained subjects in an SRT task in which each sequence element was associated with a specific

RSI value. In this sense, serial and temporal structure were tightly coupled. In the SRT task, subjects were instructed to respond as quickly and rapidly as possible to stimuli that were presented as single targets on a touch-sensitive computer display screen. In the test block, the serial structure remained unchanged, but the temporal structure was changed by systematically changing the RSI associated with each sequence element. As illustrated in Figure 1, the effect of this modification was a significant increase in reaction times in this test block. This is in agreement with the suggestion that both the serial and the temporal structure can be learned in implicit conditions, and that indeed, they are learned as a unified structure so that modification of one will yield performance impairments for the expression of the other. This is not inconsistent with the possibility of an additive effect, such that modification of both serial and temporal structure will have a greater impact than disruption of only one at a time.

## 2.2 Abstract structure

While serial and temporal structure appear to be functionally linked, what is the status of abstract structure in this relation (reviewed in Dominey, Lelekov, Ventre-Dominey, Jeannerod 1998)? As noted by Shanks and St. John (1994) the learning of rules or abstract structure is characterized by a conscious effort to discover and exploit the appropriate rules, an effort that can be invoked by specific instructions to make such a conscious effort (Gick & Holyoak 1983). This position is supported by studies of analogical transfer in problem solving that involve the extraction of an abstract structure common to several problems with different serial structures (Gick & Holyoak 1983; Holyoak, Junn, & Billman 1984; Holyoak, Novick, & Melz 1994). Such studies demonstrate that this process requires the explicit intention to find the abstract structure. In contrast, the learning of instances or serial structure is oriented towards memorization of the instances themselves, without a conscious processing effort to search for common, rule based structure (e.g. Cohen, Ivry, & Keele 1990; Curran & Keele 1993). These studies suggest the existence of dissociable mechanisms for processing serial and abstract structure.

### 2.2.1 *Abstract structure in artificial grammar learning*

In artificial grammar learning tasks, naive subjects examine a set of letter strings that have been generated by a finite state grammar. Subsequently, the same subjects are asked to judge a new set of letter strings as “grammatical” or not, based on the study set. Thus while SRT tasks require an element-by-element processing and response, AGL tasks involve processing of the sequence as a whole. Several studies of artificial grammar learning have provided convincing evidence for the existence of dissociable mechanisms for serial vs. abstract structure representations (Knowlton & Squire 1996; Gomez & Schvaneveldt 1994; Gomez 1997). A crucial aspect of these experiments is the measure of transfer of learning to a new set of letter strings generated by the finite state grammar, but with each letter systematically replaced by a different letter. Thus in such "changed letter set" conditions a string ABCBAC might become BHTHBT. In these changed letter set conditions, any learning must reflect knowledge of the abstract rather than the serial structure that has been learned. Knowlton and Squire demonstrated that both rule adherence (abstract structure) and chunk strength, i.e. similarity of letter bigram and trigram distribution (serial structure) influence grammaticality judgments. Likewise, Gomez and Schvaneveldt's (1994) results indicate that training on legal letter pairs is sufficient for classification with the same letter set, but that training with longer strings is required to allow transfer of abstract structure to a changed letter set. This suggests that pairs provide a source of serial structure, while abstract structure is only available in strings. These results thus indicate that in AGL there are dissociable forms of representation for serial and abstract structure, respectively.

Although artificial grammar learning tasks are often considered to test implicit learning, it is important to note that during the test phase the subjects are explicitly instructed to apply a set of rules to classify the new objects, and it is likely that some rule abstraction takes place during this explicit testing phase (Perruchet & Pacteau 1991; Reddington & Chater 1996). Likewise, Mathews et al. (1989) have demonstrated that this grammatical knowledge can become at least partially explicit, and that for grammars that exploit relational properties like the ones used in our current studies, learning can only occur in truly explicit conditions. This relation between explicit processing and artificial grammar learning has recently been further clarified by Gomez (1997) who demonstrated that subjects' ability to transfer abstract structure learning to changed letter sets was invariably accompanied by explicit knowledge as revealed in direct tests. Conversely, subjects who learned first-order serial structure dependencies but failed to display transfer of the abstract structure in the changed letter set condition did not differ from naive controls on the direct tests. Thus the ability to transfer knowledge of the grammatical structure appears tightly coupled to some degree of explicit awareness.

### 2.2.2 *Abstract structure in SRT learning*

Finally, Gomez (1997) demonstrated that for the same testing materials presented either in a whole-string AGL task or a letter-by-letter sequence learning task, transfer and the associated acquisition of explicit knowledge occurred only in the AGL task. This indicates that, especially in sequencing tasks, the acquisition of abstract structure and its transfer to isomorphic sequences involves explicit processing.

Dominey et al. (1998) investigated this claim in experiments that studied the possible learning of serial and abstract structure, and the various attentional conditions under which this learning could occur. In particular, this study examined how the instructions given to subjects, and the corresponding attentional states, would affect their ability to learn serial and abstract structure. Subjects in the Explicit group (N = 10) were shown a schematic representation of the rule 123–213 and asked to demonstrate knowledge of the rule by pointing to B-A-C given A-B-C. They were told to actively try to use such a rule to help them go as fast as possible in an SRT task. Subjects in the Implicit group (N = 10) were simply told to go as fast as possible, and were given no hint that there might be an underlying structure in the stimuli.

Figure 2 illustrates the performance of these two groups of subjects. In blocks 1–6, and 8, a 12-element sequence with the structure ABCBACDEFEDF is repeated nine times per block. Note that within this sequence, the abstract structure 123213 repeats twice. With respect to this abstract structure, we say that elements in the second triple 213 are predictable, while elements 123 are non-predictable. In contrast, all elements in a repeating sequence are predictable with respect to the serial structure. Block 7 is a randomly ordered sequence that allows a test of overall learning effects. Blocks 9 and 10 use an isomorphic sequence that has a different serial structure, but shares the same abstract structure. Thus, blocks 9–10 allow a specific test of the possible transfer of knowledge of the abstract structure to a new isomorphic sequence. As seen in Figure 2, both groups demonstrated the classic reduction of reaction times in the initial 6 blocks, and a confirmation of learning as revealed by the increase RTs in the random block 7. This indicates that both groups acquired knowledge of the serial structure, independent of the attentional conditions. A clear difference in the two groups is seen in terms of the difference between RTs for responses that are predictable vs. non-predictable with respect to the abstract structure. Subjects in explicit conditions demonstrate a clear advantage for predictable elements that is not seen in the implicit condition group. More importantly, in the transfer to a new isomorphic sequence in blocks 9 and 10, this advantage in the explicit group transfers to the new se-

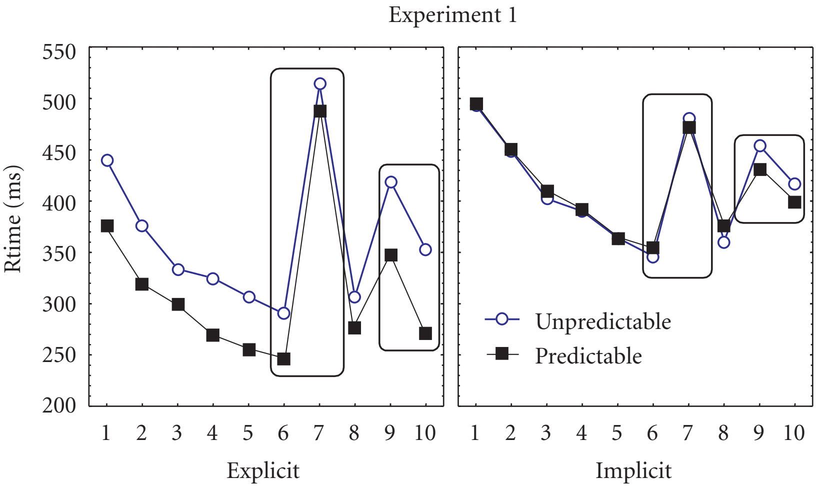


**Figure 2.** Serial and abstract structure learning: Mean RTs for responses that are predictable vs. unpredictable (with respect to the abstract structure) in the 10 blocks of trials for Explicit and Implicit subjects. Blocks 1–6, and 8 use an abstract rule of the form 123–213 that recurs in the 12 element sequence, ABCBACDEFEDF that repeats 9 times in each block. Block 7 is a random series of elements. Blocks 9 and 10 each use 9 repetitions of a new, 12 element sequence isomorphic to that used in blocks 1–6 and 8 (i.e. with the same abstract structure, but with a different serial structure). The critical blocks for learning and transfer assessment are marked in the rounded boxes. Explicit subjects learn serial and abstract structure and display transfer to blocks 9 and 10. Implicit subjects learn only serial structure, with no transfer. (From Dominey et al. 1998, Experiment 1).

quence, whereas the implicit group displays no advantage derived from transfer of abstract structure.

### 2.3 Discussion of behavioral studies

This review of behavioral studies of learning the serial, temporal and abstract structure of sensorimotor sequences indicates that serial and temporal structure can be learned without explicit awareness that there is a structure to be learned (see Jiménez et al. 1996). In contrast it appears that learning and transferring abstract structure requires this explicit awareness, with the following qualifications. It is clear from numerous studies of artificial grammar learning that some abstract knowledge can be acquired without explicit awareness of the abstract structure. The levels of knowledge that are expressed in these AGL studies are statistically greater than chance, but not at all at the level of complete transfer as seen in Dominey et al. (1998), and in Gomez (1997). Indeed, the debate continues concerning just exactly what is learned that allows transfer under implicit conditions in AGL tasks. While rather an oversimplification, one can at least say that by definition, knowledge of an abstract rule that can be fully transferred to new isomorphic sequences required explicit awareness of that rule (Gomez 1997; Dominey et al. 1998).

## 3. Neuropsychological studies

The previous experiments revealed behavioral differences in serial, temporal and abstract structure learning that hinted at, but did not shine much light on, the possibility of dissociated systems being required for learning these different dimensions. Studies of neurological patients can shed more light on these issues, particularly in the critical cases where the neurological deficit affects a brain system that is believed to support one of the potentially dissociated systems.

### 3.1 Parkinson's disease

Parkinson's disease results from a degeneration of dopamine-producing neurons in the midbrain (nucleus accumbens, and substantia nigra pars compacta of the basal ganglia). The most clinically relevant effect is the classic triplet of motor control dysfunctions that are tremor, rigidity and akinesia. Historically, it has been considered that the interaction between cortex and basal ganglia could provide a basis for sensorimotor learning and habit formation (reviewed in Dominey et al. 1995, and in Section 4). In this context, depletion of dopamine in the striatum would perturb the functioning of corticostriatal interactions and thus render sensorimotor learning impaired.

In several recent studies with implicit SRT learning tasks, PD patients displayed significant learning impairments for serial structure with respect to their age matched controls (Ferraro et al. 1993; Jackson et al. 1995; Pascual-Leone et al. 1993). Likewise, their processing of temporal structure seems also to be impaired (Pastor et al. 1992). When the task is made more explicit, however, PD patients' performance improves to that of control subjects (Pascual-Leone et al. 1993). In one explicit form, patients were informed that a sequence would be presented and were asked to concentrate on the sequence without making motor responses, and to be prepared to reproduce the sequence at the end of the test. PD patients' learning in this explicit task, as measured by the declarative knowledge of the sequence, was equal to that of the control subjects. Pascuale-Leone et al. considered that the PD patients' improvement is due to the explicit instructions, though the effects of the eliminated motor effort cannot be ignored. In a related test of the effect of explicit declarative knowledge on SRT performance, PD patients demonstrated learning equal to that of control subjects after 30 repetitions of the sequence. These results indicate that the procedural learning impairment in PD can be reduced or eliminated if tasks are made explicit.

Based on this hypothesis, one would predict that PD patients should be able to learn and transfer abstract structure under explicit conditions. To test this prediction, the sequence learning capabilities of seven (7) non-demented, non-depressed, right handed patients with early or mid-stage (duration range 1–10 years) idiopathic, levodopa-responsive Parkinson's disease were examined. In the SRT task, targets could appear in blocks of two types – random and sequence. In random blocks, 120 targets were successively presented in random order. In sequence blocks, 120 targets were successively presented in 5 repetitions of 24 element sequences of the form A-B-C-B-C-D-C-D-E-D-E-F-E-F-G-F-G-H-G-H-A-H-A-B. This sequence has the interesting property that the serial structure is quite complex and long, whereas the abstract structure is short and simple, as revealed by examining the underscored triplets. For a given triplet of sequence elements, the first two element repeat the previous two elements, and the third element is unpredictable.

Since the goal is to study the transfer of knowledge between different, isomorphic sequences, several sequences must be constructed that meet these requirements. Three such sequences were generated by using the 24 element pattern described above with three different mappings of A-H to the 8 locations on the touch sensitive screen. Thus, the three resulting 24-element sequences differ completely in their serial or verbatim ordering of the spatial targets. However, they are *isomorphic* in that they all share in a common abstract structure.

Figure 3 displays the performance of the Parkinsonian and control subjects on this task. The values presented indicate the difference in milliseconds between reaction times for predictable elements minus those for unpredictable elements. While the effect for the patients has less amplitude than that of the control subjects, both groups display a significant difference that increases over the course of the three successive sequence blocks (Dominey et al. 1997).

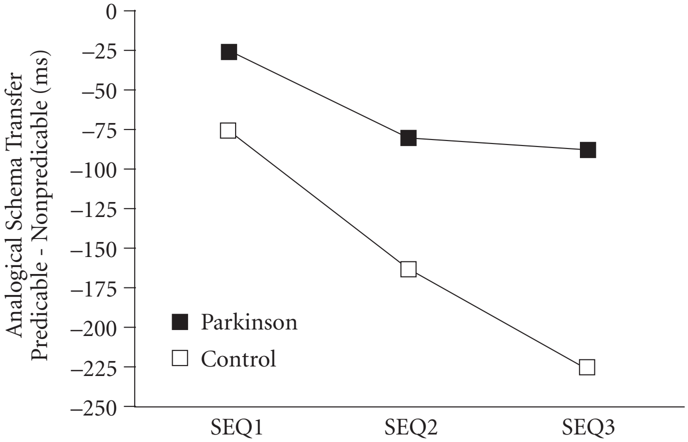


**Figure 3.** Abstract structure learning in PD. The level of abstract structure learning and transfer is displayed as the progressive change during the three sequence blocks of the level of analogical schema acquisition, as indexed by difference between Predictable minus Non-predictable RTs. For both the Control and PD groups, this measure becomes increasingly significant in the progression from SEQ1 to SEQ3, indicating a significant level of analogical transfer in both groups. (From Dominey et al. 1997).

These results indicate that under explicit learning conditions, as observed by Pascuale-Leone et al. (1994), the sequence learning capabilities of Parkinsonian patients remains largely intact, and can even be applied to the acquisition of abstract sequential structure. In contrast, related studies indicate that in implicit conditions, these patients are impaired in the acquisition of serial structure, though this remains to be confirmed. These observations thus contribute to the interpretation that the frontostriatal system plays an important role in implicit learning of serial structure, and is less involved in learning under explicit conditions such as those required for acquisition of abstract structure.

## 3.2 Schizophrenia and sequence learning

In an interesting counterpoint to Parkinsonian patients, while the cognitive performance of schizophrenic subjects is largely preserved in implicit or automatic cognitive functions such as word-stem completion, repetitive priming and procedural learning, it is impaired in explicit or effortful cognitive tasks such as the Wisconsin Card Sorting Task. It has been suggested that various cognitive impairments in schizophrenia may be related to a more central dysfunction, such as an impairment in maintaining and using contextual information, or an impairment in conscious awareness (see Dominey & Georgieff 1997).

Given this profile, one would expect a behavioral mirror image of the Parkinsonian patients to be observed in the schizophrenic patients. That is, one would predict that they should be relatively intact in learning serial structure, and relatively impaired in capitalizing on awareness for learning abstract structure. This prediction was tested in a group of 6 schizophrenic patients and ten matched control subjects. The experimental protocol was identical to that described in Section 2.2, that allows a separation of RTs for elements that are predictable, or not, based on knowledge of the abstract structure. Blocks 1–6 and 8 use a repeating 12 element sequence that can asses the learning of both serial and abstract structure. Block 7 is randomly organized, and blocks 9–10 use a sequence that is "isomorphic" to that in the initial training blocks. That is, it has the same abstract structure and a different serial structure. The six schizophrenic subjects were explicitly informed about the existence of the abstract structure. They were shown a schematic diagram depicting the abstract structure ABCBAC, and were able to successfully supply (by pointing) the missing fragment (BAC) once shown the initial fragment ABC, where the letters correspond to spatial targets as on the touch-sensitive screen. Five control subjects were tested under identical explicit conditions, and the five remaining control subjects were tested in implicit conditions in which they were simply told to point to the spatial targets as quickly and accurately as possible. The results are displayed in Figure 4.

While schizophrenic subjects displayed significant learning of the serial structure, as revealed by the RT reduction in blocks 6 and 8 with respect to random block 7, they failed to learn the abstract structure, as revealed by a lack of effect for predictability, and a lack of transfer to the isomorphic sequence in blocks 9 and 10. Indeed, the relative profile of the schizophrenic patients, who were explicitly informed, is not different from that of the control subjects that were in implicit conditions (Dominey & Georgieff 1997).

In a task that permits the simultaneous learning of serial and abstract sequential structure, Schizophrenic subjects acquired only the serial structure. This is despite the fact that these patients were fully informed, in an "explicit" experimental condition. Thus, though explicitly informed, the schizophrenic patients behave as if they were in implicit conditions. While the Implicit group had RTs that were overall faster than those for the Schizophrenic group, there was no significant group x block interaction in either the measure of serial structure learning, nor in the failed transfer to the isomorphic sequence. In

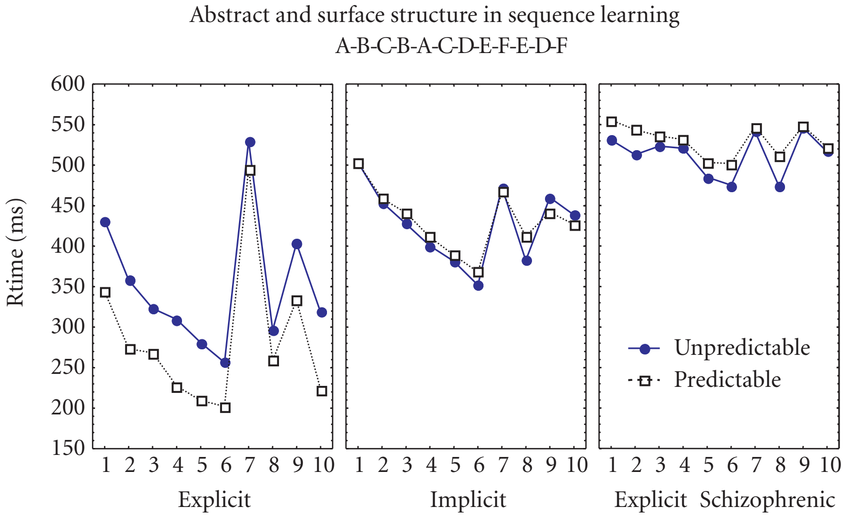


**Figure 4.** Serial and abstract structure learning in schizophrenia. Results from control subjects in Explicit, and Implicit conditions (as in Figure 2), and from schizophrenic patients in explicit condition. Schizophrenic patients display a learning profile similar to that of control subjects in implicit conditions, they learn the serial structure but not the abstract structure (From Dominey & Georgieff 1997).

other words, apart from a global vertical displacement on the time axis, the performance of the Schizophrenic group did not differ from that of the Implicit group. This is in agreement with the proposition that explicit processing is required for processing abstract structure, and that in the absence of this capability as in schizophrenia, abstract structure will not be learned.

### 3.3 Abstract structure and syntax in agrammatic patients

Linguistic syntax represents perhaps the ultimate behavioral extension of abstract sequential structure. A major open issue in cognitive neuroscience concerns the search for the functional and neurophysiological bases of this syntactic aspect of language. In almost all languages there is a standard or "cannonical" word order, such as "agent action object" in the sentence "John greeted Bill" in English. Patients with syntactic processing deficits tend to rely on this canonical word order, and typically demonstrate problems with sentences that deviate from this order, such as passive sentences that follow the non-canonical order "object action agent" as in "Bill was greeted by John" (see Caplan et al.

1985 and Grodzinsky 2000). Based on the functional parallel between the generative aspects of syntax, and these same aspects of abstract sequential structure, Dominey and Lelekov (2000) investigated this question in patients that were impaired in syntactic processing in order to determine if there is a parallel impairment in the processing of abstract sequential structure. Abstract structure processing was examined with a protocol similar to those used in studies of artificial grammar learning (Gomez 1997; Reber 1967), that tests the ability to learn and use an abstract structure to classify letter-sequences. The experiment was conducted with seven aphasic subjects. During an initial familiarization and training period of 10-15 minutes, the subjects studied a list of 10 letter-sequences (e.g. HBSBHS, YPBPYB) generated from the non-canonical abstract structure 123213. The subjects were instructed to study the list in order to decide how to complete the sequence BKT_ _ _. After this training period, subjects demonstrated their understanding of the abstract structure and the task by completing the above sequence with KBT (to form the sequence BKTKBT, following the abstract structure 123213).

In a subsequent testing period of 5 minutes, the patients were presented with 20 new sequences, and were informed that each of the 20 sequences had to be classified as corresponding, or not, to the abstract structure they extracted in the study phase. In a separate testing phase, the same procedure was performed with the canonical abstract structure 123123 for six of the seven patients, as one became unavailable for subsequent testing. Performance in this non-linguistic task was then compared with performance in syntactic comprehension as evaluated using the 9 sentence-type task of Caplan et al (1985).

Figure 5A illustrates the significant correlation between the impairments in syntactic comprehension and in the classification of non-canonical letter sequences based on their abstract structure ($r^2 = 0.86$, $p = 0.003$) for 7 aphasic patients. Note that one patient with a right peri-sylvian lesion (R in Figure 5A) scored the highest on both tasks, indicating a relative sparing of both functions in the case of right hemisphere lesion. In contrast, a patient with a left subcortical lesion (L in Figure 5A) performed poorly in the syntactic comprehension task (Alexander et al. 1987; Lieberman 1992), and also in the abstract structure classification task. This correlation remains significant when patient R is removed ($r^2 = 0.72$, $p = 0.03$), and likewise remains significant when a larger population of 11 aphasic patients ($r^2 = 0.60$, $p = 0.005$) is included.

In order to test the more specific prediction that processing of non-canonical order would be specifically impaired both for linguistic syntax and for non-linguistic abstract structures, a comparison was made between canonical vs. non-canonical performance (in terms of percentage of correct re-

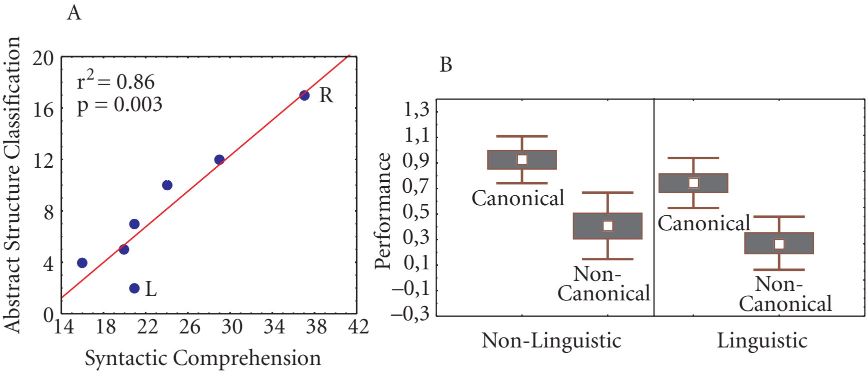


**Figure 5.** Abstract and syntactic structure learning in agrammatic Aphasia. A. Linear regression for performance on the Abstract Structure Classification task, and the Syntactic Comprehension task. Error-free scores are 20 and 45 respectively for the two tasks. L – patient 4 with left subcortical lesion, R – patient 7 with right peri-sylvian lesion. B. Selective impairment for non-canonical structure in linguistic and non-linguistic tasks. Mean performance values as percentage correct, with standard error (boxes), and standard deviation (whiskers). A. For the non-linguistic abstract structure classification task, patients performed significantly better for canonical (92%) vs non-canonical (45%). B. Likewise, for the linguistic syntactic comprehension task, patients performed significantly better for canonical (74%) vs non-canonical (34%) task.

sponses) for the linguistic and non-linguistic tasks. Figure 5B illustrates that for both tasks, non-canonical processing is significantly impaired with respect to canonical processing. Note that the observed superior performance in the non-linguistic task is likely attributed to the fact that chance is 50% in the non-linguistic task (Yes/no response), and below 25% in the linguistic task since more than half of the sentences have at least 6 possible responses (possible orderings of 3 thematic roles). More interestingly, there was also a significant effect for Order ($F(1,5) = 31.7$, $p = 0.0025$), indicating that processing of canonical order was significantly superior (83%) to processing of non-canonical order (37%). Finally and most important, the Order x Task interaction was not significant ($F(1,5) = 0.053$, $p = 0.8$), indicating that this performance impairment for non-canonical order processing holds both for linguistic (canonical 74% vs. non-canonical 30%) and non-linguistic (canonical 92% vs. non-canonical 45%) Tasks.

Lelekov et al. (2000) subsequently wanted to determine if this same functional correlation between syntactic and abstract structure processing would be maintained in schizophrenic patients following up on Dominey and Georgieff

(1997). It has been repeatedly demonstrated that schizophrenic patients are impaired in the comprehension of sentences with complex syntax. Lelekov et al. (2000) investigated the hypothesis that this syntactic comprehension impairment in schizophrenia is not a purely linguistic dysfunction, but rather the reflection of a cognitive sequence processing impairment as observed in the left-hemisphere lesioned patients. They tested schizophrenic patients (n=10) using the standard measure of syntactic comprehension, and our non-linguistic sequence processing task, both of which required simple and complex transformation processing. Patients' performance impairments on the two tasks was highly correlated ($r^2 = 0.84$), and there was a significant effect for complexity, independent of the task. These results are quite similar to those of aphasic patients with left hemisphere lesions. This suggests that syntactic comprehension deficits in schizophrenia reveal the dysfunction of cognitive sequence processing mechanisms that can be expressed both in linguistic and non-linguistic sequence tasks.

These results are consistent with the hypothesis that syntactic comprehension deficits result, at least in part, from an impairment in performing serial order transformations on non-canonical forms that is not restricted to natural language. More generally, these results demonstrate that the use of "a new, highly abstract and precise approach" provides evidence that there is a non-linguistic correlate to the transformation processing impairment described by Grodzinsky (2000), and that both within and outside of natural language, this transformation processing remains highly specific and dissociable from other sequence processing capabilities, as suggested by previous results from simulation (Dominey 1997; Dominey et al. 1998), experimental psychology (Dominey et al. 1998) and neuropsychology (Dominey et al. 1997; Dominey & Georgieff 1997; Lelekov et al. 2000).

### 3.4 Discussion of neuropsychological studies

These neuropsychological studies argue in favor of a neurophysiological dissociation between serial and temporal processing on the one hand, and abstract structure processing on the other. The former rely on the intact functioning of the corticostriatal system, and are thus impaired in Parkinson's disease. The later relies more on the distributed network that includes the peri-sylvian cortex in and around Broca's area, and is impaired both in left-hemisphere lesions associated with aphasia, as well as in conditions of more diffuse physiological dysfunction as in schizophrenia.

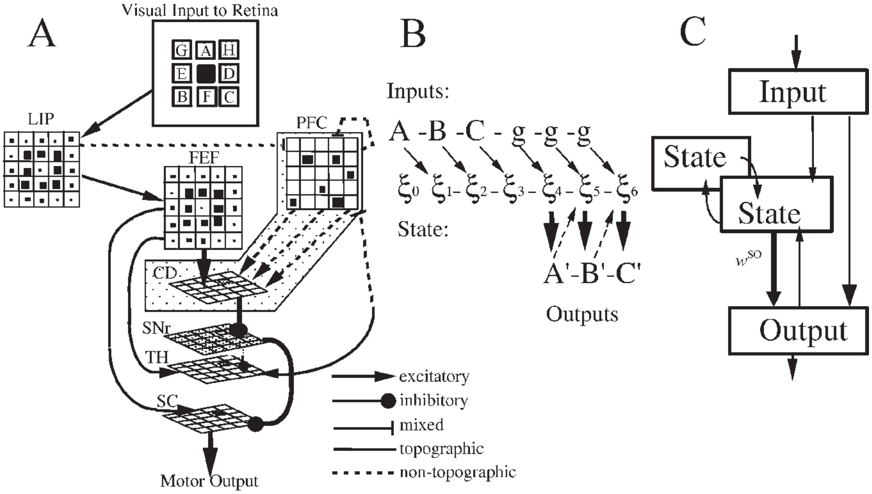


**Figure 6.** A corticostriatal model of sequence learning. A. sequence model architecture. B. Time course of a sequence reproduction task. C. Schematized sequence model. Recurrent State network encodes history of sensory inputs, and motor outputs. State-Output associative memory ($w^{SO}$ synapses) associates internal states with the correct next response in the sequence, via reinforcement learning.

## 4. Simulation studies

Together, the studies above suggest that dissociable systems may be required for the treatment of serial and temporal structure on the one hand, and abstract structure on the other. I now examine the contributions of our modeling studies to this debate. The model is based on a biologically plausible implementation of a sequence reproduction capability observed in primates (Figure 6). The model was developed (Dominey et al. 1995) to explain the capability of primates to learn sequences of oculomotor saccades (eye movements), and the corresponding sequence-encoding activity in neurons of the prefrontal cortex (PFC). Barone and Joseph (1989) recorded PFC neurons whose spatially and temporally selective activity encoded the ongoing state of sequence presentation and execution. From this perspective, the model is based on the principle that at any point during the presentation or reproduction of a sequence, the internal state of the system encodes the history of the sensory and motor events that have so far occurred. At each point during the sequence reproduction, the current state will become linked, via reinforcement learning in a simple associative memory, to the correct next output in the sequence.

In the model presented in Figure 6A, the state representation function is performed by a network corresponding to prefrontal cortex, or PFC. From a functional neuroanatomy perspective, during generation of movement (oculomotor saccade) sequences, continuous inhibition of motor response units in superior colliculus (SC) by substantia nigra (SNr) is temporarily interrupted by caudate (CD) inputs. CD saccade-related cells are influenced by topographic projections from frontal eye fields (FEF) via the parietal cortex (LIP), and also by modifiable, non-topographic projections from prefrontal cortex (PFC). PFC combines visual, motor (saccade efferent copy via thalamus – TH), and recurrent input in order to generate a time varying sequence of internal states for each pre-saccade period in the sequence reproduction task. These states or patterns of activity become associated with caudate activity for the correct saccade by reinforcement learning in an associative memory.

This associative memory allows the activity in PFC encoding sequence state to command the correct saccades corresponding to each element of the sequence. The modeled substrate of this associative memory is in modifiable synapses connecting PFC (state) to the caudate nucleus or CD (motor output influence). Neurophysiological and anatomical evidence suggest that the cortico-striatal system may provide a basis for this associative memory (Rolls & Williams 1987). This suggestion has been substantiated by experiments demonstrating that intact caudate function is required for sensory-motor association learning (Reading et al. 1991; Robins et al. 1990). Indeed, Kermadi et al (1993) recorded single units caudate whose activity reflected a conditional association between sequence state and associated motor responses. The underlying corticostriatal plasticity may be based on dopamine-related modification of corticostriatal synapses after rewards (or absence of predicted rewards) since (a) the most significant dopaminergic input to the caudate originates from the substantia nigra pars compacta (SNc), (b) there is a phasic modulation of this SNc dopamine release in the striatum related to reward (or in the absence of a predicted reward) (Ljungberg et al. 1991), and (c) dopamine originating from SNc participates in synaptic plasticity in the striatum (Calebresi et al. 1992; Centonze et al. 2001).

Figure 6C displays a more schematic representation of this system. The recurrent State network encodes history of sensory inputs, and motor outputs. State-Output associative memory ($w^{SO}$ synapses) associates internal states with the correct next response in the sequence, via reinforcement learning. Figure 6B shows the progression in time of the reproduction of sequence A-B-C. The model starts in an initial state $\xi_0$. Presentation of the first visual input A drives the system to a new state, $\xi_1$. Presentation of input B drives the system from $\xi_1$

to $\xi_2$. Then presentation of input C drives the system from $\xi_2$ to $\xi_3$. Presentation of the first go signal (g) drives the model to state $\xi_4$ and triggers the model to produce a motor output by retrieving from the associative memory the output currently associated with state $\xi_4$ (dark vertical arrow in 6C).

If the retrieved output is incorrect, the offending association in $w^{SO}$ synapses is weakened, reducing the probability that the same choice will be made again. If the output is correct, A', this association is strengthened, and the system moves on to state $\xi_5$ which retrieves B' from the associative memory by the same process, and so on. By this trial and error learning, this system will learn the state-output associations (indicated by the heavy arrows in 6B and C), and thus will reproduce spatiotemporal sequences as the concatenation of context dependent state-response associations. The next section will consider the sequence learning performance of this model in the SRT context, and in particular with respect to the serial, temporal, abstract dissociation.

### 4.1 Serial and temporal structure

As noted in Section 2.1, a number of studies have examined the interaction between serial and temporal structure in SRT learning. In particular, based on the proposal (Stadler 1995) that dual task conditions may perturb sequence learning by disrupting the temporal organization of sequences, Stadler (1995) and later Dominey (1998a) investigated the effects of manipulation of the regularity of temporal structure on SRT learning. Temporal structure was modified by changing the duration of the Response-to-Stimulus-Interval (RSI), i.e. the delay between a subject's response and the presentation of the subsequent stimulus.

Prior to Stadler's observation that dual task conditions could be mimicked by random introduction of response stimulus interval (RSI) pauses, the choice concerning which parameters to vary in the simulation of dual task conditions was a conceptual challenge for modelers (Dominey 1998b). Cleeremans and McClelland (1991), Cleeremans (1993a) simulated dual task conditions of Cohen et al. (1990) by introducing random changes (noise) to the activity of hidden units in a simple recurrent network (SRN) introduced by Elman (1990). Keele and Jennings (1992) demonstrated that manipulation of parsing-related information could also be used to simulate dual task performance impairments in a recurrent network. Cleeremans (1993b) also modeled discrete attentional and non-attentional processes to simulate dual task effects on these mechanisms as observed by Curran and Keele (1993), again by the addition of noise. In such simulation approaches the modeler must make a choice about how behavioral parameters of the dual task condition, such as tone perception, discrimination and counting, are to be represented in the model, when in fact the mapping between experimental and model parameters is not necessarily direct. Stadler's temporal processing interpretation offers an interesting alternative, in particular for models in which time, including the onset of each stimulus and the RSI for each stimulus, is a parameter that is completely specified. Such models will not be forced to make a difficult representation choice, but instead can directly simulate the temporal disorder conditions as those studied by Stadler (1995).

Simulating real-time temporal structure in recurrent networks is not always a simple matter, however. The problems occur in the following context: After an input is presented, and several network cycles occur, an output is generated and the error is evaluated and corrected. A given recurrent connection has contributed to the error, but in a different way on each successive cycle of information passing through this connection in the network. How is this connection's contribution to the error over these sequential time steps to be unraveled in order to implement the error-reducing learning? One way is to simply require that only one time step can pass between input/output events and connection updates, so that a connection's contribution is only made in one pass through the network. This is the case in the SRN (Elman 1990), in which the learning algorithm and its application to the recurrent connections requires that at each time step a new input is provided, an output is generated and the learning algorithm is applied. Thus, time steps cannot pass independent of input, output and learning processing.

Otherwise, in order to modify connections in a way that takes into account the weight's contribution to error over a number of time steps, one must calculate the effect of that weight change for each of these time steps. One way to do this is to "unfold" the recurrent network in to an n-layered network where each layer is a copy of the original network. One update cycle for this cascade network is equivalent to n time steps in the original recurrent net (Doya 1995). This and other methods of resolving the problem of learning in recurrent networks over multiple time steps are biologically implausible, however, because they are not consistent with forward running time, and/or because they have excessive computational and memory storage requirements (Werbos 1995).

In the current temporal recurrent network (TRN) model (Figure 6) there is no learning in the recurrent part of the network, only in the feedforward connections between the State units and the Output units, and only at the time that a response is evaluated. Thus, the above described complexity of recurrent learning is no longer an issue. Simulation time steps are mapped to real-time or experiment time, and are coupled to input, output and learning processing in a temporally realistic maner. Following the model's production of a given response, the response is evaluated and the reinforcement learning rule (described below) is applied. The network can then run for an arbitrary number of time steps (i.e. the experimenter-specified RSI) before the next input is presented. During this time the State activity is modified due to its recurrent connections, providing an explicit representation of the effects of time. The key point is that for this TRN, changing the temporal structure of a sequence changes the representation of the sequence itself. The model inseparably encodes the combined serial and temporal structure. While the TRN is certainly not unique in the capacity to simulate the passage of time, it is distinguished in the computational simplicity by which this is achieved (Dominey 1998a, b).

In the TRN, reaction times (RTs) are measured as the delay in simulation time steps (sts) between the onset of activation of a given Input unit, and the activation of the corresponding Out unit driven (in part) by the one-to-one connection from Input. In the SRT task, learning is unsupervised since the only possible response in Out is the one driven by the single activated Input unit. Recall that the response units in Out are leaky integrators whose response latencies are not instantaneous and depend on the strength of their driving sources. One source comes from the corresponding Input unit in a one-to-one mapping. This will activate the Out unit with some baseline RT. The other driving source for Out comes from State, which can change with learning in the $w^{SO}$ Synapses (Figure 6C). As learning occurs, RTs for elements in learned sequences will become reduced due to learning-specific influences of State on Out. This SRT learning in the model is understood in terms of three invariant observations: (1) During exposure to a repeating sequence, a given subsequence reliably generates the same pattern of neural activity in State. (2) This subsequence is reliably followed by a given element. (3) Learning results in strengthening of State-Out connections binding that pattern of activity in State to that sequence element in Out. These strengthened connections yield reduced reaction times for units in Out, for any element that is reliably preceded by the same sub-sequence, thus providing an SRT learning capability.

In the standard SRT tasks the presentation of a stimulus follows the preceding response by a fixed response-stimulus-interval (RSI). In the following simulations, the standard RSI is 20 simulation time steps. During this period the State activity continues to evolve, while no input nor output events occur. With a fixed RSI in a series of sequence trials, the resulting State activity for a given trial will be equivalent each time the sequence is repeated, thus providing the basis for learning. Changing the RSIs for each trial allows manipulation of the *temporal structure* of a sequence while leaving its serial order intact. This will yield modification of the State activity associated with each response. In simulating the "pauses" conditions where the RSI is increased on a random election of half of the trials (Stadler 1995), the RSI is augmented to 100 time steps, on half of the trials in a random fashion. During a pause the activity in State can significantly change, so that while the serial order remains intact, the representation of the sequence in State is modified. With random RSI delays, a given trial in one repetition of a sequence may not have the same RSI in the sequence's next repetition, thus changing the pattern of activity in State for that trial, disrupting the learned State-Output association, and yielding an increase RT.

#### 4.1.1 *Simulating the effects of Random RSIs (Stadler 1995)*

Based on Stadler (1995) the combined effects of temporal organization (pause, no-pause), block (sequence, random), and sequence complexity (easy, intermediate and hard) on the model were tested (Dominey 1998b). The purpose was to determine if the model demonstrates the same qualitative patterns of sensitivity to these parameters as did humans. In particular, can the effects of dual task conditions be mimicked by the randomized introduction of response stimulus interval (RSI) pauses?

Ten model "subjects" were produced by using different initial random weights to initialize network connections. Each subject was separately tested in each one of 6 conditions that derive from the possible combination of the two pause conditions (pause, no-pause) and three sequence complexity conditions (easy, intermediate, hard). Each test consisted of 9 sequence blocks of 80 trials, followed by two random blocks of 80 trials. In the pauses tests, the RSI was randomly changed from 20 to 100 simulation time steps on half of the trials, corresponding to the change introduced by Stadler. The easy sequence was BCADBCA, and the hard sequence was DADBACBCAB, as in Stadler (1995). The intermediate sequence was ABCBDC, as used by Curran and Keele (1993) in Experiments 1–3. Random blocks were made up of a randomized uniform distribution of elements A-D that was the same for each of the 6 conditions. A-D correspond to four discrete (x,y) locations on the 5x5 input array, and responses are generated by activation of the corresponding locations in the Output layer.

As seen in Figure 7B, single task (no-pause) performance (17.27 sts) was significantly better than dual task (pause) performance (18.25 sts). There was also a significant sequence learning effect with Sequence performance (14.8 sts)

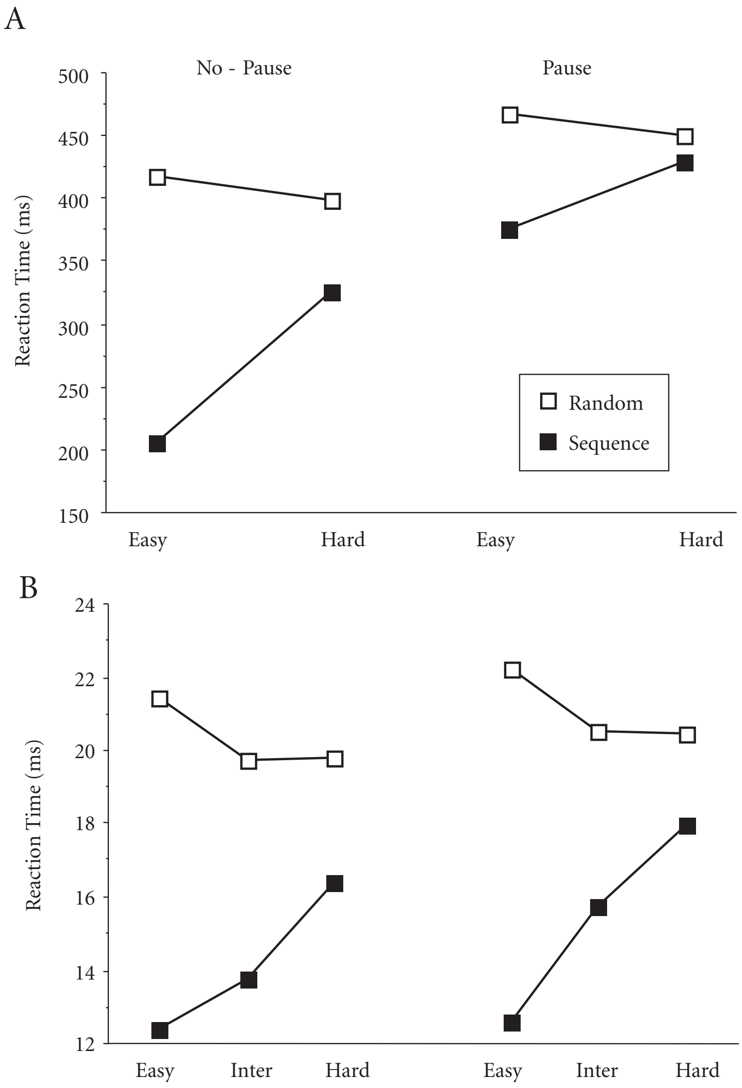


**Figure 7.** Learning effects of temporal disorganization. Performance in pause and no-pause conditions for Seq and Rand blocks with three sequences of varying complexity. Easy: BCADBCA, Intermediate: ABCBDC, Hard: DADBACBCAB. RTs for both Sequence and Random blocks are elevated in the Pauses condition with respect to No-Pauses. While this RT increase is somewhat uniform for Rand blocks, it is complexity dependent for the Seq blocks. A. Human data from Stadler (1995). B. Simulation data. For humans and simulations, temporal disruption impairs learning, with the impairment increasing with sequence complexity. (From Dominey 1998b).

superior to Random performance (20.7 sts). Easy (17.16 sts) and Intermediate (17.45 sts) RTs were faster than Hard RTs (18.66 sts).

However, as can be seen in Figure 7, the simulation data differed from the human data in one noticeable aspect. Whereas learning in humans was considerably greater in the No-pause condition than in the Pause condition for both the easy and hard sequences, for the simulation data this was so only for the intermediate and hard sequence conditions and not for the easy sequence condition. This is due to the fact that the easy sequence was in a sense too easy, and was not sensitive to the temporal structure perturbation. Nevertheless, the simulation data reveal that learning is more effectively eliminated in the pauses condition as the sequence complexity increases, reflecting the interaction between distraction and complexity observed by Cohen et al. (1990).

Most importantly for the present purposes, the model is also sensitive to the effects of temporal disorder induced by introducing RSI increases randomly in half of the trials which produces a dual task effect as demonstrated by Stadler (1995). The model displays the classic learning decrement in pause (dual task) conditions for the sequences of complexity equal or above that used by Curran and Keele (1993) in Experiments 1–3 (ABCBDC). It is thus reasonable to consider this RSI pause method as a way to mimic in the model the effects of a dual task on sequence learning performance. This provides a direct method to test Stadler's (1995) suggestion that it is this disruption of temporal order in a single learning mechanism that may be the explanation for dual-task transfer results in the experiments performed by Curran and Keele (1993). Indeed, I have quantitatively demonstrated that modifications of temporal structure induce corresponding changes in the internal representation (in State) of the sequence, effectively yielding a different sequence (Dominey 1998a, b).

#### 4.1.2 *Simulating dual task effects (Curran & Keele 1993) with Random RSIs*

The simulation described above demonstrates that indeed, the introduction of random RSI structure can produce performance impairments by changing the serial-temporal structure of the sequence. The question remains, however, as to whether an ensemble of behavioral results attributed to dual task effects can also be explained in terms of disruption of temporal structure. This was the goal of part of Dominey (1998b) in addressing the results of a set of four experiments that were presented by Curran and Keele (1993) as evidence for dissociable forms of sequence learning.

Curran and Keele (1993), studied the degree of transfer of sequence knowledge between single and dual task conditions (and vice versa) in an effort to determine if dissociable attentional and non-attentional mechanisms could be isolated. They observed that the improved learning acquired in single task conditions did not transfer to dual task testing. They interpreted this as evidence that an attentional form of sequence learning that yielded superior performance in single task conditions was blocked in dual task conditions due to the attentional load.

They also observed that after learning in dual task conditions, there was no performance improvement when subsequent testing occurred in single task conditions. This was interpreted as evidence that the attentional mechanism was blocked in the initial dual task training and thus unable to learn. Thus even when liberated in the subsequent single task testing, it could not express knowledge that it had never attained. These data together supported the existence of two dissociable forms of sequence learning: One that required attention both for learning and expression, and the other that required attention for neither learning nor expression, but had a low complexity limit.

I set out to determine if these results could be explained in terms of the temporal disruption of sequences (Dominey 1998b). In a simulation based on Experiment 1 of Curran and Keele (1993), initial training takes place in single task (no-pause) conditions, and then testing occurs in dual task (pauses) conditions. This allows the measure of transfer of sequential knowledge acquired in single task conditions to testing in dual task conditions (Figure 8A).

A population of 10 model subjects was tested. The experiment consisted of 15 blocks of 80 trials each (Figure 8B). Blocks 1–11 used a fixed RSI of 20 simulation time steps (sts). Blocks 1–9, and 11 each consisted of 80 trials using the repeating sequence of the form ABCBDC. Block 10 had randomly organized trials. Blocks 12–15 used a randomized schedule of RSIs of 20 time steps on half of the trials, and 100 on half. In this "dual task" (pause) portion, block 14 followed the same sequence as in blocks 1–9, and 11. The other blocks in the "dual task" (pause) testing were randomly organized.

Figure 8 compares C&K's human results (A) with simulation results (B). Learning that occurs in the No-Pause condition does not completely transfer to the Pauses condition, similar to C&K's results for transfer from single to dual task conditions. The difference is that for C&K, these variations were linked to dissociable learning systems, whereas in the current simulations the variations are functionally attributed to the disruptions of temporal structure.

Figure 9 summarizes the learning measures obtained from the four experiments in Curran and Keele (1993) with those obtained in the Dominey (1998b) simulations. C&K's Experiment 1 compared learning expressed in single (no-pause) vs dual (pause) task conditions after training in single task conditions.

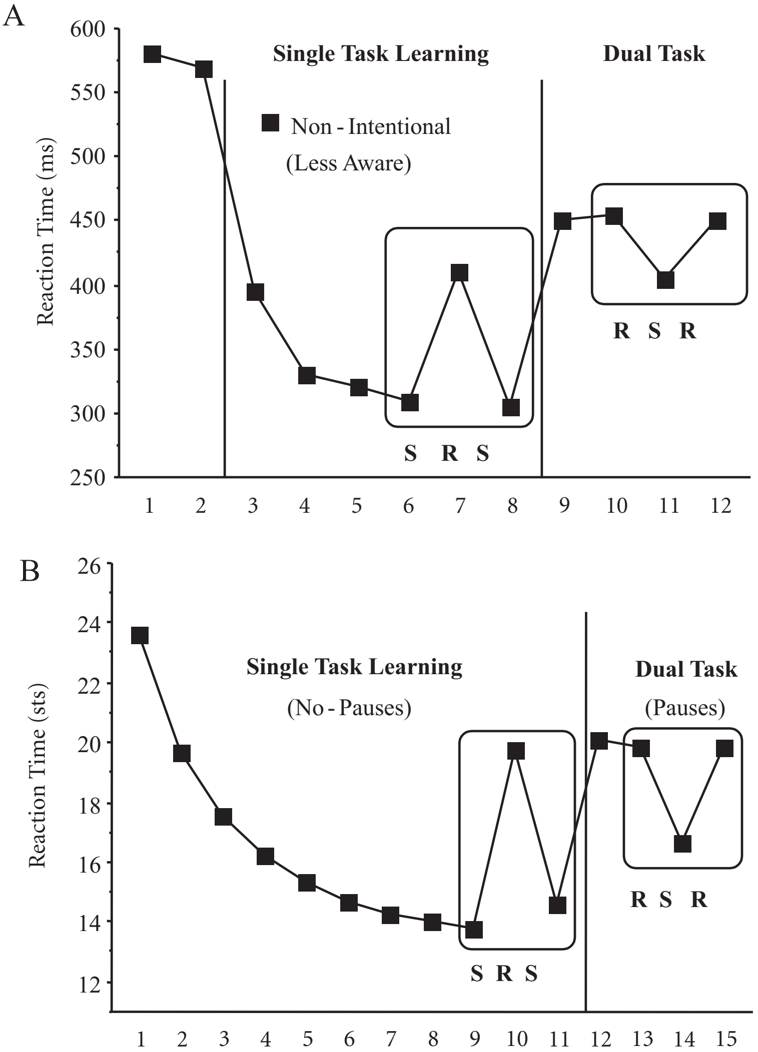


**Figure 8.** Dual task processing as temporal disorganization. Learning acquired in single task conditions does not completely transfer to dual task testing. The blocks used for learning and transfer assessment are marked in the curved boxes. R – random block, S – sequence block. A. Human data from Curran and Keele (1993) Experiment 1 (Figure 1), median block RTs. Blocks 1, 2, 7, 9, 10, 12 Random; 3–6, 8, 11 Sequence. B. Simulation of Curran and Keele (1993) Experiment 1. Mean simulated RTs. Blocks 10, 12, 13 and 15 are random, all others are sequence. For both humans and the TRN model, the expression of learning observed in the dual task (pauses) conditions is significantly reduced with respect to that observed in single task (no pauses) conditions. (From Dominey 1998b).

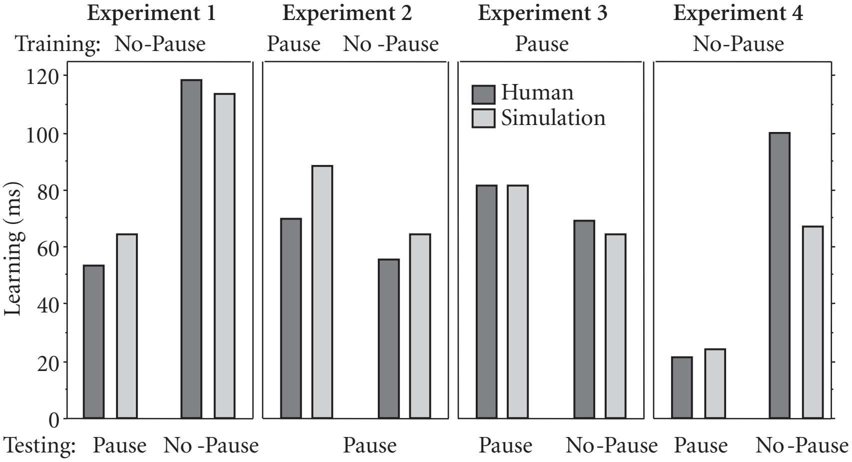


**Figure 9.** Summary of dual task simulations. Comparison of human and simulated learning in dual (pause) and single (no-pause) task manipulations. Learning (Rand-Seq) in milliseconds for the single and dual task related conditions in each of Exps 1–4 of C&K 1993 and the corresponding simulations. Learning cannot optimally occur, nor can it be optimally expressed, in dual task (pauses) conditions. (From Dominey 1998b).

For both the model and humans, the expression of learning was significantly greater in single (no-pause) than dual (pause) conditions, indicating that the advantage of single task learning was not fully transferred to the dual task conditions. In C&K's Experiment 2, one group was trained in dual task and the other in single task conditions. Both were then tested in dual task conditions. Both groups displayed significant learning with no Block x Group interaction, suggesting equivalent sequential knowledge for the two groups. This was true for both the human and simulation data. Again, this indicates that in both systems, the advantage of learning without distraction/pauses is lost in the subsequent presence of distraction/pauses.

In C&K's Experiment 3, subjects were trained in dual task conditions, and tested in dual and then single task conditions. For both the human and simulation data the performance advantage lost in dual task training conditions was not regained in subsequent single task testing. In C&K's Experiment 4, subjects were trained with a more complex sequence in single task conditions and then tested in single and dual task conditions. The learning expressed in single task conditions was significantly greater than that in dual task conditions, both in the human and simulation results, indicating again that the advantage of single task training does not fully transfer to dual task testing.

In summary, the model demonstrates the following behavior of interest: sequence disruption by RSI pauses yields reduced learning, and impairs transfer of prior learning that occurred during preceding single task (no-pauses) training (Dominey 1998b), thus providing a possible explanation for these dual task effects that does not rely on the necessary existence of multiple learning systems with different attentional requirements.

This issue remains complex however, as it combines two topics that are themselves questions of ongoing debate. The first topic concerns the functional disturbance induced by dual task conditions, and the second concerns the relation between dual task effects and random RSI effects. While Curran and Keele (1993) demonstrated that dual task conditions effect both the learning and the expression of sequence knowledge, Frensch, Lin and Buchner (1998), subsequently claimed that dual task conditions effects expression, but not learning. Shanks (this volume) however, using carefully structured test conditions argues that indeed sequence learning is impaired in dual task conditions. Finally, Jiménez and Méndez (1999) argue that dual-task conditions do not interfere either with learning nor with its expression when the sequences are probabilistic. From the perspective of the simulation studies reviewed above, to the extent that RSI randomization mimics dual task effects, it appears clear that dual tasks impair both learning and expression, due to the disruption of the sequence representation. It remains to be investigated if for probabilistic sequences, the presence of an inherent "disruption" (due to the probabilistic structure) will reduce or eliminate this effect.

With respect to the effects of RSI manipulation, while Stadler (1995) first suggested that RSI manipulation will yield a disorganization of the sequence itself, with resulting impairments in learning and expression, Willingham, Greenberg, and Thomas (1997) consider that RSI manipulation does not affect learning, but only performance. From the perspective of the simulation studies reviewed above, this remains consistent with the idea that the global organization of a sensorimotor sequence is disrupted by random RSI changes. Future studies should address these issues in more detail.

#### 4.1.3 *Simulating the learning of structured RSIs (Dominey 1998)*

Section 2.1 reviewed evidence that even coherent and systematic changes in temporal structure can disrupt previously acquired learning. That is, when the initial training occurs with one coherent temporal structure, and then a different, coherent temporal structure is used in the transfer test, a performance impairment is observed in the transfer phase. Here I report on simulation of this effect with the TRN.

The SRT task allows the comparison between RTs for sequential series using the same serial structure but different temporal structures. Two coherent temporal structures T1 and T2 were studied, each consisting of a repeating series of 10 RSI values of either 0.1 or 0.5 simulation time units (20 or 100 time steps, respectively). In T1, elements A and D are always preceded by RSIs of 20 time steps, while elements B and C by are preceded by RSIs of 100, and the opposite in T2.

A set of 5 instances of the model were generated by using different seed values for the random number generator to initialize the weights in $w^{SO}$. The 5 models were replicated to yield 2 equivalent groups of 5, according to the temporal structure used for their training (T1 or T2). The two groups were trained on 6 blocks of 80 trials with the repeating 10-trial sequence B-C-B-D-C-A-D-A-C-D using temporal structures T1 and T2, respectively. The trained models were then tested with the same serial structure while using the same temporal structure as that used for training (blocks 6 and 8), and with the same serial but different temporal structure from the one used for training (block 7), and finally with random serial and the same temporal structure (block 9).

The model's performance in the SRT task is displayed in Figure 1, where the RTs for responses after training in sequence blocks 6 and 8 are reduced with respect to those in the transfer block 7 that uses the different temporal structure, and also with respect to the random block 9. Thus, like the human subjects, the model was significantly perturbed by a change in the temporal structure in block 7, despite the fact that the "perturbing" temporal structure was coherent, rather than random. Given sufficient exposure, this new temporal structure would likely be learned.

### 4.2 Abstract structure

While the TRN is sensitive to serial and temporal structure, it fails to learn abstract structure as defined in Section 2.2 (Dominey et al. 1998). In order to permit the representation of abstract structure as we've defined it, the model must be capable of comparing the current sequence element with previous elements to recognize repetitive structure (i.e. u, u, u, n-2, n-4, n-3 for ABCBAC, where "u" signifies unpredictable, and "n-2" indicates a repetition of the element 2 places behind, etc.). These functions would rely on the more non-sensorimotor associative areas of the anterior cortex, and would permit the generalization of grammar-like rules to new, but "legal" sensorimotor sequences (Dominey 1997). To make this possible, as illustrated in Figure 10, a short term memory (STM) mechanism is introduced that is continuously updated to store the

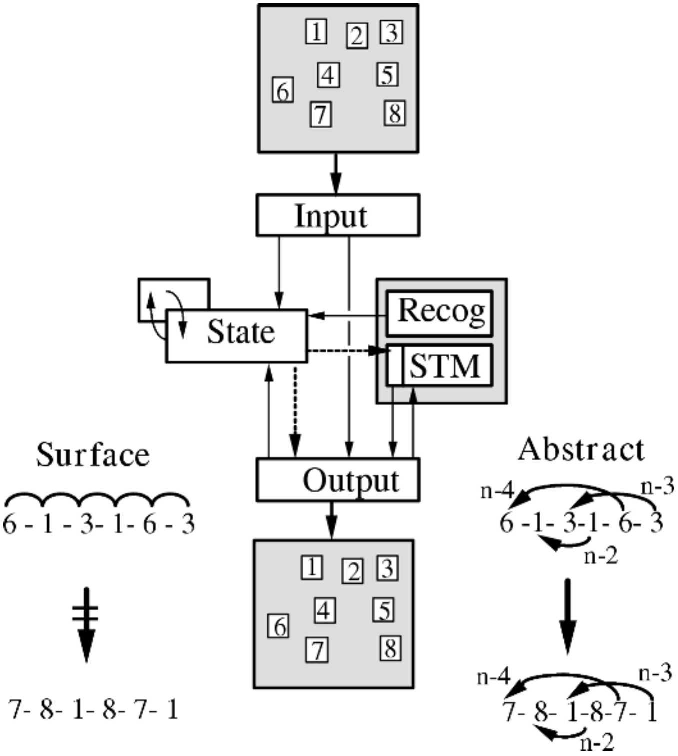


**Figure 10.** Schematic representation of an anatomically structured model for learning serial and abstract structure, the Temporal Recurrent Network (TRN) and Abstract Recurrent Network (ARN) respectively. TRN: Presentation of sequence stimuli to Input activates both State and Output. State is a recurrent network whose activity over time encodes the sequence context, i.e. the history of previous sensory (Input) and motor (Output) events. At the time of each Output response there is a specific pattern of activity, or context, in State. Connections from State to Output (dotted line) are modified during sequence learning, thus binding each sequence context in State with the corresponding response in Output for each sequence element, yielding reduced RTs. Abstract model (ARN) extension: A short term memory (STM) encodes the 7 ± 2 previous responses. Recog encodes repetitions between current response and previous responses in and provides this input to State. Modifiable connections (dotted line) from State gate the contents of appropriate STM elements to Output when repetitive structure is predictable, reducing RTs for predictable elements in isomorphic sequences. Left: The TRN can learn the serial order of sequence elements 613163 but cannot transfer this knowledge to isomorphic sequence 781871. Right: The ARN learns the relations between repeating elements in 613163 as u, u, u, n-2, n-4, n-3, the abstract structure which transfers completely to the isomorphic sequence 781871 (see text) (From Dominey et al. 1998).

previous 7 ± 2 responses, and a Recognition mechanism that compares the current response to the stored STM responses to detect any repeated elements (Dominey 1997, et al. 1998).

These modifications for the Abstract Recurrent Network (ARN) permit the recoding of sequences in terms of their abstract structure that is now provided as input to State. Thus, in terms of the recoded abstract structure representation provided to State, the two sequences ABCBAC and DEFEDF are equivalent: u, u, u, n-2, n-4, n-3. For sequences that follow this "rule," the pattern of activation (context) produced in State by sub-sequence u, u, u, n-2 will reliably be followed by that context associated with n-4. To exploit this predictability, the system should then take the contents of the STM for the n-4th element and direct it to the output, yielding an RT reduction. This is achieved in the following manner: For each STM element (i.e. the structures that store the n-1, n-2, .. responses) there is a unit that modulates the contents of this structure to Output. If one of these units is active, the contents of the corresponding STM structure is directed to Output.

Now, during learning, each time a match is detected between the current response and an STM element, the connections are strengthened between State units encoding the current context and the modulation unit for the matched STM element. The result is that the next time this same pattern of activation in State occurs (i.e. before a match n-4 corresponding to the learned rule), the contents of the appropriate STM will be directed to Output in anticipation of the predicted match, thus yielding a reduced reaction time. In the same sense that the TRN learns to anticipate specific elements that define a given sequence, the ARN learns to anticipate repetitive structure that defines a class of isomorphic sequences. Two formal models for the treatment of serial/temporal and abstract structure have thus now been defined, the TRN and ARN respectively, that together make up the dual process model.

Figure 11 displays the comparison between human subjects in explicit conditions (see Figure 2), and the combined effects of the TRN and ARN. While these two systems together display performance that approaches that of humans, it was also clearly demonstrated (Dominey et al. 1998; Dominey & Ramus 2000) that without the ARN, the representation capability necessary to learn and transfer abstract structure is impossible.

### 4.3 Discussion of simulation studies and dissociable systems

In theory, different types of information structure must be treated by different processes, and the inverse: a given processing architecture must be capable of

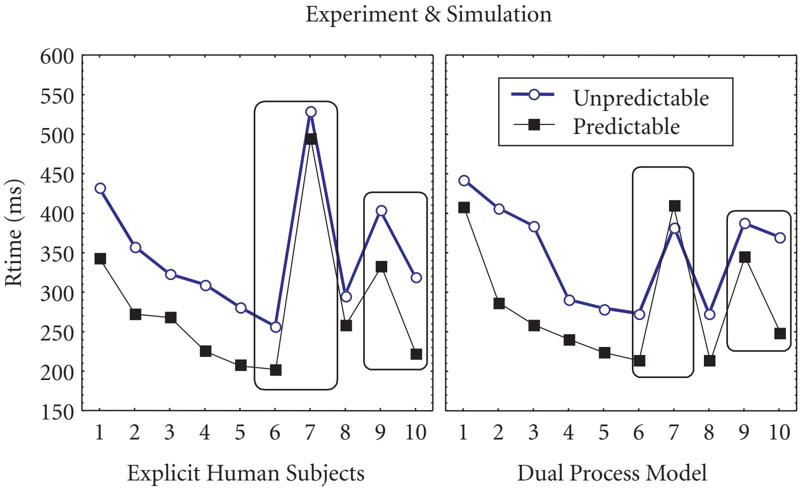


**Figure 11.** Human and simulation performance in abstract structure learning. The combined "dual process" model for serial and abstract structure, like humans in explicit conditions, expresses knowledge of both the surface and abstract structure, and transfers knowledge of the abstract structure to the isomorphic sequence in blocks 9 and 10. Compare to Figure 2. (From Dominey et al. 1998).

treating some but not other information structures. These simulation results demonstrate that for serial and abstract structure, as defined here, there are two corresponding sequence learning models that are capable of learning, respectively, only one of these types of sequential structure. Simulation studies have shown that the combined model that includes the TRN and ARN is effective in covering substantial behavioral ground. Indeed, though it is beyond the scope of this chapter, Dominey and Ramus (2000) have demonstrated that this dual system model is capable of simulating the behavior of human infants in their sensitivity to serial, temporal and abstract structure in language acquisition, and how this capability can be extended through training, to a form of adult syntactic comprehension (Dominey 2002).

With respect to dissociable systems, the simulations of dual task SRT conditions as timing (RSI) changes suggest that it is not necessary to invoke dissociable attentional and non-attentional systems to explain dual task results. However, a dissociable system was required for abstract structure learning. Indeed, it may seem strange to dismiss the Curran and Keele (1993) two-system model and then propose a highly similar two-system model. Why couldn't the

ARN be invoked to do everything? Interestingly, the ARN will fail on sequences that have the simplest structure, i.e. those that have no internal repetition with each element followed by a unique successor, as in ECBAD. The ARN fails here as each element will be represented as a unique element that does not match a recent predecessor. That is, the sequence has no abstract structure. Still, it is clear that multiple sequence learning systems exist. While the TRN simulations demonstrated that a single system could produce performance deficits similar to those observed in dual task conditions, the results clearly do not deny the existence of multiple sequence learning systems, nor the relation between awareness and the functioning of these systems.

## 5. General discussion

At the outset of this chapter I announced the goal of identifying dissociable dimensions of sequential structure and the corresponding neurophysiological underpinnings. I also wanted to determine how this partition was associated with varying degrees of intention or awareness, and finally whether the sequence processing aspects of language could fit into this framework. These issues were separately addressed from the perspectives of human behavior, neuropsychology and simulation results.

The first conclusion from this multidimensional approach is that serial and temporal structure can be processed by a common system, whose functional architecture corresponds to that of the TRN. This learning system appears to rely on the intact functioning of the cortico-striatal system, and can operate in the absence of overt attention or awareness of what is to be learned.

The second conclusion is that, in contrast, the learning of abstract structure is functionally distinct, corresponding to the ARN model. This learning system appears to exploit processing capabilities related to those involved in particular aspects of syntactic processing, and requires additional processing capabilities and attentional awareness of the type of structure to be learned. This conclusion appears to be in conflict with the position evoked by Knowlton & Squire and others, that limited knowledge of abstract structure can be acquired in the absence of explicit awareness, and by amnesic patients. However, as pointed out in Section 2.2, while the levels of transfer in AGL experiments are significantly above chance, they are far from the robustness required for transfer to multiple isomorphic sequences as in the SRT task of Section 2.2, or the sequence classification task in Section 3.3.

The third related conclusion is that there is a functional relation between syntactic processing, and the processing of abstract non-linguistic sequences. Our limited neuropsychological data suggest that this system relies on a distributed network whose principal component is the left peri-sylvian cortex in and around Broca's area. This was revealed in particular by the correlation between impairments in syntactic and abstract structure processing in aphasic patients. If this is the case, it would indicate that to the extent that syntactic processing overlaps with abstract structure processing, and to the extent that abstract structure processing relies on explicit awareness, then syntactic processing will also rely on explicit awareness. While it is often considered that language acquisition is an implicit process, observation of young children in the acquisition phase reveals that they are actively and explicitly engaged in the communicative process of language. Moreover, self observation reveals that both in the production and comprehension of language, one must be attentively, explicitly and actively involved. This is not to be confused with the argument that while one can form and recognize well formed sentences, one may not necessarily be able to describe explicitly the rules by which those sentences were formed. While likely valid, this argument does not contradict the assertion that language acquisition and use is an active, explicit process. This conclusion implies common neurophysiological processes, then, for these respective aspects of syntactic and cognitive sequence processing, and will form the basis for future investigation of the relations between structure and function in cognitive sequence learning (Hoen & Dominey 2000).

## Notes

* This work has been supported by the Cognitique project of the French Education and Research Ministry, and by the Human Frontiers Science Program. I thank Luis Jiménez, Eliot Hazeltine, Greg Ashby and Michael Casale for comments on a previous version of the manuscript.

## References

Alexander, M. P., Naeser, M. A., & Palumbo, C.L. (1987). Correlations of subcortical CT lesion sites and aphasia profiles. *Brain*, *110*, 961–991.

Barone, P. & Joseph, J. P. (1989) Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research, 78*, 447–464.

Calabresi, P., Maj, R., Pisani, A., Mercuir, N. B., & Bernardi, G. (1992). Long-Term Synaptic Depression in the Striatum: Physiological and Pharmacological Characterization. *Journal of Neuroscience, 12*, 4224–4233.

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, *21*, 117–175.

Centonze, D., Picconi, B., Gubellini, P., Bernardi, G., & Calabresi, P. (2001). Dopaminergic control of synaptic plasticity in the dorsal striatum. *European Journal of Neuroscience, 13*, 1071–1077.

Cleeremans, A. (1993a). *Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing.* Cambridge, MA: MIT Press.

Cleeremans, A. (1993b). Attention and Awareness in Sequence Learning. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp.330–335). Hillsdale, NJ: Erlbaum.

Cleeremans, A. & McClelland, J. L. (1991). Learning the Structure of Event Sequences. *Journal of Experimental Psychology: General, 120*, 235–253.

Cohen, A., Ivry, R. I., & 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.

Dominey, P. F. (1995). Complex Sensory-Motor Sequence Learning Based on Recurrent State-Representation and Reinforcement Learning. *Biological Cybernetics, 73*, 265–274.

Dominey, P. F. (1997). An Anatomically Structured Sensory-Motor Sequence Learning System Displays Some General Linguistic Capacities. *Brain and Language*, *59*, 50–75.

Dominey, P. F. (1998a). A shared system for learning serial and temporal structure of sensori-motor sequences? Evidence from simulation and human experiments. *Cognitive Brain Research, 6*, 163–117.

Dominey, P. F. (1998b). Influences of Temporal Organization on Transfer in Sequence Learning: Comments on Stadler (1995) and Curran and Keele (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 234–248.

Dominey, P. F. (2002). A model of learning syntactic comprehension for natural and artificial grammars. In E. Witruk, A. Friederici, and T. Lachmann (Eds.), *Basic Functions of Language, Reading, and Reading Disability* (pp. 61–78). Dordrecht: Kluwer.

Dominey, P. F., Arbib, M. A., & Joseph, J. P. (1995). A Model of Cortico-Striatal Plasticity for Learning Oculomotor Associations and Sequences, *Journal of Cognitive Neuroscience, 7*, 311–336.

Dominey, P. F. & Georgieff, N. (1997). Schizophrenics learn surface but not abstract structure in a serial reaction time task. *NeuroReport, 8*, 2817–2822.

Dominey, P. F. & Lelekov, T. (2000). Non-Linguistic Transformation Processing in Agrammatic Aphasia, Comment on Grodzinsky: The neurology of syntax: language use with and without Broca's area. *Behavioral and Brain Sciences*, *23*, 34.

Dominey, P. F., Lelekov, T., Ventre-Dominey, J., & Jeannerod, M. (1998). Dissociable Processes for Learning the surface and abstract structure sensorimotor sequences. *Journal of Cognitive Neuroscience*, *10*, 734–751.

Dominey, P. F. & Ramus, F. (2000). Neural network processing of natural lanugage: I. Sensitivity to serial, temporal and abstract structure of language in the infant. *Language and Cognitive Processes*, *15*, 87–127.

Dominey, P. F., Ventre-Dominey, J., Broussolle, E., & Jeannerod, M. (1997). Analogical Transfer is Effective in a Serial Reaction Time Task in Parkinson's Disease: Evidence for a Dissociable Sequence Learning Mechanism. *Neuropsychologia, 35*, 1–9.

Doya, K. (1995). Recurrent networks: supervised learning. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks* (pp. 796–800). Cambridge, MA: MIT Press.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14*, 179–211.

Ferraro, F. R., Balota, D. A., & Connor, L. T. (1993). Implicit Memory and the Formation of New Associations in Nondemented Parkinson's Disease Individuals and Individuals with Senile Dementia of the Alzheimer Type: A Serial Reaction time (RT) Investigation. *Brain and Cognition, 21*, 163–180.

Frensch, P. A. & Miner, C.S. (1994). Effects of presentation rate and individual differences in short-term memroy capacity on an indirect measure of serial learning. *Memory & Cognition, 22*, 95–110.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological Research, 61*, 83–98.

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and Analogical Transfer. *Cognitive Psychology, 15*, 1–38.

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology, 33*, 154–207.

Gomez, R. L. & Schaveneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 396–410.

Grodzinsky, Y. (2000). The neurology of syntax: language use without Broca's area. *Behavioral and Brain Sciences,* 23, 1–21.

Hoen, M. & Dominey, P. F. (2000). ERP analysis of cognitive sequencing: a left anterior negativity related to structural transformation processing. *Neuroreport. 28*, 3187–3191.

Holyoak, K. J., Junn, E. N., & Billman, D. O. (1984). Development of Analogical Problem-solving Skill, *Child Development, 55*, 2042–2055.

Holyoak, K. J., Novick, L. R., & Melz, E. R. (1994). Component processes in analogical transfer: Mapping, pattern completion, and adaptation, In K. Holyoak and J. Barnden (Eds.), *Advances in Connectionist and Neural Computation Theory, Vol 2: Analogical Connections* (pp. 113–180). Norwood NJ: Ablex.

Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard, C. (1995). Serial Reaction Time Learning and Parkinson's Disease: Evidence for a Procedural Learning Deficit. *Neuropsychologia, 33*, 577–593.

Jiménez, L. & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 236–259.

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 948–969.

Keele, S. W. & Jennings, P. J. (1992). Attention in the representation of sequence: Experiment and Theory. *Human Movement Science, 11*, 125–138.

Kermadi, I., Jurquet, Y., Arzi, M., & Joseph, J. P. (1993). Neural activity in the caudate nucleus of monkeys during spatial sequencing. *Experimental Brain Research*, *94*, 352–356.

Knowlton, B. J. & Squire, L. R. (1993). The learning of categories: Paralles brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.

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

Knowlton, B. J. & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 169–181.

Lieberman, P. (1992). Could an autonomous syntax module have evolved? *Brain and Language*, *43*, 768–774.

Ljungberg, T., Apicella, P., & Schultz, W. (1991). Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Research, 567*, 337–341.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of Implicit and Explicit Processing in Learning From Examples: A Synergistic Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1083–1100.

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

Pascual-Leone, A., Grafman, J., Clark, K., Stewart, M., Massaquoi, S., Lou, J. S., & Hallett, M. (1993). Procedural Learning in Parkinson's Disease and Cerebellar Degeneration. *Annals of Neurology, 34*, 594–602.

Pastor, M. A., Artieda, J., Jahanshahi, M., & Obeso J. A. (1992). Time estimation and reproduction is abnormal in Parkinson's disease. *Brain, 115*, 211–225.

Perruchet, P. & Pacteau, C. (1991). Implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. *Journal of Experimental Psychology: General, 120*, 112–116.

Reddington, M. & Chater, N. (1996). Transfer in Artificial Grammar Learning: A Reevaluation. *Journal of Experimental Psychology: General, 125*, 123–138.

Reading, P. J., Dunnet, S. B., & Robins, T. W. (1991). Dissociable roles of the ventral, medial and lateral striatum on the acquisition and performance of a complex visual stimulus-response habit. *Behavioral Brain Research, 45*, 147–161.

Robins, T. W., Giardini, V., Jones, G. H., Reading, P. J., & Sahakian, B. J. (1990). Effects of dopamine depletion from the caudate-putamen and nucleus accumbens septi on the acquisition and performance of a conditional discrimination task. *Behavioral Brain Research, 38*, 243–261

Rolls, E. T. & Williams, G. V. (1987). Sensory and Movement-Related Neuronal Activity in Different Regions of the Primate Striatum. In J. S. Schneider and T. I. Lidsky (Eds.), *Basal Ganglia and Behavior* (pp. 37–59). Seattle: Hogrefe & Huber Publishers.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of Dissociable Memory Systems. *Behavioral and Brain Sciences, 17*, 367–447.

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

Stadler, M. A. (1995). The Role of Attention in Implicit Learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 674–685.

Werbos, P. J. (1995). Backpropagation: Basics and new developments. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks* (pp. 134–139). Cambridge, MA: MIT Press.

Willingham, D. B., Greenberg, A. R., & Thomas, R. C. (1997). Response-to-stimulus interval does not affect implicit motor sequence learning, but does affect performance. *Memory & Cognition, 25*, 534–542.

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.

# Temporal effects in sequence learning

Arnaud Destrebecqz and Axel Cleeremans*
Université Libre de Bruxelles

## 1. Introduction

Implicit learning is a complex domain in which many issues central to the cognitive neurosciences arise. Indeed, over the past decade or so, the field has come to embody ongoing questioning about three fundamental issues, namely (1) consciousness, and specifically how we should conceptualize and measure the relationships between conscious and unconscious cognition; (2) modularity and the architecture of the cognitive system, and in particular whether one should think of implicit learning as being subtended by specific, dedicated regions of the brain or not; and (3) mental representation, and in particular the complex issue of abstraction.

In this paper, we focus essentially on the first two issues. Based on a theoretical perspective (e.g., Cleeremans & Jiménez 2002) that takes it as a starting point that the differences between implicit and explicit knowledge might be differences in degree rather than in kind, we explore how the development of each can be selectively influenced by the time made available to participants to process each stimulus in a sequence learning situation. We also describe a computational model of performance in our experimental situation that offers a *graded* and *single-process* account of the interactions between implicit and explicit knowledge. The central idea embodied by the model is that developing high-quality, explicit representations of the stimulus material takes time. We start by offering some background on the reasons why assessing the respective contributions of implicit vs. explicit knowledge in implicit learning situations is such a difficult problem, and continue this introduction by discussing temporal effects in the context of sequence learning situations.

## 1.1 Direct and indirect measures in sequence learning

Sequence learning has now become the best behavioral paradigm through which to study implicit learning (see Clegg, DiGirolamo, & Keele 1998; Cleeremans, Destrebecqz, & Boyer 1998 for reviews). A sequence learning task is typically divided into two phases: (1) a training phase, during which subjects perform a serial reaction time (SRT) task, and (2) a test phase, during which (explicit) knowledge of the sequence is assessed. On each trial of the SRT task, a stimulus appears at one of several locations (typically 4 to 6) on a computer screen. Participants are simply asked to respond by pressing as fast and as accurately as they can on one of several keys (each corresponding to one of the locations at which the stimulus may appear) arranged in a spatially compatible layout. Unknown to them, the sequence of successive stimuli either follows a simple repeating pattern (Nissen & Bullemer 1987), or is produced by generation rules that describe legal transitions between successive sequence elements (e.g., Cleeremans & McClelland 1991).

In the context of SRT tasks, sensitivity to the sequential structure of the material can be revealed through comparisons between reaction times to various types of stimuli. All these measures assume that reaction time, in general, will tend to reflect the extent to which the stimulus that is responded to is predictable or familiar in its temporal context. Thus for instance, reaction times tend to progressively decrease with practice on a given sequence, but to increase when the repeating pattern is modified in any way (Reed & Johnson 1994) – thus indicating that whatever knowledge participants have acquired over the course of training is specific to the sequence they have been exposed to. Reaction times also tend to be faster when the sequence is repeated rather than random (Frensch & Miner 1994). Finally, participants also respond faster to grammatical than to ungrammatical stimuli when the sequence is generated by an artificial grammar (Cleeremans & McClelland 1991). Importantly, all these measures are *indirect measures* to the extent that participants are not told that the material contains sequential structure. The discriminations they reflect are therefore not directly required by task instructions (Jiménez, Méndez, & Cleeremans 1996), and the corresponding learning effects are thus best described as incidental.

In the same way, various tasks performed during the test phase have been used as *direct measures* of sequence knowledge. The most widely used such direct measures are generation and recognition tasks. In a generation task, participants are asked to indicate the location of the next stimulus of the sequence rather than to react to the current one. In a recognition task, participants are presented with small sequence fragments and asked to classify them as instances of the training material or not. Most of the studies using such forced-choice tasks have shown strong associations between performance on these tests and the indirect measure of learning in the SRT task (e.g., Perruchet & Amorim 1992; Shanks & Johnstone 1999). These associations are usually taken as evidence for explicit rather than implicit sequence learning, for they suggest that whatever knowledge people have acquired over training on the SRT task can also be deployed intentionally, in response to explicit instructions to do so.

However, this conclusion that sequence learning is based on the conscious acquisition of sequential knowledge depends on the plausibility of a critical assumption, namely that the direct task used to assess knowledge of the sequence constitutes an exclusive and exhaustive measure of explicit knowledge. This assumption, however, is simply unwarranted (see Jiménez 1997, for a detailed analysis). There is now extensive evidence that no task can be taken to be process-pure (Reingold & Merikle 1988; Jacoby 1991), that is, that no task can be taken to exclusively reflect the operation of a single cognitive process. Rather, any task will always tend to involve many different cognitive processes operating simultaneously. In the context of sequence learning research, for instance, Shanks and Johnstone (1998) have shown that participants who perform above chance in a generation task may nevertheless believe that they are guessing the sequence of locations – thus suggesting that whatever knowledge was expressed during the generation task might also depend on implicit influences rather than exclusively on explicit recollection.

Assessing the relative contributions of implicit and explicit knowledge on sequence learning is therefore a critical issue that requires sensitive methods to be used. With this goal in mind, we recently adapted the process dissociation procedure (PDP; Jacoby 1991) to sequence learning (Destrebecqz & Cleeremans 2001; see also Buchner, Steffens, Erdfelder, & Rothkegel 1997; Buchner, Steffens, & Rothkegel 1998; Goschke 1997, 1998, for similar attempts). The process dissociation procedure was originally devised to make it possible to dissociate implicit and explicit forms of memory. It is based on comparing performance on two tasks that differ only by the instructions: the inclusion and the exclusion task. In the context of sequence learning, consider for instance a generation task performed under inclusion instructions. Participants are told to produce a sequence that resembles the training sequence. To do so, they can either explicitly recollect the regularities of the training sequence, or they can guess the location of the next stimulus based on intuition or familiarity. Hence, under inclusion instructions, both implicit (e.g., intuition) and explicit (e.g., recollection) processes can contribute to improve performance. Consider now the same generation task, this time performed under exclusion instructions. Participants are told to generate a sequence that differs as much as possible from the training sequence. Implicit and explicit influences are now set in opposition, for the only way to successfully avoid producing familiar sequence elements is to consciously know what the training sequence was and to produce something different. Observing the continued generation of familiar elements under exclusion instructions would thus clearly indicate that generation is automatically influenced by implicit knowledge.

Based on these ideas, we applied this logic in a previous study (Destrebecqz & Cleeremans 2001) in which we also attempted to manipulate the extent to which learning was explicit. To do so, we explored how changes in the response-to-stimulus interval (RSI; the amount of time that elapses between the response and the onset of the next stimulus) in the SRT task influenced sequence learning. Our main hypothesis was that reducing the value of the RSI to 0 ms might selectively impair the development of conscious expectations about the location of the next stimulus. This hypothesis was confirmed. Indeed, while all participants were able to learn about the sequential regularities contained in the material and to project this knowledge in a generation task performed under inclusion instructions, only participants trained with an RSI of 250 ms were able to successfully exclude their knowledge in a generation task performed under exclusion instructions. Participants trained with an RSI of 0 ms instead tended to continue to preferentially reproduce the regularities of the training sequence under exclusion instructions. We suggested that these participants lacked conscious control over their sequential knowledge. This impression was further strengthened by the results of a recognition task showing that these participants were unable to correctly discriminate between old and new sequence fragments. In contrast, participants trained with an RSI of 250 ms were perfectly capable of discriminating between novel and familiar sequence fragments.

Based on these results, we therefore concluded (1) that direct tasks such as generation or recognition may indeed be influenced by implicit knowledge, (2) that our data offered clear evidence that learning can be unconscious to the extent that the relevant knowledge may influence performance yet remain outside conscious control and recollection, and (3) that the time available for processing each stimulus during the SRT task is critical in determining to extent to which sequence knowledge is available to conscious awareness.

This third conclusion, – that explicit sequence learning depends on the time available to process each event over the course of the SRT task – is somewhat speculative because existing studies do not offer a coherent picture of the influence of temporal factors on sequence learning. In the remainder of this introduction, we would therefore like to focus on the importance and nature of these temporal effects in sequence learning.

## 1.2 Temporal effects in sequence learning

The results of the experiment described above are consistent with some, but not with all of the previous studies that explored the importance of temporal factors on sequence learning (e.g., Hsiao & Reber 2001). Such temporal factors were proposed as a way of understanding the effects of a secondary task on sequence learning performance. In the "dual-task" version of the SRT experiment, introduced by Nissen and Bullemer (1987), either a low- or a high-pitched tone is produced during the RSI. Instructions require participants not only to respond to each stimulus location (the primary task) but also to keep a running count of how many low-pitched tones have occurred during each block (the secondary task). In their original study, Nissen & Bullemer argued that a secondary tone-counting task impairs sequence learning because it exhausts participants' attentional resources.

Other authors have instead suggested that the detrimental effect of the tone-counting task is due to scheduling conflicts between performing the main and secondary tasks, rather than to attentional load. Stadler (1995), for instance, argued that a secondary tone-counting task impairs sequence learning, not because it divides attention, but because it introduces variability in the RSI. Stadler pointed out that the secondary task lengthens the RSI only for the target trials in which the tone count must be updated. This incidental lengthening of the RSI would then have effects similar to those resulting from actually inserting a pause between those trials and the next. The pauses would disrupt participants' ability to parse the sequence into consistent chunks – a process which, according to Stadler, is essential to sequence learning.

Frensch and Miner (1994), in contrast, attribute the detrimental effects of the secondary task to short-term memory limitations: Secondary tasks impair sequence learning not so much because they make it hard for participants to chunk the sequence consistently, but simply because they lengthen the response-to-stimulus interval (RSI), and that this lengthening makes it more difficult for participants to link together the memory traces corresponding to successive elements of the sequence in short-term memory. Consistently, Frensch and Miner (1994) reported that sequence learning is impaired when the RSI is increased to the unusual value of 1500 ms.

Yet another hypothesis about the effects of the RSI on sequence learning was put forward by Willingham et al. (1997), who argued that lengthening the RSI does not impair sequence learning *per se*, but only the expression of knowledge about the sequence. Willingham et al. reported that participants trained with a 1500 ms RSI showed impaired sequence learning as compared to participants trained with a 500 ms RSI. However, when transferred to a shorter RSI, the former group showed the same level of sequence learning as the latter.

At first sight, the Destrebecqz and Cleeremans (2001) study described above appears to contradict these results. Indeed, we found that a higher value of the RSI tends to improve *explicit* sequence learning. We argued that people trained with an RSI of 250 ms are given more opportunities to link together high-quality memory traces and to develop stronger representations of the sequential constraints of the training material – an account that is totally inconsistent with the findings of Frensch and Miner (1994), for instance.

To further explore and clarify the role of temporal factors on sequence learning, we therefore conducted a new experiment, presented below, in which we manipulated the RSI over three different values (RSI = 0, 250, or 1500 ms), and in which performance was assessed through a wider array of objective and subjective measures. This study had two main goals: (1) to confirm the fact that sequence learning is implicit when the RSI is reduced to 0 ms, and (2) to explore the effects on sequence learning of a major increase of the RSI, from 250 ms to 1500 ms. Several predictions are possible about the effects of such a large RSI on sequence learning performance.

First, we might observe reduced sequence learning, based on the notion put forward by Frensch and Miner that working memory limitations will make it more difficult for participants to link together relevant memory traces during learning. If this were the case, we would thus expect to observe impaired performance on both the SRT task and on the subsequent generation and recognition tasks under RSI = 1500 ms conditions.

Second, it might also be the case, as Willingham et al. argued, that increasing the RSI from 250 ms to 1500 ms would have no effect on sequence learning in and of itself, but only influence SRT performance. In other words, under RSI = 1500 ms conditions, we would thus expect to observe a deterioration in SRT performance, but intact generation and recognition performance.

Finally, based on our own previous findings, we might expect to observe improved *explicit* sequence learning under RSI = 1500 ms condition. Indeed, our theory predicts that people develop stronger (and hence more explicit) representations with higher values of RSI. If this theory were correct, we would thus expect improved *explicit* sequence learning under RSI = 1500 ms condition as assessed namely by generation performance.

## 2. Method

### 2.1 Participants

72 participants, all undergraduate students of the Université Libre de Bruxelles were randomly assigned to one of three experimental conditions.

### 2.2 Material

The experiment was run on Macintosh computers. The display consisted of four dots arranged in a horizontal line on the computer's screen and separated by intervals of 3 cm. Each screen position corresponded to a key on the computer's keyboard. The spatial configuration of the keys was fully compatible with the screen positions. The stimulus was a small black circle 0.35 cm in diameter that appeared on a white background, centered 1 cm above one of the four dots.

### 2.3 Procedure

The experiment consisted of 15 training blocks during which participants were exposed to a serial four-choice RT task. Each block consisted of 96 trials, for a total of 1440 trials. On each trial, a stimulus appeared at one of the four possible screen locations. Participants were instructed to respond as fast and as accurately as possible by pressing on the corresponding key. The target was removed as soon as a key had been pressed, and the next stimulus appeared after either a 0 ms (RSI 0 condition), 250 ms (RSI 250 condition), or 1500 ms (RSI 1500 condition) interval depending on the condition. Erroneous responses were signaled to participants by means of a tone. Short rest breaks occurred between any two experimental blocks. Participants were presented with one of the following twelve elements sequences: 342312143241 (SOC1), 341243142132 (SOC2). Each experimental block consisted of eight repetitions of the sequence. These sequences consisted entirely of "second order conditional" transitions or SOCs (Reed & Johnson 1994). With SOC sequences, two elements of temporal context are always necessary to predict the location of the next stimulus. Both sequences were balanced for stimulus locations and transition frequency but differed in terms of the subsequences of three elements that they contained. For instance, the transition ‘34’ was followed by location 2 in SOC1 and by location 1 in SOC2. In each condition, half of the participants were trained on SOC1 during the first 12 blocks and during blocks 14 and 15; and on SOC2 during block 13. This design was reversed for the other half of the participants. Increased RTs during block 13 are thus expected only if participants have acquired SOC knowledge during training over blocks 1–12.

### 2.3.1 *Verbal reports*

After the SRT task, participants were presented with the following five propositions and asked to indicate the one that best described the sequence of stimuli:

1. The sequence of stimuli was random.
2. Some positions occurred more often than others.
3. The movement was often predictable.
4. The same sequence of movements would often appear.
5. The same sequence of movements occurred throughout the experiment.

This questionnaire is identical to the one used by Curran (1997). After completing the questionnaire, all participants were informed that the sequence of stimuli was not random. They were then introduced to either the generation or the recognition task. In each condition, half of the participants performed the recognition task before the generation task. The opposite order was used for the other half of the participants to control for a possible order effect.

### 2.3.2 *Generation task*

In the generation task, participants were presented with a single stimulus that appeared in a random location, and asked to freely generate a series of 96 trials that “resembled the training sequence as much as possible”. They were told to rely on intuition when feeling unable to recollect the location of the next stimulus. After this generation task – performed under inclusion instructions – participants were asked to generate another sequence of 96 trials, this time under exclusion instructions. They were told that they now had to try to avoid reproducing the sequential regularities contained in the training sequence. In both generation tasks, participants were also told not to repeat responses. The stimulus moved whenever participants had pressed one of the keys, and appeared at the corresponding location after a delay of either 0 ms, 250 ms or 1500 ms depending on the condition.

#### 2.3.3 *Confidence test*

After completion of both generation tasks, participants were asked to perform a confidence test in which they had to rate, on a scale from 0 to 100, how confident they were in their generation performance. They had to evaluate how well they thought they had been able to reproduce the sequence during the inclusion task, and how well they thought they had been able to avoid reproducing the sequence under exclusion instructions.

#### 2.3.4 *Recognition test*

In the recognition task, participants were presented with 24 fragments of three trials. Twelve were part of SOC1 and twelve were part of SOC2. Participants were asked to respond to the stimuli as in the SRT task, and to subsequently provide a rating of how confident they were that the fragment was part of the training sequence. Ratings involved a six points scale where 1 = "I'm certain that this fragment was part of the training sequence", 2 = "I'm fairly certain that this fragment was part of the training sequence", 3 = "I believe that this fragment was part of the training sequence", 4 = "I believe that this fragment was not part of the training sequence", 5 = "I'm fairly certain that this fragment was not part of the training sequence", and 6 = "I'm certain that this fragment was not part of the training sequence" (Shanks & Johnstone 1999). It was also emphasized to participants that they had to respond as fast as possible to the dots. Both ratings and reaction times were recorded.

## 3. **Results**

### 3.1 Reaction time task

As the two sub-groups of participants presented in each condition with either SOC1 or SOC2 were trained identically, their reaction times were combined for subsequent analyses.

Figure 1 shows the average reaction times obtained over the entire experiment, plotted separately for the three conditions. To analyze the data, we performed an ANOVA with Blocks [15 levels] as a within-subject variable and Condition [three levels] as a between-subjects variable. This analysis revealed significant effects of Blocks [$F(14, 966) = 29.182$, $p < .0001$, $Mse = 54543.548$] and of Condition [$F(2, 69) = 11.49$, $p < .0001$, $Mse = 1197844.526$]. The Block × Condition interaction also reached significance [$F(28, 966) = 3.038$, $p < 0.001$, $Mse = 5678.831$]. The significant effect of Condition indicates that

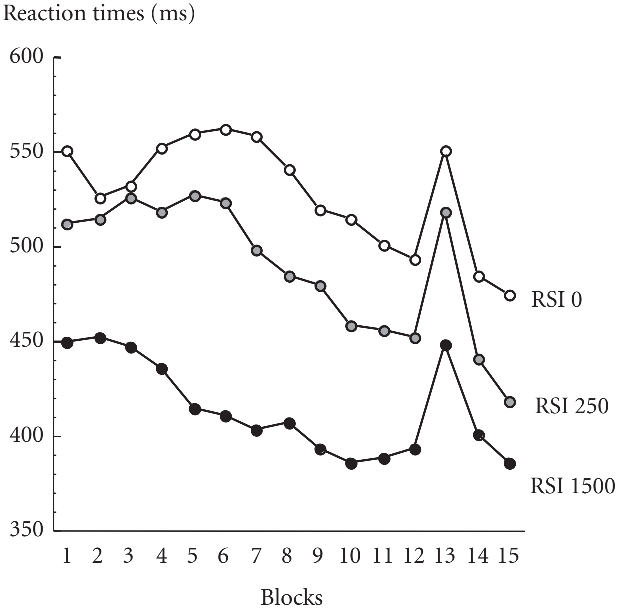


**Figure 1.** Mean reaction times for the 15 blocks of the SRT task, plotted separately for the three conditions.

participants respond reliably faster when the RSI increases. This effect however may simply stem from improved motor performance rather than from improved learning.

Indeed, the increase in reaction time between blocks 12 and 13 suggests that participants learn the sequence in all three conditions. To validate this impression and assess the magnitude of the transfer effect, we compared RTs obtained for the transfer block (block 13) with the average of RTs obtained for blocks 12 and 14. An ANOVA with Condition [three levels] as a between-subjects variable, and Block [2 levels] as a within-subject variable revealed significant effects of Block [$F(14, 69) = 86.894$, $p < .0001$, $Mse = 136571.450$] and of Condition [$F(2, 69) = 7.672$, $p < .01$, $Mse = 1197844.526$]. The Block × Condition interaction, however, was not significant. These results therefore confirm that participants learn about the sequential structure of the material in all three conditions, and that the magnitude of the transfer effect (our indirect measure of learning) is not significantly influenced by the value of the RSI.

## 3.2 Verbal reports

In this task, participants were presented with a five items questionnaire describing the structure of the training sequence, and asked to choose the proposition that best described their subjective appraisal of the material. The five propositions (1 to 5) described increasing degrees of sequential structure. To quantify participants' subjective knowledge, we simply scored each participant by recording the rank of the proposition they chose. A high score is therefore expected if participants noticed a systematic repeating pattern. Low scores would indicate that participants considered that the sequence was random. The average scores of subjective knowledge were 3.83, 3.92, and 3.67 for the RSI 0, RSI 250, and RSI 1500 respectively. A one-way ANOVA performed on these scores failed to reveal a significant effect of Condition ($F < 0.5$).

### 3.2.1 *Generation task*

Because tasks order failed to exert any effect on generation or recognition performance, the data for both groups of participants were collapsed in all three conditions.

To assess generation performance, we computed the number of generated chunks of three elements that were part of the training sequence in both inclusion and exclusion tasks. Since the generated sequences consisted of 96 trials, the maximum number of correct chunks that can be produced is 94. To obtain inclusion and exclusion scores for each participant, we therefore divided the corresponding number of correct chunks by 94. As participants were told not to produce repetitions, chance level is 0.33.

Figure 2 shows average inclusion and exclusion scores for the three conditions. An ANOVA with Condition [3 levels] as a between-subjects variable and Instructions [2 levels, inclusion vs. exclusion] as a within-subject variable only revealed a significant effect of Instructions [$F(1, 66) = 44.885$, $p < .0001$, Mse = 1.039]. The Instructions × Condition interaction was marginally significant [$F(2, 66) = 2.814$, $p < 0.07$, Mse = 0.065]. Condition failed to reach significance.

*T* tests were used to compare generation scores to chance level. Inclusion scores are significantly above chance level in all three conditions, $t(23) = 4.57$, $p < .001$, $t(23) = 8.71$, $p < .001$, $t(23) = 4.45$, $p < .001$ for the RSI 0, RSI 250 and RSI 1500 respectively. Planned comparisons performed on inclusion scores did not reveal any significant differences (all *F*s < 1.7). Exclusion scores are also *above* chance level in the RSI 0 condition [$t(23) = 2.23$, $p < .05$]. By contrast, they are significantly *below* chance level in the RSI 250 condition [$t(23) =$ -

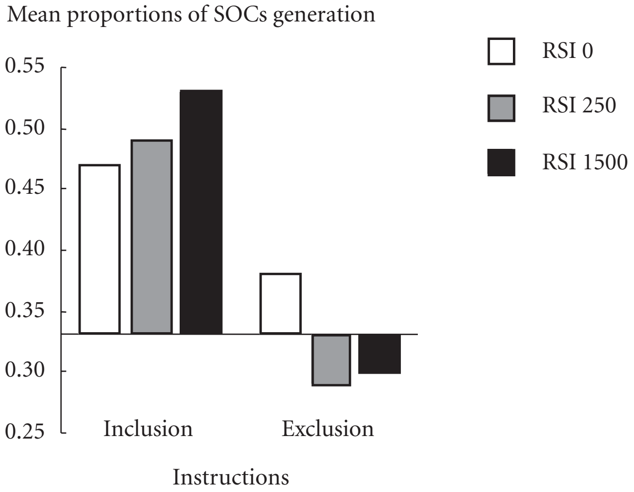


**Figure 2.** Mean generation scores for the three conditions and under inclusion and exclusion instructions.

2.18, p < .05]. Exclusion scores did not differ from chance level in the RSI 1500 condition [*t*(23) = -1.22]. Planned comparisons showed that exclusion scores were significantly higher in the RSI 0 condition than in both the RSI 250 [F(1, 66) = 7.53, p < .01, Mse = 0.0118] and the RSI 1500 condition [F(1, 66) = 2.814, p < .05, Mse = 0.0118]. Exclusion scores in the latter conditions did not differ from each other.

Planned comparisons further revealed that the number of generated chunks corresponding to the training sequence decreased significantly from inclusion to exclusion instructions in the RSI 250 [F(1,66) = 20.83, p < .0001, Mse = 0.0231] and RSI 1500 conditions [F(1,66) = 25.863, p < .0001, Mse = 0.0231]. However, this difference was only marginally significant in the RSI 0 condition [F(1,66) = 3.08, p < .06, Mse = 0.0231].

To summarize, these results suggest that participants have control on their knowledge in the RSI 250 and RSI 1500 conditions, but not in the RSI 0 condition. It is only in this latter condition, indeed, that participants produced triplets of the training sequence above chance level in both inclusion and exclusion tasks.

We further analyzed exclusion performance by comparing the number of produced triplets that are congruent either with the training or the transfer sequence. As noted by Reed and Johnson (1994) and by Shanks and Johnstone (1999), this analysis is important, for the following reason: Assume participants notice, over training, that the material contains regularities such as "the number of reversals is low". Crucially, such regularity is not specific to the training sequence, but applies just as well to the transfer sequence. During free generation, participants may thus produce responses that reflect such general knowledge of the material, and hence end up performing better than chance (0.33) would predict, despite not having learned the specific SOCs present in the training material. If this were the case, the "correct" chance level might in fact be different from 0.33. To control for this possibility, we evaluated how many generated triplets are congruent with the *transfer* sequence. The number of generated triplets that are congruent with the transfer sequence is the appropriate control because the training and transfer sequences differ by every SOC transition but are identical in all other structural aspects. Thus, for instance, they contain exactly the same number of reversals (such as "121" in SOC1 and "323" in SOC2).

If participants had learned the training sequence explicitly, then they should be able to control the expression of their sequential knowledge. We would thus not expect that they produce more triplets from the training sequence than from the transfer sequence in the exclusion task. That is the pattern of results we observed for RSI 250 and RSI 1500 participants, who produced equal numbers of training and transfer triplets under exclusion instructions (*p*s > 0.6). By contrast, RSI 0 participants produced more training (35.75) than transfer triplets (29.79) in the exclusion task [$t(23) = 2.45$, $p < .05$, bilateral]. This result can only be attributed to the automatic influence exerted on generation performance by sequence knowledge acquired implicitly during the SRT task in this latter condition.

### 3.2.2 *Subjective tests*

After each generation task, participants had to rate how confident they were in their performance on a scale ranging from 0 to 100. Figure 3 shows mean inclusion and exclusion confidence ratings for the three conditions. The figure suggests that participants are more confident in their exclusion than in their inclusion performance. This impression was confirmed by the results of an ANOVA conducted on these confidence ratings with Condition [three levels] as a between-subjects variable and Instructions [2 levels] as a within-subject variable. This analysis revealed a significant effect of Instructions [$F(1,66) = 66.482$, $p < .0001$, Mse = 1.937]. Neither Condition nor the Instructions × Condition interaction reached significance (*F*s < 1.6).

To obtain a more detailed appraisal of the relationships between generation performance and confidence, we also computed correlations between genera-

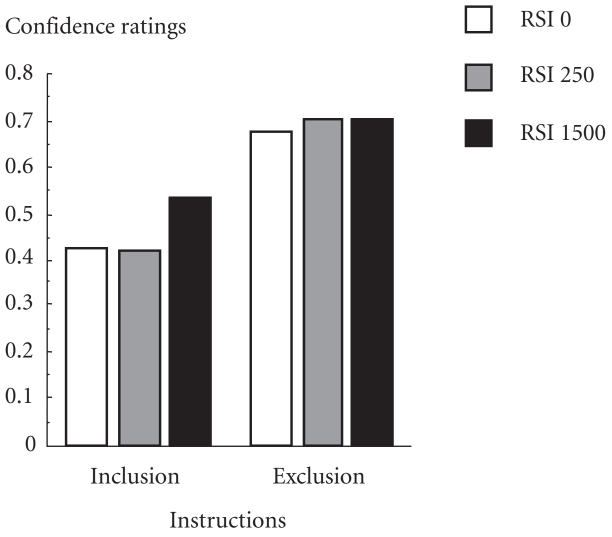


**Figure 3.** Mean confidence ratings for inclusion and exclusion tasks plotted separately for the three conditions.

**Table 1.** Correlations between generation tasks and the corresponding confidence level for the three conditions

| | | *Confidence in inclusion* | *Confidence in exclusion* |
|---|---|---|---|
| | *Inclusion score* | 0.239 | |
| ***RSI 0*** | *Exclusion score* | | –0.343 |
| | *Inclusion score* | –0.082 | |
| ***RSI 250*** | *Exclusion score* | | –0.358 |
| | *Inclusion score* | 0.464* | |
| ***RSI 1500*** | *Exclusion score* | | 0.213 |

*Note.* * indicates a significant correlation ($p < 0.05$)

tion scores and confidence levels in both inclusion and exclusion tasks (see Table 1). This analysis showed a significant correlation between inclusion scores and confidence ratings, but only in the RSI 1500 condition.

### 3.2.3 *Recognition task*

Participants were presented with 24 short sequences of three elements. Half of these triplets were part of the training sequence and the other half were part of the transfer sequence. For each triplet, participants first responded to each of

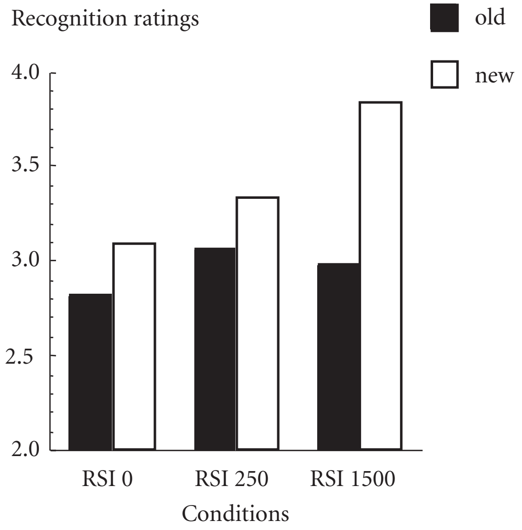


**Figure 4.** Mean recognition ratings for the three conditions and for old and new sequences of three elements. A high rating (between (4 and 6) is expected for a new sequence, and a low rating (between 1 and 3) is expected for an old sequence.

its three elements just as in the SRT task, and were then asked to rate the extent to which they felt the triplet was part of the training sequence or not.

To control for the possibility that recognition ratings are influenced by perceptual-motor fluency during the test itself (as could be the case if participants responded faster to familiar triplets), we first compared RTs to old and new items. Because old and new triplets differed exclusively by their third stimulus, our analysis was limited to the RTs associated with responses to the third element of each triplet. An ANOVA with Condition [3 levels] as a between-subjects variable and Sequence [2 levels, old vs. new] as a within-subject variable conducted on these data revealed no significant effect. We conclude that recognition scores can therefore be safely attributed to explicit recollection.

Mean recognition ratings for the three conditions and for both types of sequence are shown in Figure 4. An ANOVA with Condition [3 levels] as a between-subjects variable and Sequence [2 levels, old vs. new] as a within-subject variable conducted on these data revealed significant effects of Sequence [$F(1, 66) = 23.018$, $p < .0001$, $Mse = 7.793$] and of Condition [$F(2, 66) = 3.475$, $p < .05$, $Mse = 2.441$]. The Sequence × Condition interaction also reached significance [$F(2, 66) = 3.825$, $p < .05$, $Mse = 1.295$]. Planned comparisons revealed a significant difference between old and new triplets in the RSI 250 [F(1, 23) = 4.764, p < .05, Mse = 0.878] and RSI 1500 [F(1, 23) = 13.735, p < .01, Mse = 8.56] conditions. Sequence was only marginally significant in the RSI 0 condition [F(1, 23) = 3.531, p < .08, Mse = 0.946]. These recognition results suggest that participants trained in the RSI 0 condition lacked explicit sequence knowledge.

We also compared mean differences between recognition ratings attributed to old and new triplets. A one-way ANOVA performed on these data revealed a significant effect of Condition [F(2, 69) = 3.613, p < .05, Mse = 2.59]. Contrasts indicated that the mean difference between old and new triplets was higher in the RSI 1500 condition than in the RSI 0 and RSI 250 conditions (*p*s < .05), but that it did not differ between the two latter conditions.

## 4. Discussion

In this experiment, participants were trained on a standard SRT task under three different conditions differing by the length of the interval separating their response to each stimulus and the onset of the next stimulus (RSI = 0, 250, or 1500 ms).

Our results show that participants were able to learn about the second-order contingencies present in a sequence of visual stimuli independently of the value of the RSI. This is made plain by the fact that all participants exhibited a significant transfer effect when switched to a different sequence during Block 13. While reaction times were systematically faster for higher values of the RSI, the magnitude of the transfer effect failed to differ between the three conditions. These findings stand in contrast with the results of previous studies showing faster overall reaction times for shorter values of RSI (Frensch & Miner 1994; Willingham et al. 1997). At this point, we cannot offer a clear explanation of this discrepancy. It remains possible that various factors along which this study and previous ones differ – such as the complexity of the sequence or the amount of practice – interact with the value of the RSI.

Turning now to our direct measures, verbal reports are suggestive that participants in all three conditions acquired some knowledge of the fact that target movements were regular in the SRT task. Verbal reports do not indicate, however, that RSI influences sequence learning: On average, all participants expressed that "the same sequence of movements would often appear".

In contrast with the indirect measure of learning and verbal reports, performance on the generation tasks was clearly suggestive that manipulating the

RSI influenced sequence learning. Indeed, while inclusion performance was comparable in all three conditions – indicating that participants were able to use their knowledge of the sequence when directly instructed to do so, exclusion task results revealed that participants trained in the RSI 0 condition had no control on their sequential knowledge. In this condition only, participants tended to keep reproducing the training sequence above chance level in exclusion, despite instructions to the contrary. Moreover, these participants also tended to produce more second-order transitions characteristic of the training sequence than transitions characteristic of the transfer sequence, – an indication that their knowledge of these regularities was unconscious. Inclusion scores were quantitatively higher than exclusion scores but the difference was only marginally significant. It may therefore be the case that RSI 0 participants have acquired some explicit sequence knowledge, but their exclusion performance is undoubtedly attributable to the uncontrolled influence of implicit knowledge.

By contrast, in the RSI 250 and RSI 1500 conditions, participants exhibited control over their sequential knowledge: Inclusion scores exceeded exclusion scores, and did not appear to be influenced by implicit knowledge, since (1) exclusion performance was at chance in both conditions, and (2) training sequence triplets were not produced more frequently than transfer triplets. These conclusions are also supported by our recognition results, where only RSI 250 and RSI 1500 participants were able to discriminate between old and novel sequence fragments.

Did the increase of the RSI between 250 ms and 1500 ms influence sequence learning? We aimed to contrast several possible hypotheses: (1) increasing the RSI would impair sequence learning because working memory limitations would make it harder for the relevant memory traces to be linked together; (2) the increase might have no effect on sequence learning itself; (3) the increase might actually improve explicit sequence learning by making it possible for stronger representations of the sequential representations to develop during training.

Our results seem to favor this latter hypothesis. One way to ascertain the extent to which knowledge is explicit is to consider the extent to which confidences judgments relate with generation performance. We only found a significant correlation for RSI 1500 participants between confidence and inclusion performance. This result is suggestive that these participants only had relevant meta-knowledge about their knowledge. Following Dienes and Perner (1996), only RSI 1500 participants have *attitude explicitness* about their knowledge, that is, know that they know something about the sequence. Another way to make the same point is to consider the *zero-correlation criterion* (Dienes & Berry 1997), according to which learning is above the subjective threshold when confidence is related to performance and below the subjective threshold in the opposite case. Applying this analysis to our data suggests that knowledge is above the subjective threshold only for RSI 1500 participants.

Importantly, the results of the confidence judgment task do not imply in and of themselves that learning was fully explicit in the RSI 1500 condition and fully implicit in the RSI 0 and RSI 250 conditions. Rather, they suggest that the knowledge acquired when the RSI was set at 1500 ms is qualitatively different, and based on stronger representations of the sequence than with shorter values of RSI. This notion is also supported by the pattern of results we obtained in the recognition task, in which the difference between ratings attributed to old and new triplets was higher in the RSI 1500 than in the RSI 250 condition.

To summarize, we replicated the results of our previous study: RSI 0 participants who were denied preparation to the next stimulus in the SRT task learned the sequence but were unable (1) to refrain from expressing their knowledge under exclusion instructions and (2) to project this knowledge in a recognition task. These findings confirm that sequence learning can occur unconsciously. We also found new evidence suggestive of increasing degrees of explicitness as the RSI is increased. When the RSI was set at a standard 250 ms, participants acquired explicit sequence knowledge that they could both control in the exclusion task as well as recollect in the recognition task. Increasing the RSI to 1500 ms further allowed participants to acquire meta-knowledge about their knowledge, as suggested by the pattern of correlations between inclusion performance and confidence.

What sort of computational mechanisms might account for our data? In the next section, we propose a novel model of sequence learning that takes it at a starting point that implicit and explicit learning involve the very same processes, and explore the extent to which it is capable of simulating the temporal effects we have observed in both the SRT task an in the generation tasks to which our human participants were exposed.

## 5. Mechanisms of sequence learning

A natural starting point from which to begin thinking about how to capture our empirical results is the simple recurrent network (SRN) first introduced by Elman (1990) and subsequently adapted to sequence learning by Cleeremans and McClelland (1991). The SRN (see Figure 5) is a connectionist network that

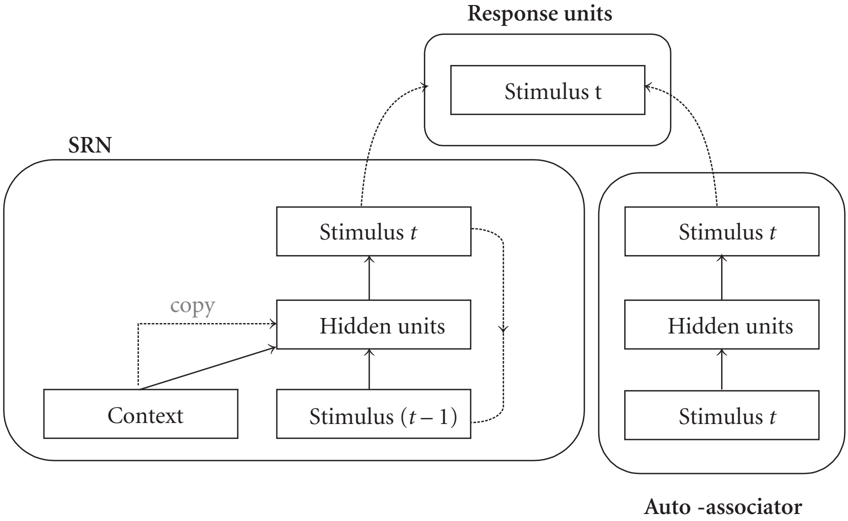


**Figure 5.** The network includes several elements: (1) an auto-associator (on the right), his task consists in identifying the current stimulus *t*, (2) the SRN (on the left) which is trained to predict the stimulus *t* based on the stimulus (*t* – 1) and a representation of the temporal context that the network has itself developed, and (3) a set of response units (on top) connected with the output layers of the SRN and the auto-associator by a series of fixed ”one-to-one” connections. The response units integrate the information brought by the SRN, on the one hand, and the auto-associator, on the other hand, in order to determine the model's response.

is trained, through backpropagation (Rumelhart, Hinton, & Williams 1986), to predict the next element of a sequence based on the current element. To perform this prediction task, the network includes a pool of so-called *context units*, which, on each time step, contains a copy of the pattern of activation that existed over the network's hidden units during the previous time step. Over training, these patterns of activation come to represent associations between one element and the next. When fed back over the context units, these representations then enable the network to base its prediction responses not only on the constraints that exist between the current input and the next element, but also on a representation of the temporal context that the network has developed itself over training. In this manner, the SRN becomes progressively capable of making prediction responses that take into account the information contained in an increasingly large and self-developed "temporal window" extending over several elements of the sequence (see Figure 5).

To simulate sequence learning performance, Cleeremans and McClelland (1991) assumed that the prediction task performed by the network represents preparation to the apparition of the next stimulus in human participants, and that reaction time is inversely related to the activation of the output unit corresponding to the element being responded to. These simple assumptions have proven adequate to account for the results of a wide variety of sequence learning experiments (see e.g., Cleeremans 1993; Cleeremans & Jiménez 1998).

As a model of sequence learning, however, the SRN suffers from two important limitations. First, the model as it stands is unable to account for the time course of processing during a single trial. In contrast to other models (see e.g., Dominey, this volume), the model, as most other networks based on backpropagation, indeed assumes that activation is propagated through the network in a single step upon presentation of each input. This of course renders the SRN *prima facie* incapable of accounting for the RSI effects we have described in this chapter.

Second, the SRN also assumes that all responses involve prediction. This is, however, not only inconsistent with the task demands that characterize the SRT task (in which subjects are required to respond to the current element of a sequence), but also makes the model incapable of accounting for the difference between identification (as in the SRT task) and prediction (as in the generation task). The model instead assumes that both identification and prediction involve exactly the same processes.

A better way to conceptualize the relationships between the SRT task and the generation task is to consider interactions between perception, memory, and action in both tasks. The SRT task essentially involves an encoding operation that takes a visual stimulus as input (perception) and produces a motor response as ouput (action). Going from perception to action can be influenced by memory, as countless sequence learning experiments demonstrate. The generation task, in contrast, involves a prediction operation that is potentially based on both memory and perception, but in which memory plays a predominant, if not exclusive role whenever the material contains higher-order structure, as is the case for SOC sequences.

In the following, we introduce a novel model of sequence learning that is aimed at addressing both of these shortcomings. To enable the model to capture the time course of processing during a single trial, we adapted the cascade algorithm initially introduced by McClelland (1979). To enable the model to capture the difference between SRT performance and generation performance, we modified its architecture in such a way that identification and prediction now involve separate components of the model.

## 5.1 Capturing the time course of processing: The cascade algorithm

McClelland (1979) proposed a relatively simple algorithm (the cascade model), which assumes that instead of being transmitted all at once, activation in a network can be propagated by small increments, each representing a set proportion of the total activation to be transmitted. Cohen, Dunbar, and McClelland (1990) further described how the cascade algorithm can be adapted in the context of backpropagation networks in order to include adaptive components and make it possible to capture the effects of experience.

In the cascade model, each trial is decomposed in a series of processing cycles during each of which only a fraction of the activation sent by each unit is propagated, depending on a constant rate of transmission ($\tau$). As a consequence the activation level of each unit changes proportionally to the rate of transmission (Cohen, Dunbar, & McClelland 1990).

As Equation 1 shows, in the cascade model, the activation of each unit is a running average of its net input over time. The net input to each unit is computed on each cycle, and depends on the information transmitted on the current cycle ($t$) and on the average value of the net input on the preceding cycle:

$$\mathrm{a}_j(t) = \mathrm{Net}_j(t) = \tau\,\mathrm{net}_j(t) + (1-\tau)\mathrm{Net}_j(t-1) \tag{1}$$

where $\mathrm{Net}_j(t)$ corresponds to the running average of the net input to the jth unit, and $\mathrm{net}_j(t)$ is the net input to the jth unit at time $t$. In backpropagation networks, the activation function needs to be non-linear. Typically, the logistic function is used, and the cascade equation is thus modified as in equation 2 to reflect this (see Cohen et al. 1990):

$$\mathrm{a}_j(t) = 1/\left[1 + e^{-\mathrm{Net}_{j(t)}}\right] \tag{2}$$

where $\mathrm{Net}_j(t)$ is computed according to equation 1.

## 5.2 Differentiating between identification and prediction

To make it possible for the model to account for the difference between identification and prediction tasks, we used the architecture shown in Figure 5. The model incorporates three components: A “memory” component consisting of an SRN network, a “perception” component consisting of an auto-associator network, and an “action” component consisting of a small pool of response units. As a simplifying assumption, memory and perception are taken to be completely independent, and the corresponding SRN and auto-associator networks are thus trained separately. Both networks produce outputs that correspond to element *t* of a sequence.

In the perception component, the auto-associator network is presented, on each trial, with an input corresponding to element *t* of the sequence, and is trained to produce this same element on its pool of output units. This process can be thought of as corresponding to the simple encoding operation that occurs in the context of a simple reaction time task. It also captures the task demands imposed upon subjects in the context of sequence learning tasks, for these demands make no mention of the fact that the material contains sequential structure. We know, however, that in such tasks, responses are also influenced by the temporal context. This influence is captured by the memory component of the model.

In this memory component, the SRN network is trained to produce element *t* of the sequence on its pool of output units, just as for the perception component. However, the SRN has to do so based only on a representation of the temporal context in which successive sequence elements occur. In our model, this representation of the temporal context consists of element $t-1$ of the sequence, together with the SRN's set of context units.

In our model, to capture the notion that both memory and perception influence action, the outputs of both the SRN and the auto-associator jointly influence the activation of a set of response units, as described in the next section

## 5.3 Response selection process

The action component of the model consists of four "leaky integrator" units (Usher and McClelland 2001), each corresponding to one of the four possible responses. Each of these units only receives input from the corresponding output units of both the SRN and the auto-associator units. For instance, the activation of the first response unit is only influenced by the first output unit of the SRN and by the first output unit of the auto-associator. Thus, each response unit can be seen as an accumulator that accumulates evidence in favor of the particular response it stands for (Cohen et al. 1990). The relative contribution of the SRN and of the auto-associator to the activation of a response unit is set by a balance parameter $\rho$. In our model, this parameter is constant. The connections between output and response units are thus not subject to learning, and no error term is computed at this level.

The net input received by each response units is computed on each processing cycle according to the following equation:

$$\mathrm{net}_j = \rho\, a_{j(\mathrm{srn})} + (1-\rho)a_{j(\mathrm{aa})} \tag{3}$$

where $\mathrm{net}_j$ corresponds to the net input of unit *j*, $a_{j(\mathrm{srn})}$ and $a_{j(\mathrm{aa})}$ are the activation of the corresponding output unit of the SRN and of the auto-associator respectively. To introduce competition between units, the net input to each unit is then divided by the sum of the net inputs to all four response units (Luce 1963). Equation 4 shows how activation is updated at each cycle, as follows:

$$a_j(t) = a_j(t-1) + [1 - a_j(t-1)]\mathrm{net}_j(t) \tag{4}$$

where $a_j(t)$ represents the activation of unit *j* at cycle *t*. Units thus asymptotically reach their minimum (0.0) or maximum (1.0) activation values. Each activation value is then normalized by applying Luce's choice rule once again.

This updating procedure is applied up until the activation level of the response unit corresponding to the current stimulus reaches a given threshold. To simulate reaction time performance, we considered that the number of cycles required to reach the threshold corresponds to reaction time. This procedure is undoubtedly a simplification, given that it does not make it possible to capture error performance. Another, important simplification is that we assume fixed connections between output and response units. One can indeed imagine that these connections weights, i.e. the respective influence of temporal context and identification processes are subject to learning, as suggested in Usher and McClelland (2001). Our goal, however, was not to account for every aspect of participants' performance in the experiment described in the first part of this chapter, but simply to simulate the effects of the RSI on (1) reaction times in the SRT task and on (2) generation performance.

## 5.4 Reaction time task

To simulate performance during the SRT task, processing occurs in the following way: Elements $t-1$ and $t$ of a sequence are presented simultaneously to the SRN and to the auto-associator network respectively. Activation is propagated concurrently in both the SRN and in the auto-associator networks according to the cascade equations. At some point during the trial, the activation of the output units of both the SRN and of the auto-associator networks start to change, thus reflecting the constraints contained in the input. These changes in turn start influencing the activation levels of the response units. Processing stops when the activation level of the response unit corresponding to the current input reaches a threshold. At that point, back-propagation occurs, independently in both the SRN and in the auto-associator networks. The copy operation then takes place in the SRN network, activations are reset to their resting values, and a new input is presented to the model.

Importantly, we can now simulate the effects of different values of the RSI on performance, in the following manner. When the RSI is equal to 0 ms, processing in both the SRN and in the auto-associator start at the same time: Memory and perception influence action at the same time. When the RSI is greater than 0 ms, processing can start in the SRN before the element *t* of the sequence is presented: Memory can start influencing responses before perception.

## 5.5 Generation task performance

To simulate generation performance, we considered the outputs of the SRN network only. We simply interpreted the output of the SRN as a series of possible responses when generating a sequence of stimuli rather than preparation to the next stimulus in the SRT task (see also Christiansen & Chater 1999). As performance does not depend on reaction time in this task, processing was not cascaded during generation. As for the participants, the generation task begins by the presentation of a randomly chosen stimulus. One of the responses is then selected based on activation levels and presented as the next stimulus to the SRN: the activation of the corresponding input unit is set to one while the activation of the three other input units are set to zero. The same procedure is repeated for every trial in the inclusion and exclusion task.

The response selection procedure varies between inclusion and exclusion instructions. In inclusion, the next input corresponds to the most activated output units at the previous trial, by contrast, in exclusion, this particular response is excluded and the next stimulus is randomly chosen between the three other possible responses.

In the following, we describe a simulation performed to contrast reaction time and generation performance with different values of RSI in the SRT task.

## 5.6 Method and parameters

Twelve different networks in each of three conditions (no RSI, small RSI, high RSI) were each initialized with random weights. To capture the fact that participants enter the experiment knowing how to perform a simple reaction time task, each network was pre-trained on the identification task by exposing it, prior to training on the task itself, to 10,000 elements of a random sequence.

Each sequence element was randomly selected among the four possible stimuli, with the constraint that simple repetitions were forbidden. All networks were then trained on the same stimulus material as human participants were exposed to, and for the same number of trials.

Each trial was divided in 300 processing cycles. The transmission rate ($\tau$) was fixed at 0.05. Response threshold was set at 0.45. The SRN and the auto-associator contributed equally to the activation of response units ($\rho = 0.5$). After any response unit had reached the response threshold, spread of activation was interrupted in the network and error information was back-propagated. After weights had been modified, the activation of response units was then reset to 0.0. The previous stimulus ($t - 1$) was then presented to the SRN. The current stimulus ($t$) was presented to the auto-associator either immediately (no RSI condition) or 50 or 290 cycles later (small RSI and high RSI condition respectively). The values of the other parameters were as follows: slow learning rate = 0.1, momentum = 0.9, fast learning rate = 0.45, fast weight decay = 0.5 (see Cleeremans & McClelland 1991). In order to introduce variability – inherent in human performance – into the model, normally distributed random noise ($\sigma = 0.2$) was added to the net input of each receiving unit (except for the response units). Generation performance was assessed as described above, based on the trained networks resulting from exposure to the SRT task.

## 5.7 Results

Figure 6 (left panel) shows simulated reaction times for the three conditions. The figure makes it clear that changes in the RSI influences the model's response times in the same way as for human participants. Response times are smaller when the RSI increases. Response times increase when the sequence is switched to a different one during the 13th block of trials. The increase in response time appears to be less important for the model than for participants in the high RSI condition.

To further explore network's performance, we plotted the mean square error (i.e., the difference between the target and the actual output) committed by the SRN for the 15 training blocks in all three conditions. As figure 6 (right panel) illustrates, the mean square error tends to decrease from the sixth practice block in the small RSI and high RSI conditions, but remains relatively stable in the no RSI condition. The error rate increases dramatically in all three conditions during the transfer block, but the increase is roughly twice as important in the small RSI and in the high RSI conditions than in the no RSI

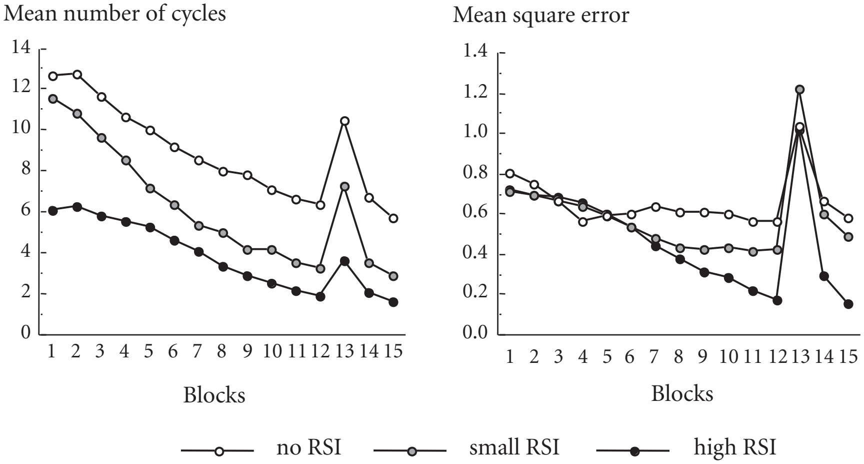


**Figure 6.** Mean simulated reaction times (i.e., the number of processing cycles needed to reach the response-threshold) in the three conditions.

condition. The error also decreases more with practice in the high RSI than in the small RSI condition.

As a whole, these observations suggest that the SRN develops better representations of the sequence when presentation of the next stimulus to the auto-associator network is delayed. Indeed, in the small RSI and high RSI conditions, more activation can reach the SRN's output units than in the no RSI condition. As a result, the activation of the SRN's output units is allowed to be closer to the activation it would have reached in the absence of the response selection procedure. Sequence learning is thus improved with an RSI, for two different reasons. First, the SRN can influence the activation of response units even before the current stimulus is presented, thus resulting in faster response times as long as the correct prediction has been made by the SRN. Second, and more importantly, the SRN is given more of a chance to develop strong, high-quality representations of the sequence with an RSI because error back-propagation takes places on better (i.e., more asymptotic) representations. Further, the more the SRN develops stronger representations of the sequence, the more its influence on the response selection process tends to become more important.

These differences in training regimen also influence generation performance. Figure 7 shows mean simulated generation scores for the three conditions. The figure shows that the model can offer a good qualitative account

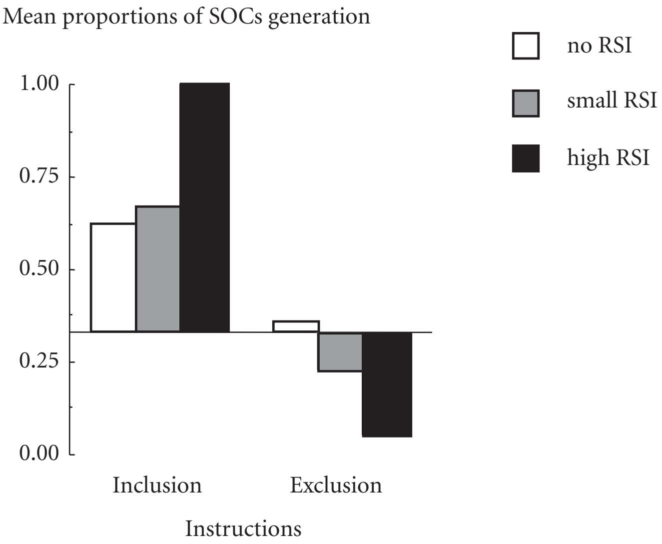


**Figure 7.** Mean simulated inclusion and exclusion scores in the three conditions.

of participants' behavior in the generation task. Indeed, inclusion scores tend to be higher than exclusion scores and, more importantly, changes in the RSI tend to exert opposite effects on inclusion and exclusion scores: increasing the RSI results in higher inclusion scores and in lower exclusion scores.

These results suggest that the stronger representations developed by the network in the RSI conditions during training on the SRT task improve both inclusion and exclusion performance, given that the SRN produced respectively more and fewer training triplets in these condition than in the no RSI condition. The simulation is far from perfect, however. Exclusion scores are lower in the high RSI than in the small RSI condition, while this was not the case for human participants. Overall however, the model seems to behave in the same way as human participants, even though it appears to be more sensitive to RSI manipulations. Over all three conditions, the model accounts for about 75 % of the variance in the SRT task and for about 90 % in the generation task, using a linear fit.

While these results are encouraging, we have to stress that our simulation work contains several simplifications. First, the model does not account for errors in the SRT task. This is not so much of a concern in sequence learning experiments where the number of incorrect responses is very low, but it is clearly inadequate as a general model of reaction time (see Ratcliff, Van Zandt

& McKoon 1999). Second, we used fixed one-to-one connections between the output units of both the SRN and the auto-associator, and the response units. This procedure ensures that the SRN always influences the decision process. We observed, in pilot simulations, that when using trainable connections between output and response units, the temporal context fails to influence the responses of the network. This makes sense, for in a way perception always transmits better information than memory does. Solving this problem probably requires entirely different learning procedures, such as for instance temporal differences learning (Sutton 1988). A third simplification concerns the procedure we used to account for generation performance. Only the SRN is used to simulate this task, given that performance depends exclusively on the temporal context. One may object that we used different strategies to simulate SRT and generation tasks. The particular procedure we used, however, is not made mandatory by the model's architecture, and we are exploring alternatives in which the perception component of the model is also involved during generation.

Overall, even though the model suffers from the limitations listed above, our simulation results are suggestive that subsequent versions of the model that address these limitations might constitute a significant advance in understanding the mechanisms of sequence learning. As it stands, the model is able to account for the difference between direct and indirect measures of sequence learning, and to capture some aspects of the time course of information processing. This makes it possible for the model to account for the temporal effects we have observed, and possibly for other effects that remain to be simulated, such as the effects of the temporal organization of the sequence (Stadler 1995) or, in double task settings, for the effects of the timing between the main and the secondary task (Hsiao & Reber 2001).

## 6. Concluding remarks

The results of the experiment we described in this chapter undoubtedly show that manipulating the RSI influences the extent to which sequence learning is explicit. Beyond the fascinating fact that such a small change in the pacing of the stimulus material can result in large changes in the acquired knowledge, our results also suggest that the differences between implicit and explicit learning might in fact be best viewed as resulting from continuous, gradual changes in a single dimension involving "quality of representation". According to this perspective, representations that are more stable, stronger, and more distinctive are more available to conscious control. Developing such high-quality representations requires time, however, over the course of a single trial, or over development and learning (see also Cleeremans & Jiménez 2002 for a principled presentation of these ideas).

Our simulation data are congruent with these ideas, for the model we introduced in this chapter does not include separate "implicit or "explicit" components. Rather, the extent to which the knowledge acquired by the model over training on the SRT task becomes available to control (as expressed in the generation task) depends on the dynamics of the interactions between memory, perception, and action made possible by different values of the RSI. As such, our model contrasts with other simulation works where implicit and explicit processes are associated with different processing modules (e.g., Wallach & Lebiere, this volume; Sun, Merrill, & Peterson 2001).

Independently of modeling studies, many theories of human learning and memory claim that distinct neural and cognitive systems subtend conscious and unconscious processing (e.g., Willingham 1998). In sequence learning, most of these proposals stems from neuropsychological studies of amnesic (Reber & Squire 1998) or Parkinson patients (Jackson, Jackson, Harrison, Henderson, & Kennard 1995). The results of many brain imaging studies have also led authors to conclude that conscious and unconscious cognition involve different brain areas (Grafton, Hazeltine, & Ivry 1995; Rauch et al. 1995). Most of these studies indeed report that explicit and implicit sequence learning processes involve very different and almost non-overlapping brain networks.

We would like to stress however that neuropsychological and neural imaging approaches do not obviate the need for sensitive behavioral methods, because one has still to carefully assess the conscious versus unconscious character of the knowledge expressed in any task. In the PET scan experiment conducted by Rauch et al., for instance, implicit sequence learning processes were exclusively associated with a phase of the SRT task performed under incidental instructions while the brain regions in charge of explicit processes were identified by the hemodynamic response during a second phase of learning, after participants had been told that a sequence was present. As the authors themselves acknowledge however, this procedure does not ensure that explicit and implicit components of learning were effectively dissociated during the two training phases. It is therefore possible that explicit processes have contaminated the first phase of training, and that implicit learning continued during the second phase of the SRT task.

We believe that the adapting the process dissociation procedure to sequence learning constitutes a possible solution to this "contamination problem" in brain imaging studies. While our model suggests that implicit and explicit learning processes may depend on the same neural substrate, it remains possible that some brain regions are specifically involved in conscious processing (Clegg et al. 1998). Based on these ideas, we recently performed a PET study using the process dissociation design described in this chapter. The results indicated that anterior cingulate and medial frontal cortex are specifically involved in supporting conscious control and recall of newly acquired sequence knowledge in a generation task (Destrebecqz et al. 2000). Further studies will attempt to identify the brain regions supporting implicit knowledge. An open issue is whether these regions will involve neural networks distinct from those involved in conscious processing.

## Notes

* Arnaud Destrebecqz is a postdoctoral researcher of the National Fund for Scientific Research (Belgium). Axel Cleeremans is a research associate of the National Fund for Scientific Research (Belgium). This work was supported by a grant from the Université Libre de Bruxelles in support of IUAP program P4/19 and by grant HPRN-CT-2000-00065 from the European Commission. We are grateful to Géraldine Vanholsbeek for conducting the experiment reported in this chapter.

## References

Buchner, A., Steffens, M. C., Erdfelder, E., & Rothkegel, R. (1997). A multiple model to assess fluency and recollection in a sequence learning task. *The Quarterly Journal of Experimental Psychology, 50A*, 631–663.

Buchner, A., Steffens, M. C., & Rothkegel, R. (1998). On the role of fragmentary knowledge in a sequence learning task. *The Quarterly Journal of Experimental Psychology, 51A*, 251–281.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*, 157–205.

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

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

Cleeremans, A. & Jiménez, L. (1998). Implicit sequence learning: The truth is in the details. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learning* (pp. 323–364). Thousand Oaks: Sage Publications.

Cleeremans, A. & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French and A. Cleeremans (Eds.), *Implicit Learning and Consciousness: An Empirical, Computational and Philosophical Consensus in the Making?* (pp. 1–40). Hove, UK: Psychology Press.

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

Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in Cognitive Science, 2*, 275–281.

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

Cohen, A., Ivry, R. I., & 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.

Curran, T. (1997). Effects of aging on implicit sequence learning: Accounting for sequence strucure and explicit knowledge. *Psychological Research, 60*, 24–41.

Destrebecqz, A. & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review, 8(2)*, 343–350.

Destrebecqz, A., Peigneux, P., Maquet, P., Degueldre, C., Luxen, A., Van der Linden, M., & Cleeremans, A. (2000). Neural correlates of consciousness of sequence generation using the process dissociation procedure. Paper presented at the 6th Annual Meeting of the Organization for Human Brain Mapping. Abstract published in *Neuroimage, 11(5)*.

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

Dienes, Z. & Perner, J. (1996). Implicit knowledge in people and connectionist networks. In G. Underwood (Ed.), *Implicit Cognition* (pp. 227–256). Oxford: Oxford University Press.

Dominey, P. F. (2002). Structure and fonction in sequence learning, *This volume*.

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

Frensch, P. A. & Miner, C. S. (1994). Effects of presentation rate and individual difference in short-term memory. *Memory and Cognition, 22*, 95–110.

Goschke, T. (1997). Implicit learning and unconscious knowledge: Mental representation, computational mechanism, and brain structures. In K. Lambert & D. Shanks (Eds.), *Knowledge, Concepts and Categories* (Vol. 8, pp. 247–333). Hove, UK: Psychology Press.

Goschke, T. (1998). Implicit learning of perceptual and motor sequences. In P. A. French and M. A. Stadler (Eds.), *Handbook of Implicit Learning* (pp. 401–444). Thousand Oaks: SAGE.

Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional Mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience, 7*, 497–510.

Hsiao, A. T. & Reber, A. S. (2001). The dual-task SRT procedure: Fine-tuning the timing. *Psychonomic Bulletin & Review, 8*, 336–342.

Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard, C. (1995). Serial reaction time learning and parkinson's disease: Evidence for a procedural learning deficit. *Neuropsychologia, 33*, 577–593.

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

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measure of sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 948–969.

Luce, R. D. (1963). Detection and recognition. In R.D. Luce, R. R. Busch, and E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 1, pp. 103–189). New York: Wiley.

McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems in cascade. *Psychological Review, 86*, 287–330.

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

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.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.

Rauch, S. L., Savage, C. R., Brown, H. D., Curran, T., Alpert, N. M., Kendrick, A., Fischman, A. J., & Kosslyn, S. M. (1995). A PET Investigation of implicit and explicit sequence learning. *Human Brain Mapping, 3*, 271–286.

Reber, P. J. & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience, 10*, 248–263.

Reed, J. & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 585–594.

Reingold, E. M. & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 44*, 563–575.

Rumelhart, D. E. Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing* (pp. 318–362). Cambridge: MIT Press.

Shanks, D. R. & Johnstone, T. (1998). Implicit knowledge in sequential learning tasks. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of Implicit Learning* (pp. 533–572). Thousand Oaks: Sage Publications.

Shanks, D. R. & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a serial reaction time task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

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

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.

Sutton, R. (1988). Learning to predict by the method of temporal differences. *Machine Learning, 3*, 9–44.

Usher, M. & McClelland, J. L. (2001). On the Time Course of Perceptual choice: The leaky competing accumulator model. *Psychological Review, 108*, 550–592.

Wallach, D. & Lebiere, C. (2002). Implicit and explicit learning in a unified architecture of cognition, *This volume*.

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

Willingham, D. B., Greenberg, A. R., & Cannon Thomas, R. (1997). Response-to-stimulus interval does not affect implicit motor sequence learning, but does affect performance. *Memory* & *Cognition, 25*, 534–542.

# Implicit and explicit learning in a unified architecture of cognition

Dieter Wallach[*] and Christian Lebiere
University of Applied Sciences, Kaiserlautern /
Carnegie Mellon University

## 1. Introduction

The last thirty years saw a number of different experimental paradigms investigating the concept of implicit learning in domains as diverse as learning of artificial grammars (Reber 1967), sequence learning (Willingham, Nissen, & Bullemer 1989), anagram learning (Gardiner, Alison, & Sutton 1989), acquisition of invariant features (McGeorge & Burton 1990), probability learning (Reber & Millward 1968), perceptual learning (Kolers & Roediger 1984), or learning to control complex systems (Berry & Broadbent 1984). While literally dozens of different definitions of *implicit learning* and its relation to *explicit learning* have been proposed in the literature (Frensch 1998), they all focus on the conjecture that people seem to learn more about the structural properties of a stimulus environment than they are able to convey, resulting in empirically demonstrated dissociations between observable task performance and verbalizable knowledge.

Definitions of implicit learning often go beyond a mere descriptive formulation of the phenomenon, and reflect particular theoretical preconceptions of the respective researchers. In this sense, many authors equate implicit learning with *unconscious* learning and thus refer to *consciousness* as a central construct in their conceptualizations of implicit learning. While recent interest in implicit learning can be traced back to Reber's classic study on artificial grammar learning (Reber 1967), the topic of learning without conscious awareness attracted the attention of researchers very early in the history of Psychology (Hull 1920; Jenkins 1933; Thorndike & Rock 1934). As Brandstädter (1991) notes, disputing the role of consciousness has been a central topic in psychology for over a century (James, 1892; Wundt, 1896) and demonstrations of learning without phenomenal awareness have unceasingly fascinated researchers ever since. It is therefore not surprising to find striking similarities in Thorndike and Rock's (1934: 1) discussion of learning "without awareness of what is being learned or intent to learn it" or Jenkins' (1933: 471) "learning that occurs in the absence of specific intent to learn it" and the definition of implicit learning as "...designating cases where some knowledge is (1) acquired without intention to learn ... (2) capable of influencing behavior unconsciously" (Cleeremans & Jiménez 1998).

While some authors interpret the available data as striking evidence for "the sophistication, ubiquity, and most of all, entirely nonconscious nature of implicit learning processes" (Lewicki, Czyzewska, & Hill 1997), others dispute that implicit learning is an unconscious process. Seemingly convincing demonstrations of dissociations and assessments of implicit learning, they argue, are fundamentally flawed in that they are subject to methodological problems concerning the sensitivity of applied assessment instruments or misconceptions about the assumed necessity of task knowledge (Shanks & St. John 1994; St. John & Shanks 1997). Consequently, some researchers prefer not to make use of the term *consciousness* in their definitions of implicit learning (Neal & Hesketh 1997). Instead, they propose to separate the question of implicit learning from issues of conscious access to acquired knowledge (Mathews & Roussel 1997). Following this position, research should focus on questions of different properties of implicit and explicit learning processes and the encoding, representation and retrieval of their products (see Berry 1997).

On a coarse level, at least three theoretical approaches that aim at explaining implicit learning by differentiating mental processes and their resulting structures can be distinguished. Researchers like Reber (1989, see Berry & Broadbent 1988; Manza & Reber 1997) define implicit learning as a "situation-neutral induction process" (Reber 1993: 12) that results in the acquisition of "abstract knowledge" about the structure of an environment. In line with this *rule-induction hypothesis,* Lewicki et al. (1989) propose that implicit learning leads to abstract representations that retain a domain's underlying structural characteristics while abstracting from specific surface information. According to the rule-induction hypothesis, the assumed abstractive nature of implicit learning is hypothesized to be grounded in the (unconscious) acquisition of rules that capture covariation patterns of physical stimuli, rather than recording details of a single episode. By contrast, researchers following a position that can be described as the *episodic chunks hypothesis* explain implicit learning as the encoding and retrieval of representations of literal instances of stimuli and their successive order (St. John & Shanks 1997) without assuming an underlying inductive abstraction process. In a variant on this position, the *distributed fragments hypothesis,* Dulany (1996: 523) equates implicit learning with the acquisition of "…evocative mental episodes. It consists of the establishment and use of evocative relations among non-propositional but fully conscious contents", arguing for the distributed character of the representational basis of implicit learning.

The positions briefly sketched above all refer to high-level mental processes and structures to explain the phenomenon of implicit learning. However, the respective positions are not empirically distinguishable without additional assumptions about properties of the human cognitive system. A central lesson that was learned from the imagery debate (Anderson 1978) is that phenomena that are supposedly explained by the assumption of certain representational formats can also be accounted for by different representational formats when making specific assumptions about their processing. To tackle this problem, we define implicit and explicit learning in this paper in terms of the architectural mechanisms and structures of a unified theory of cognition. The computational approach advocated ensures the use of precisely defined theoretical concepts, and provides a comprehensive framework for exploring the interplay of representational assumptions and proposed cognitive processes to account for phenomena found in research on implicit learning.

While various approaches to formulate computational models for several research fields of implicit learning have been proposed (Cleeremans 1993; Dienes & Fahey 1995; Sun 1999; Mathews & Roussel 1997), a major shortcoming of these models is their failure to account for explicit learning and for the difference between implicit and explicit learning (Stadler & Roediger 1998). In this paper, we attribute implicit and explicit learning to distinct architectural mechanisms of an integrative theory of cognition. We explore the scope of the proposed approach across two subfields of research on implicit learning. In two models of complex process control the role of implicit and explicit learning processes in accounting for results gathered with two well-known tasks is discussed. Then we present the application of a general sequence learning model (Lebiere & Wallach 2000) to data reported in a recent study by Destrebecqz and Cleeremans (2001).

In the following section we discuss the relevance of unified theories of cognition as theoretical frameworks for the explanation of implicit and explicit learning. We then give a brief overview of the ACT-R architecture (Anderson & Lebiere 1998) that provides the theoretical foundation for the approach discussed in this chapter and will introduce its central concepts.

## 2. On the role of unified theories of cognition

Computational modeling based on a unified theory of cognition provides a promising starting point towards a theory of implicit and explicit learning for the following reasons:

- A unified theory of cognition provides a comprehensive theoretical framework for the interpretation and explanation of empirical findings. Unified theories are *complete* theories in the sense that they encompass assumptions about the encoding, representation and processing of stimuli. Such a framework also promises to explain empirical results observed in different research fields of implicit learning by tracing them back to a common underlying set of architectural mechanisms and structures.
- Computational models provide precise quantitative predictions of empirical phenomena on the basis of a formally specified set of underlying constructs and mechanisms. While the precision of computational models in their predictions and definitions of proposed theoretical concepts is considered to be a general advantage of formal theorizing, it is of particular importance in a field that can adequately be characterized by the proliferation of different meanings of its central concepts (see also Frensch 1998). By accurately clarifying explanatory theoretical concepts, an integrative computational theory not only allows the derivation of detailed quantitative predictions for conducting new experiments, but also significantly facilitates communication among scientists in the field. In this sense, computational theories can be used as «hypothesis generators» to precisely formulate empirical predictions on the grounds of well-specified sets of theoretical constructs.
- As Perruchet and Gallego (1997: 135) note, the literature is "full of examples in which initial claims for unconscious learning have subsequently been discounted, because it later became apparent that performance was grounded on knowledge that was not considered in the original studies" (for a classical example see Dulany 1961 in his reconsideration of data on verbal operant conditioning). Consequently, Perruchet and Gallego (1997) propose to carefully analyze the type of structural information embedded in the task under study, as well as the representation prerequisites for suc-

> cessfully mastering it. By analytically encoding knowledge in the representational structures of an integrative computational model, this approach can be used to explore the impact of available knowledge on task performance. In the Transportation model, discussed later in this chapter, we demonstrate that a model is fully capable of accounting for empirically observed performance without including knowledge that was regarded as a necessary precondition for successful task performance.

Before presenting examples of learning models formulated in the Act-R framework, the next section first introduces the central concepts of the Act-R cognitive architecture (Anderson & Lebiere 1998).

## 3. Act-R: An integrative cognitive architecture

Act-R is a hybrid production system that distinguishes between a permanent procedural memory and a permanent declarative memory. Procedural knowledge is encoded in modular condition-action rules (*productions*) that represent potential actions to be taken when certain conditions are met. Declarative structures called *chunks* are used to store factual knowledge in declarative memory. Chunks encode knowledge as structured, schema-like configurations of labeled slots that can be organized hierarchically. A representation of goals 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 which chunks are retrieved from memory. These quantities reflect past statistics of use of the respective symbolic knowledge structures and are learned by Bayesian learning mechanisms derived from a *rational analysis* of cognition (Anderson 1990). Subsymbolic learning allows Act-R to adapt to the statistical structure of an environment.

Subsymbolic activation processes make a chunk active to the degree that past experience and the present context (as given by the current goal) indicate that it is useful at this particular moment. Retrieving a chunk results in its immediate reinforcement through Act-R's base-level activation learning mechanism to reflect its frequency of use. Formally, activation reflects the log posterior odds that a chunk is relevant in a particular situation. The activation $A_i$ of a chunk $i$ is computed as the sum of its *base-level activation* $B_i$ plus its *context* *activation:*

$$A_i = B_i + \sum_j W_j S_{ji} \qquad \textbf{Activation equation}$$

In determining the context activation, $W_j$ designates the attentional weight or source activation given the context element *j*. An element *j* is in context if it is part of the current goal chunk (i.e. the value of one of the goal slots). $S_{ji}$ stands for the strength of association from element *j* to a chunk *i*. Act-R assumes that there is a limited capacity of source activation and that each goal element has an equal amount. Source activation capacity is typically assumed to be 1, i.e. if there are *n* source elements in the current goal each receives a source activation of 1/n (Anderson, Reder, & Lebiere 1996). The *associative strength* $S_{ji}$ between an activation source *j* and a chunk *i* is a measure of how often chunk *i* was needed (i.e. retrieved in a production) when source *j* was in the context. Associative strengths provide an estimate of the log likelihood ratio measure of how much the presence of a cue *j* in a goal slot increases the probability that a particular chunk *i* is needed for retrieval to instantiate a production.

The base level activation of a chunk is learned by an architectural mechanism to reflect the past history of use of a chunk *i*:

$$B_i = \ln \sum_{j=1}^{n} t_j^{-d} \approx \ln \frac{nL^{-d}}{1-d} \qquad \textbf{Base-level learning equation}$$

In the above formula $t_j$ stands for the time elapsed since the *jth* reference to chunk *i* while *d* is the memory decay rate and *L* denotes the life time of a chunk (i.e. the time since its creation). As Anderson and Schooler (1991) have shown, this equation produces the Power Law of Forgetting (Rubin & Wenzel 1996) as well as the Power Law of Learning (Newell & Rosenbloom 1981). Strengths of associations are learned through a similar mechanism that records the statistics of co-occurrence between sources and chunks retrieved (for details, see Anderson & Lebiere 1998).

When retrieving a chunk to instantiate a production, Act-R selects the chunk with the highest activation $A_i$. However, stochasticity is introduced in the system by adding gaussian noise of mean 0 and standard deviation $\sigma$ to the activation $A_i$ of each chunk. In order to be retrieved, the activation of a chunk needs to reach a fixed retrieval threshold $\tau$ that limits the accessibility of declarative elements. If the gaussian noise is approximated with a sigmoid distribution, the probability $P$ of chunk $i$ to be retrieved by a production is:

$$P = \frac{1}{1 + e^{-\frac{A_i - \tau}{s}}} \qquad \textbf{Retrieval probability equation}$$

where $s = \sqrt{3}\sigma/\pi$. The activation of a chunk $i$ is directly related to the latency of its retrieval by a production $p$. Formally, retrieval time $T_{ip}$ is an exponentially decreasing function of the chunk's activation $A_i$:

$$T_{ip} = Fe^{-A_i} \qquad \textbf{Retrieval Time Equation}$$

where $F$ is a time scaling factor. In addition to the latencies for chunk retrieval as given by the *Retrieval Time Equation*, the total time of selecting and applying a production is determined by executing the actions of a production's action part, with a value of 50 ms typically assumed for elementary internal actions. External actions, such as pressing a key, usually have a longer latency determined by the ACT-R/PM perceptual-motor module (Byrne & Anderson 1998).

While the declarative chunks are called symbolic, their slot values often belong to continuous domains, such as numbers or quantities, or more generally display a similarity structure. Instead of only retrieving chunks that perfectly match the production conditions, ACT-R's *partial-matching* mechanism can retrieve any chunk to the degree that it matches the condition. Specifically, the chunk with the highest match score is retrieved, where the match score $M_{ip}$ is a function of the activation of chunk $i$ in production $p$ and its degree of mismatch to the desired values:

$$M_{ip} = A_i - MP \sum_{v,d} \left(1 - Sim(v, d)\right) \qquad \textbf{Partial matching equation}$$

In the above formula $MP$ is a mismatch penalty constant, while $Sim(v, d)$ is the similarity between the desired value $v$ held in the goal and the actual value $d$ held in the retrieved chunk, allowing the representation of continuous quantities. Thus even if no chunk in memory perfectly matches the current context, a likely occurrence with an infinite amount of continuous values, the chunk holding the closest value can be retrieved if its match score after subtracting the mismatch between values from its activation is still higher than the retrieval threshold (and the match scores of competing chunks).

A shortcoming of partial matching is that while it generalizes the matching process to handle continuous quantities, it can only return a value already present in some chunk. Lebiere (1999) proposed a generalization of the retrieval mechanism called *blending* which allows the retrieval and averaging of values from multiple chunks rather than a single one, providing for sequences of continuous values. This is a powerful kind of interpolation that has proved useful for a range of paradigms of implicit learning (Wallach & Lebiere 1999; Gonzalez, Lebiere, & Lerch 1999). Specifically, the value $V$ retrieved is:

$$V = Min \sum_i P_i \, (1 - Sim(V, V_i))^2 \qquad \textbf{Blending equation}$$

where $P_i$ is the probability of retrieving chunk $i$ and $V_i$ is the value held by that chunk.

After having presented the central concepts of the Act-R architecture in this section, we will now turn to discuss its application to modeling implicit learning in the domains of implicit process control (Section 4) and sequence learning (Section 5).

## 4. Implicit learning in process control

Before concluding that subjects are unaware of the information that they have learned and that is influencing their behavior, it must be possible to establish that the information the experimenter is looking for in the awareness test is indeed the information responsible for performance changes (Shanks & St. John 1994: 373).

The concept of implicit learning in process control was first explored by Berry and Broadbent (1984) who reported negative correlations between task performance and the ability to answer specific questions about a system's behavior (see also Broadbent 1977). In this section we discuss an explanation for the reported dissociation between knowledge and performance based on the acquisition and retrieval of instances. As Neal and Hesketh (1997) have argued, it seems clear after 20 years of controversy that implicit learning tasks can be performed by either relying on prior instances or by structural knowledge of the task – "however, researchers still disagree over the representation of these types of knowledge and whether either or both can exert an implicit influence on task performance" (Neal & Hesketh: 35). To explore the role of instances in more detail, we present an Act-R model of instance-based learning and compare it to a well-known model proposed by Dienes and Fahey (1995, 1998).

## 4.1 SugarFactory

In an influential study, Berry and Broadbent (1984) used the SugarFactory task, a dynamic system that has subsequently been applied in a number of studies to investigate complex problem-solving processes (Berry 1991; McGeorge & Burton 1989; Squire & Frambach 1990). SugarFactory is a computer-simulated task in which participants are told to imagine that they are factory managers and can control the production of sugar *sp* by determining the number of workers *w* employed on each of a number of trials. Unbeknown to the participants, the behavior of SugarFactory is governed by the following equation:

$$sp_t = 2 \bullet w_t - sp_{t-1} \qquad (1)$$

Basically, sugar production is proportional to the number of workers employed, which is intuitive enough, but inversely related to the sugar production at the previous step, a difficult and counterintuitive relation to infer. The value entered for the workers hired ($w_t$) can be varied in 12 discrete steps $1 \leq w_t \leq 12$, while the sugar production $sp_t$ changes discretely between $1 \leq sp_t \leq 12$. To allow for a more realistic interpretation of *w* as the number of workers and *sp* as tons of sugar, these values are multiplied in the actual computer simulation by 100 and 1000, respectively. If the result according to the equation is less than 1000, *sp* is simply set to 1000. Similarly, a result greater than 12000 always leads to an output of 12000 tons of sugar. Finally, a random component of ±1000 is added in 2/3 of all trials to the result derived from equation (1).

### 4.1.1 SUGARFACTORY *automaton*

Buchner, Funke and Berry (1995) proposed to think of this control task in terms of a *finite state automaton*. A finite state automaton (Partee, ter Meulen & Wall 1990: 458) can generally be described as a quintuplet $< K, \Sigma, \sigma, q_0, F >$ with:

- $K$ is a finite set of states
- $\Sigma$ is a finite set, referred to as the alphabet
- $q_0 \in K$, the initial state
- $K \supseteq F$, the set of final states

$\sigma$ is a finite function from $Kx\Sigma$ into $K$, the transition function. The following alternative formulation of equation (1) exemplifies an application of the definition above to describe a finite state "SugarFactory automaton". This

**Table 1.** Transition matrix of the SugarFactory automaton

| | | Input ($w_t$) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | Output ($sp_t$) | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 | 1100 | 1200 |
| 1 | 1000 | 1 | 3 | 5 | 7 | 9 | 11 | 12 | 12 | 12 | 12 | 12 | 12 |
| 2 | 2000 | 1 | 2 | 4 | 6 | 8 | 10 | 12 | 12 | 12 | 12 | 12 | 12 |
| 3 | 3000 | 1 | 1 | 3 | 5 | 7 | 9 | 11 | 12 | 12 | 12 | 12 | 12 |
| 4 | 4000 | 1 | 1 | 2 | 4 | 6 | 8 | 10 | 12 | 12 | 12 | 12 | 12 |
| 5 | 5000 | 1 | 1 | 1 | 3 | 5 | 7 | 9 | 11 | 12 | 12 | 12 | 12 |
| 6 | 6000 | 1 | 1 | 1 | 2 | 4 | 6 | 8 | 10 | 12 | 12 | 12 | 12 |
| 7 | 7000 | 1 | 1 | 1 | 1 | 3 | 5 | 7 | 9 | 11 | 12 | 12 | 12 |
| 8 | 8000 | 1 | 1 | 1 | 1 | 2 | 4 | 6 | 8 | 10 | 12 | 12 | 12 |
| 9 | 9000 | 1 | 1 | 1 | 1 | 1 | 3 | 5 | 7 | 9 | 11 | 12 | 12 |
| 10 | 10000 | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 6 | 8 | 10 | 12 | 12 |
| 11 | 11000 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 5 | 7 | 9 | 11 | 12 |
| 12 | 12000 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 4 | 6 | 8 | 10 | 12 |

formulation will allow for a straightforward presentation of an Act-R model that sheds light on the reported knowledge-performance dissociation. Table 1 shows the transition matrix of the SugarFactory automaton.

At any point in time the SugarFactory automaton is in one of a finite set of states $K$. In Table 1 the first column shows this set $K$ of states that the automaton could be in, numbered from #1 to #12. Each of these states is at the same time considered to be an end state, i.e. a transition from state $x$ to state $y$ finishes the automaton's computation for that trial. The second column indicates the output signals that the automaton emits when being in the respective state. Computation of the SugarFactory automaton starts in the initial state $q_0 \in K$, where the automaton is in state #6 and emits an output of 6 tons of sugar. The input values for the number of workers form the alphabet $\Sigma$ of the automaton, the transition function $\sigma$ is given in extensional form as a transition matrix $K \times \Sigma$ into $K$. The transition matrix in Table 1 displays the (12 x 12) cells that result from the combination of all possible states (with their associated output signals $sp_t$,) and input values ($w_t$, input signals). To illustrate the processing of this automaton consider a situation in which the system is in state #5, emitting an output signal $sp_t$ of 5000. If the input signal $w_t$ in this state is set to 8 (i.e. 800 workers are employed), the automaton changes to state #11 and outputs a sugar production $sp_t$ of 11000 tons. The target state in the experiment, reaching an output of 9000 tons, is marked as state #9. To account for the random component mentioned, a value of ±1000 is added to the output signal in 2/3 of all cases.

By referring to the formulation of SugarFactory as an automaton, Buchner et al. (1995: 169) interpret the reported dissociation between knowledge and performance when controlling the SugarFactory as a *sampling problem*:

> Good controllers will, by definition, reach the target states more frequently than "bad controllers". As a consequence, the "good controllers" will necessarily experience fewer transitions that are different from those leading to the target state. … In contrast, "bad controllers" will not frequently reach the target states. As a consequence, these subjects may experience a larger number of different state transitions. In other words, it is plausible that the "bad controllers" explore a larger selection of the system's transition matrix. Thus the two types of subjects may have different learning experiences.

This argument provides a very simple and straightforward interpretation of the knowledge-performance dissociation: If "bad controllers" experience a broader range of different system states, they should be better in answering questions about the behavior of the system after introducing input signals in given states. "Good controllers", on the other hand, being often in the target state (#9), should experience fewer different system transitions and can thus be assumed to be worse in answering questions about samples of system transitions.

### 4.1.2 *An instance-based learning approach: Dienes and Fahey (1995)*

For an account that refers to the processing of such a "look-up table" (Broadbent, Fitzgerald, & Broadbent 1986) of state-effect contingencies, it is necessary to specify mechanisms for the acquisition and retrieval of knowledge that encodes the effects of actions given certain states of the system. According to the arguments by Buchner et al. (1995), a computational model that learns by acquiring and deploying instances of system transitions should be successful in modeling the performance empirically observed. Based on Logan's instance theory (1988, 1990), Dienes and Fahey (1995) developed a computational model (called the D&F model in the remainder of this section) to account for the data they gathered in an experiment using the SugarFactory task.

According to Logan's instance theory, encoding and retrieval are intimately linked through attention: encoding a stimulus is an unavoidable consequence of attention, and retrieving what is known about a stimulus is also an obligatory consequence of attention. Logan's theory postulates that each encounter of a stimulus is encoded, stored and retrieved using a separate memory trace. These separate memory traces accumulate with experience and lead to a "gradual transition from algorithmic processing to memory-based processing" (Logan 1988: 493). Any model that relies on the retrieval of instances either needs an established base of instances that can be retrieved, or algorithmic, rule-like knowledge to build up representations of instances.

In an experiment with the SugarFactory task, Dienes and Fahey (1995: 862) observed that 86% of the first ten input values that participants entered into SugarFactory can be explained by the following rules:

**Rule 1:**

`For the very first trial, enter a work force of 700, 800 or 900.`

**Rule 2:**

`If the sugar production is below (above) target, then increase (decrease) the amount of workers with 0, 100, or 200.`

**Rule 3:**

`If the sugar production is on target, then respond with a workforce that is different from the previous one by an amount of -100, 0, or +100 with equal probability.`

Consequently, the authors assumed this algorithmic knowledge to be available prior to the representation of instances that could be retrieved to solve a problem. In their model, Dienes and Fahey encoded this rule-like knowledge by a constant number of prior instances that could be retrieved in any situation. The number of prior instances encoded is a free parameter in the D&F model that was fixed to give a good fit to the data reported below.

Logan's instance theory predicts that every encounter of a stimulus is stored. The D&F model, however, deviates from this assumption in that it only stores instances for those situations in which an action successfully leads to the target production of 9000 tons. All other situations are postulated to be ignored by the model – an assumption which not only lacks plausibility, but also violates Logan's instance theory that supposedly forms the theoretical foundation of the D&F model. Complicating the modeling basis further, the D&F model uses a definition of what a successful action is that was not available to participants. Since, due to the random component in the SugarFactory equation, the outcome calculated by the SugarFactory formula may vary by ±1000 tons, the D&F model only stores instances about actions that were successful according to this criterion.

In the D&F model each stored instance "relevant" to a current situation races against others and against prior instances representing the algorithmic knowledge given by rules 1–3. The fastest instance determines the action of the model. An instance encoding a situation is regarded as "relevant" by the D&F model if it either matches the current situation exactly or if it does not differ from it by more than 1000 tons of sugar in either the current output or the desired output, analogous to the "loose" range discussed above.

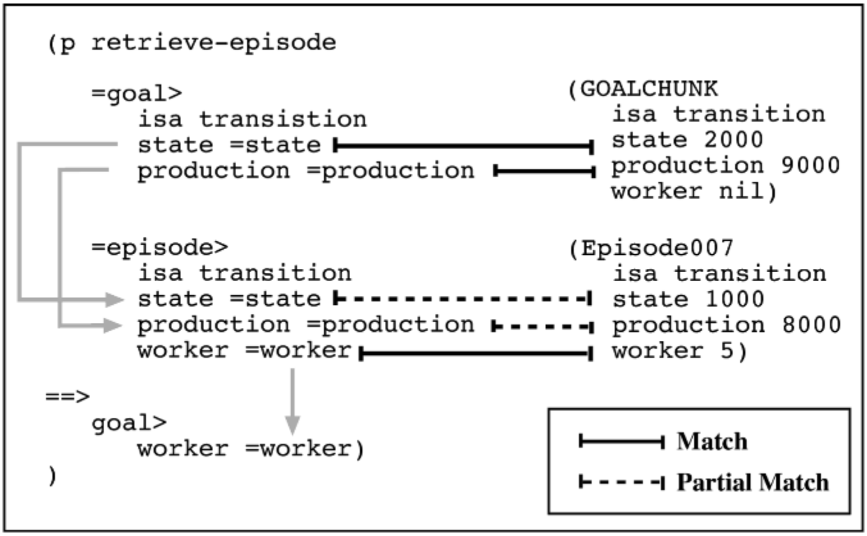
### 4.1.3 *Instance-based learning in* ACT-R

We developed an instance-based ACT-R model for the SUGARFACTORY task. In contrast to the D&F model, we encoded the algorithmic knowledge (rules 1-3; see the previous section) needed for "bootstrapping" an initial base of instances by simple ACT-R production rules. While the D&F model selectively encoded only "successful" situations, the ACT-R model learned a chunk for every situation that the model encountered, irrespective of its result. This is not only perfectly in line with Logan's view on the encoding of instances, but also does not require the assumption that subjects selectively discard episodes that are not successful according to a scoring scheme that they were not aware of. Both of these differences are direct implications of the ACT-R theory: The procedural-declarative distinction specifies the form of algorithmic knowledge as production rules while the theory of chunk creation as a function of goal completion specifies that all problem-solving episodes lead to new declarative chunks. The following chunk is an example of an instance stored by the ACT-R model:

```
Transition123
ISA transition
STATE 3000
WORKER 800
PRODUCTION 12000
```

Chunks of type transition encode the state of the system (`STATE`), a record of the subjects' own action (`WORKER`) as well as the system's response (`PRODUCTION`). The above chunk (the name `TRANSITION123` is arbitrary) represents a situation in which an input of 800 workers, given a current production of 3000 tons, led to subsequent sugar production of 12000 tons.

Retrieval in the ACT-R model is governed by similarity-based matches between the present situation and encoded episodes experienced in the past. On each trial, a memory search is initiated based on the current state and the target state '9000 tons' as cues to retrieve an appropriate intervention. As outlined in the previous section, chunk retrieval in ACT-R is governed by the activation level of memory elements. Generally, productions aim at retrieving instances that exactly match the situation pattern formulated in its condition part. If



```
(p retrieve-episode

   =goal>
      isa transistion
      state =state
      production =production

   =episode>
      isa transition
      state =state
      production =production
      worker =worker

==>
   goal>
      worker =worker)
)
```

```
(GOALCHUNK
   isa transition
   state 2000
   production 9000
   worker nil)

(Episode007
   isa transition
   state 1000
   production 8000
   worker 5)
```

**Figure 1.** Partial matching of chunks

such chunks do not exist in memory, or if the activation level of an exactly matching chunk is too low, the model might retrieve a highly active memory element that only partially matches the condition pattern of the retrieving production. Chunks that do not exactly match the current situation will, however, be penalized by having their activation lowered for each mismatching slot by an amount proportional to the degree of mismatch using the following equation:

$$\textit{Activation Penalty} = MP \sum_{s=\textit{slots in matched chunk}} (1 - \textit{sim}(\textit{required } s, \textit{actual } s))$$

For each chunk pattern of the production that is matched to a slot of a candidate chunk for retrieval, the similarity between their respective contents is calculated. If this similarity is perfect (i.e. *sim*=1), no penalty is subtracted. When the similarity is lower than 1, a corresponding proportion of *MP* is subtracted from the activation to yield the chunk's match score.[1] Thus any chunk can potentially be retrieved by partial matching, not only those that differ by a given amount, as in the D&F model. However, all things being equal, the greater the amount of mismatch, the lower the probability of retrieval.

Figure 1 shows an example of this process of partial matching. In Figure 1 the chunk `GOALCHUNK` represents a situation in which the current sugar output ("2000") is encoded in the slot `STATE`, while the target state ("9000") is encoded in slot `PRODUCTION`. Chunk `EPISODE007` is retrieved by production

`RETRIEVE-EPISODE` using partial matching of the values “1000” vs. “2000” in the state slot and the values “8000” vs. “9000” in the `PRODUCTION` slot. In the action part of production `RETRIEVE-EPISODE` the number of the `WORKER` slot of `EPISODE007` (“5”) is then used to modify the current `GOALCHUNK` by the retrieved value for its `WORKER` slot.

While the ACT-R model and the D&F model share strong similarities, the D&F model makes assumptions regarding the storage and the retrieval of instances that can hardly be justified on either a theoretical or empirical basis. Dienes and Fahey (1995: 865) point out that these critical assumptions are essential to the performance of their model:

> The importance to the modeling of assuming that only correct situations were stored was tested by determining the performance of the model when it stored all instances. (...) This model could not perform the task as well as participants: the irrelevant workforce situations provided too much noise by proscribing responses that were in fact inappropriate (...) If instances entered the race only if they exactly matched the current situation, then for the same level of learning as participants, concordances were significantly greater than those of participants.

The ACT-R model, on the other hand, does not postulate these assumptions , it can be regarded as simpler than the D&F model, demonstrating how instance-based learning can be captured by the basic mechanisms of a unified theory of cognition.

As Figure 2 and Figure 3 demonstrate, both models are equally successful in their empirical predictions. Figure 2 plots the trials on target when participants controlled SUGARFACTORY over two phases, consisting of 40 trials each. While the ACT-R model slightly overpredicts the performance reported by Dienes and Fahey (1995) in the first phase, the D&F model slightly underpredicts the performance of the participants in the second phase.

After the participants controlled the SUGARFACTORY in the experiment of Dienes and Fahey (1995), they were required to answer a questionnaire task. Again they had to determine the work force in 80 situations, but this time they did not receive feedback, but just moved on to a new, unrelated situation. The 80 situations presented were the last 40 situations from the first part of the experiment mixed with 40 new situations, i.e. situations which participants did not encounter while controlling the system. Figure 3 shows the percentage of times *(concordance)* participants chose the same work force in this second task (questionnaire) as they did in the first task (control). The baseline level represents the chance that both choices are equal due to random choice. This

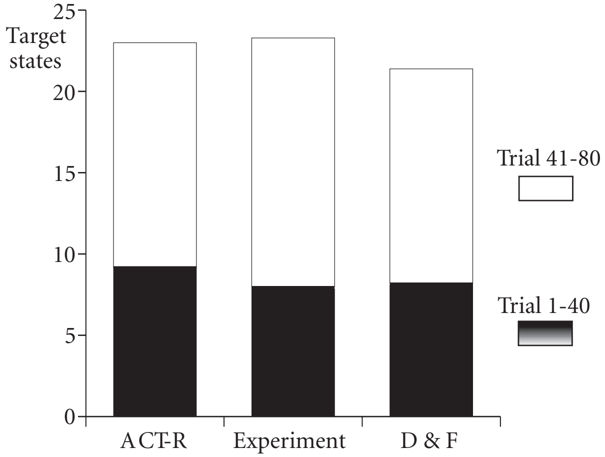


Figure 2. Control performance with the SugarFactory

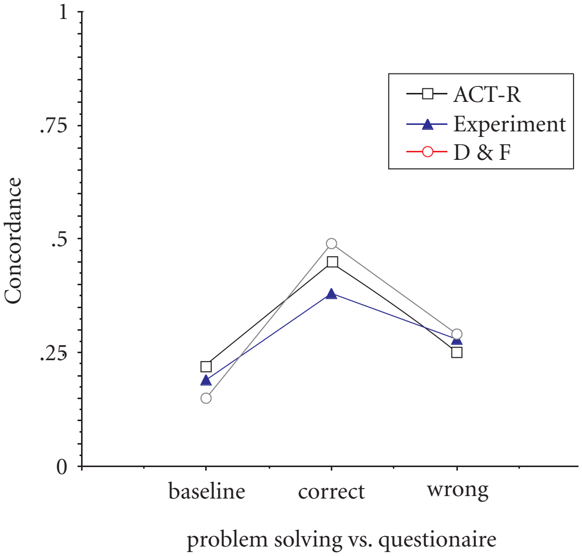


Figure 3. Concordance data

chance is higher than 1/12, because some choices are made more often during the experiment than others. The column labeled “correct“shows how often the same work force was chosen if this lead to a correct output, the “wrong” column shows the same for the incorrect outputs. Choices are close to base level for "wrong" answers, while they are significantly higher for "correct" answers, indicating a better memorization of "correct" answers. While this is a trivial consequence from not storing "wrong" instances in the D&F model, no special mechanisms to arrive at this result are required in the Act-R model. Again, both models seem to do similarly well in modeling the data, on a qualitative and quantitative level.

In this section we have demonstrated that an instance-based learning approach is sufficient to successfully account for the data pattern reported with the SugarFactory task. We also provided an explanation for observed dissociations between knowledge and performance and have shown that a simple model developed in the Act-R cognitive architecture is not only capable of modeling the data with the same precision as the model of Dienes and Fahey (1995), but that it does so without ad hoc assumptions about the storage and retrieval of instances.

The performance of the Act-R model for the SugarFactory task is based on the encoding and retrieval of declarative chunks. Although no abstraction mechanisms as proposed by Reber (1989) is involved, similarity-based chunk retrieval provides some form of implicit generalization of stored instances, inasmuch as a given instance can be applied to new but similar situations. Knowledge about the link between a specific state of the SugarFactory (i.e. the current production level), a given response, and its resulting sugar production output is assumed to be explicitly available as a chunk. In line with Shanks and St. John's (1994) argument of implicit retrieval, the process of retrieving instances based on subsymbolic activation levels is, however, assumed to be beyond conscious control. Instance retrieval is thus dependent on the ability of a certain cue to activate the respective memory chunk. In an experiment with the SugarFactory task, Dienes and Fahey (1998) found that in a recognition task participants are unable to recognize as "old" situations in a recognition task that they have previously experienced in the control phase. According to the argument above, the recognition task presents cues that fail to reinstate the encoding context and thus to activate the respective knowledge (see also Dienes & Fahey, 1998: 609).

In this section it was shown that activation-based instance retrieval in Act-R is capable of generalizing to situations not previously experienced, providing a variant of an abstraction process without the acquisition of rules. In the next section this approach is generalized in a model of Broadbent's Transportation task.

## 4.2 Implicit learning in process control: TRANSPORTATION

Broadbent's TRANSPORTATION task (Broadbent & Aston 1978; Broadbent 1977) has been used in a large number of studies to investigate implicit learning processes in the control of dynamic systems. In the TRANSPORTATION task participants can vary two input variables: [$t$] (the time interval between buses entering the city) and [$f$] (the fee charged for the use of the city's parking lots). Altering these quantities affects two output variables: [$L$] (load on the buses) and [$S$] (number of empty spaces remaining in the parking lots). Unbeknownst to the subjects the behavior of TRANSPORTATION is governed by the following equations:

1. $L = 220t + 80f$
2. $S = 4.5f - 2t$

In a typical experiment using the TRANSPORTATION task, subjects are asked to manipulate the input variables [$t$, $f$] to produce given value pairs [$L$, $S$] of the output variables. Previous research by Broadbent and his colleagues found no correlation between ability to control the system (as judged by the number of attempts to reach specific target values) and scores on a post-task questionnaire. While most subjects in Broadbent's experiments discovered the direct (or salient) influence of $t$ on $L$ and $f$ on $S$, they were not able to verbalize knowledge of the non-salient (i.e. weaker) influence of $f$ on $L$ and $t$ on $S$. Berry and Broadbent (1987: 9), however, assume:

> If, however, the equations are such that there is a unique pair of input values for each output pair subjects must take the crossed relationships, as well as the direct ones, into account when controlling the system. This is a feature of the task that ensures that successful performance cannot be based on the salient relationships alone.

As Berry (1993: 20) reports, questionnaire scores on the crossed relationship questions actually deteriorated over time, "even though performance required subjects to take these relationships into account." Since subjects seem to lack verbalizable knowledge of the non-salient relationships, Broadbent and Berry refer to an *implicit* learning mechanism to explain the performance of the participants. In short, the authors assume that direct (or salient) relationships are learned by explicit learning, while the indirect (or non-salient) relationships are implicitly learned (Dienes & Berry 1997: 10; Berry 1993: 26).

To analyze the validity of the assumption that "successful performance cannot be based on the salient relationships alone" we generalized the instance-based learning approach outlined in the previous section. The proposed ACT-R

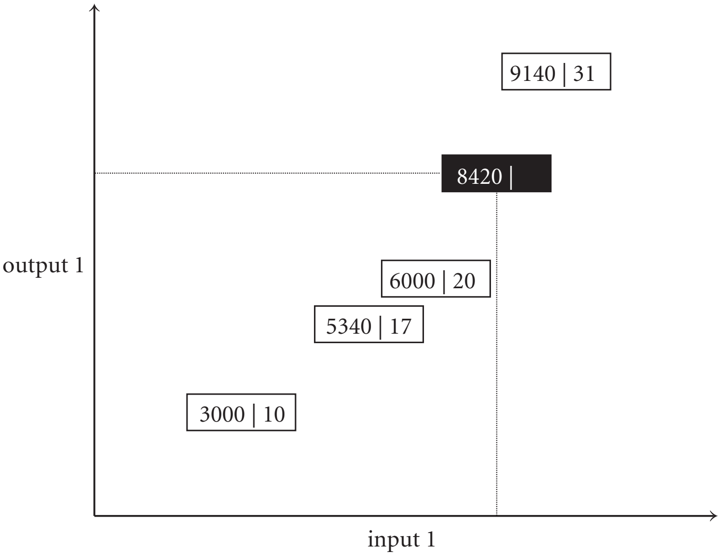


**Figure 4.** Illustration of the blending mechanism

model challenges the view of Berry and others (see Broadbent & Berry 1987) by substantiating that instance-based learning that only represents pairs of encountered input-output values without explicitly encoding structural knowledge about causal relationships between variables is clearly sufficient to successfully control the TRANSPORTATION task. The model makes use of the ACT-R *blending mechanism* (Lebiere 1999) that does not retrieve a single chunk from declarative memory, but rather the value(s) that best satisfy the constraints expressed by an entire set of chunks, with each chunk weighted by its probability of retrieval. This allows relations that are available in declarative memory to have a bearing upon the similarity-based retrieval process without needing to be explicitly formulated. Figure 4 illustrates the blending mechanism showing a situation where memory is probed for an input value to achieve a passenger load of 8420.

The TRANSPORTATION model encodes and retrieves separate pairs of input/output values of the saliently connected system variables. This representation assumes that subjects are aware of the *salient relationship* between $t \rightarrow L$ and $f \rightarrow S$, respectively. The separate representation can be justified by the results of a questionaire that was given to participants of the *explicit* condition of an experiment reported below *prior* to having access to the TRANSPORTATION task. The data of the experiment revealed that 37 out of 40 participants

**Table 2.** The two central productions of the Transportation model

```
(p fee-retrieval
   =goal>
      isa spaces-fee
      spaces =spaces
   =encoded-fact>
      isa spaces-fee
      spaces =spaces
      fee =fee
==>
   =goal>
      fee =fee
)
```

```
(p interval-retrieval
   =goal>
      isa load-interval
      load =load
   =encoded-fact>
      isa load-interval
      load =load
      interval =interval
==>
   =goal>
      load =interval
)
```

assumed a positive relationship between $f$ and $S$ and 33 out of 40 participants assumed a positive relationship between $t$ and $L$ *before* they were exposed to the Transportation task. In a post-task questionaire virtually all subjects were able to verbalize knowledge of the respective salient relationships (40 out of 40 participants assumed a positive relationship between $f \rightarrow S$, 39/40 participants assumed a positive relation of $t \rightarrow L$). Table 2 shows the central productions of the Act-R model for the Transportation task. The productions fee-retrieval and interval-retrieval probe declarative memory for encoded values for the *spaces* or *load* targets, and retrieve a weighted average for the respective input values. Note the direct similarity between these productions and the retrieve-episode production for the SugarFactory task.

To empirically evaluate the performance of the model, we compared it to data that we gathered in a study using the Transportation task. 40 subjects (20 male, 20 female, mean age: 22,3 years, SD=2.9) from the Saarland University participated in the study for course credit. To manipulate the learning orientation of the participants, we introduced two experimental conditions in an initial training phase: an *implicit condition* and an *explicit condition.* In the *implicit* condition participants (N=20) were given 2 problems (i.e. required target combinations of output values for $L$ and $S$) as training trials with the instruction to reach the desired target values for $L$ and $S$. To establish an incidental learning situation, no instruction was given to actively explore the underlying system structure, i.e. the relations between its variables. In the *explicit* group, participants were instructed to freely explore the internal structure of the system in the training phase. They could make interventions and observe the resulting outcomes to test hypotheses about the internal workings of the Transportation task for a maximum of 30 trials. In contrast to an incidental learning situation, explicit learning is generally characterized as the use of deliberate strategies such as generating or testing hypothesis. Consequently, the experimental conditions were intended to induce two learning modes.

After the learning phase, both groups then worked in a subsequent control phase on 6 problems (combinations of values for $L$ and $S$) that were also used in Broadbent's initial study (Broadbent & Aston 1978; Broadbent 1977). Control performance in this phase was measured by the number of trials necessary to achieve the respective target-value pairs. Since no statistical difference on acquired knowledge (as determined by post-task questionnaires) or achieved task performance between the experimental conditions was found, the following empirical evaluation of the Act-R Transportation model does not differentiate between the respective experimental groups. At the end of this section we will briefly discuss a possible interpretation for the lack of difference between the two conditions.

As mentioned in the previous section, every instance-based learning account needs as a preliminary basis for processing either initial memory representations of instances that can be retrieved or algorithmic knowledge to generate instances of encoded episodes. Since the individual problem-solving traces of all participants in the experiment sketched above were recorded, we created individualized Act-R models that encode the input-output pairs that each assigned subject produced in the training phase of the experiment. The basic model, essentially comprising the two productions shown in Table 2, was thus individually assigned to every single participant in the experiment, encoding the respective participant's problem-solving episodes from the training phase in declarative chunks. In this sense, the individualized model and the respective modeled participant share the same history of problem-solving episodes from the training phase. As illustrated in Table 2, the resulting chunks encode the input-output values of the directly related variables only.

To predict the performance in the *control phase* of the experiment, the model was run on the original six target values for the output values of $L$ and $S$. One architectural parameter (*activation noise*) was globally set to 0.25 to fit the data observed in the experiment.[2] The model was evaluated by comparing its performance to the participant's average number of trials to target as well as to error data. As shown in Figures 5 and 6, the model's performance is, despite its lack of knowledge of the crossed relationship between $f \rightarrow L$ and $t \rightarrow S$, nevertheless quite close to the performance of the participants. Figure 5 demonstrates that model data and experiment data on the number of trials to reach the respective target values correspond well on a qualitative and quantitative level.

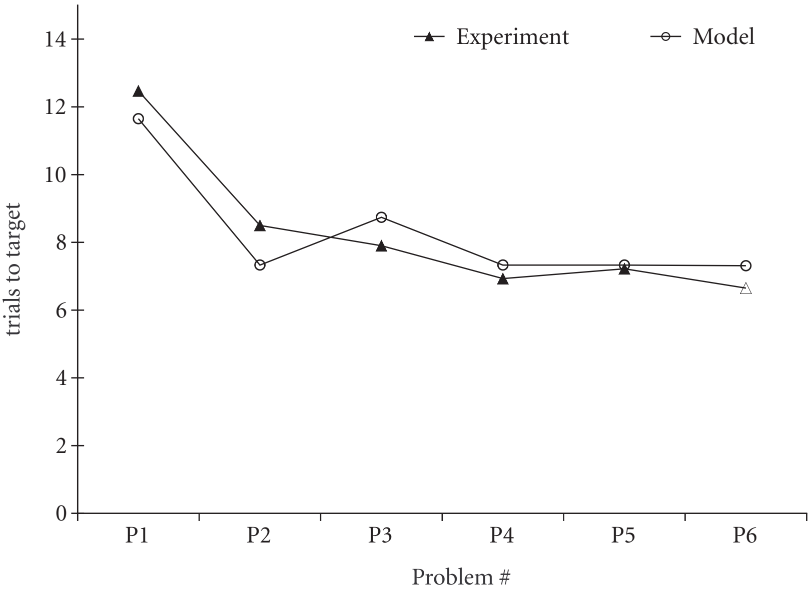


**Figure 5.** Average number of trials to target

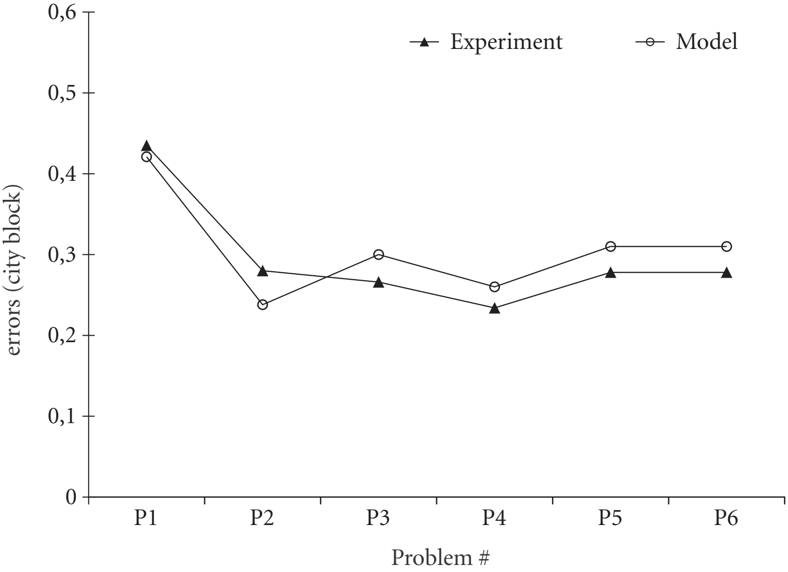


**Figure 6.** Average number of errors

As Figure 6 exemplifies, the average number of errors, defined as an increase of the distance to the required target values from trial $n$ to trial $n + 1$ according to a City-Block-Metric, is also within the empirically observed range.

We can conclude from the comparison of empirical data and model data that the ACT-R model is not only successful in capturing the observed data pattern, but that it does so without encoding the non-salient relationships that were claimed to be essential for successful performance. Thus, the model demonstrates that learning the non-salient, or indirect, relationships is not necessary for explaining the performance in the experiment. Instead, a simple ACT-R model that relies on the similarity-based retrieval of chunks representing input-output pairs of the saliently connected variables is sufficient to account for the observed performance. As with the SUGARFACTORY model, chunks encoding the effects of certain input values on output variables form the model's explicit knowledge, while implicit processing comes into play in the retrieval of knowledge units based on subsymbolic activation quantities.

As noted above, the two experimental conditions to induce implicit or explicit learning in the TRANSPORTATION study did not lead to different scores in control performance or acquired knowledge. This result might, however, not be surprising since it seems to be questionable whether experimental manipulations of the participant's learning orientations in task control are practically effective. Even in the implicit condition participants seem to be actively searching for the rules underlying the task. Since participants assume that knowledge about the system structure may be helpful in achieving given target values, it can hardly be excluded that participants actively try to learn about the task structure. In fact, the semantic cover story of TRANSPORTATION already suggests a positive relationship between the fee and the number of free spaces in the car park (i.e. the more expensive the fee, the more parking spots available), as well as between the interval and the bus load (i.e. the longer people have to wait, the more passengers will be waiting to fill the bus). Knowledge about these salient relationships was, however, shown to be sufficient for successful performance. It can thus be questioned whether control tasks provide appropriate instruments for establishing an incidental learning situation. A paradigm that overcomes the problems mentioned is the investigation of sequence learning in the serial reaction time task, which is explored in the next section.

## 5. Implicit learning in the serial reaction time task

Sequence learning was recently described as the best paradigm for studying implicit learning (Destrebecqz & Cleeremans 2001). In a typical sequence learning experiment subjects are exposed to visuospatial sequences in a compatible response mapping serial reaction time task. Establishing an incidental learning situation, participants are introduced to the task as a reaction time experiment where they are required to react to 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 usually 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 that features regular transitions between successive stimuli.

In their seminal experiment Nissen and Bullemer (1987) used a recurring loop of a "D-B-C-A-C-B-D-C-B-A" sequence for ten successive repetitions for each of a number of blocks, where each letter designates a specific horizontal position on the screen. The succession of signals appearing at the spatial positions results in a continuous stream of events to which subjects have to respond, whereby a 200-500ms response-to-stimulus interval (Rsɪ) is typically used. Sequence learning is said to have occurred when (a) subjects exposed to systematic event sequences show faster response latencies and produce fewer errors than those responding to random event signals, or (b) response times of subjects increase significantly when systematic sequences are temporarily switched to random signals. The faster response times are interpreted as resulting from acquired knowledge about the pattern of stimuli that allows the 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 number of studies found that participants – despite showing a significant speed-up in the sequence blocks in comparison to random blocks – often failed to express verbalizable knowledge about the sequence pattern (Willingham, Nissen, & Bullemer 1989; Curran & Keele 1993; Cohen, Ivry, & Keele 1990). As in other paradigms of implicit learning, the dissociation between performance in the serial reaction time task and in subsequent direct memory tests is interpreted as evidence for the implicit nature of the underlying learning process.

## 5.1 An ACT-R theory of sequence learning

We have recently proposed (Lebiere & Wallach 2000) an integrative theory of sequence learning based on the ACT-R cognitive architecture that was successfully applied to data from a number of classic studies in the field (Willingham, Nissen, & Bullemer 1989; Perruchet & Amorim 1992; Curran & Keele 1993). The experimental conditions in these studies vary widely with regard to the type of sequence used, the length of the sequence, the distribution of systematic vs. random sequence 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 ("tone counting"). Model validation was not restricted to comparing model-generated and empirical data on a single dimension, but included a comparison of latencies, learning trajectories, errors, stimulus anticipations and individual differences as well as the structure of acquired chunks.

A basic assumption of the theory is that the mappings between stimulus locations and response keys in an experiment are encoded as declarative chunks. Each chunk associates the respective stimulus location on the screen to a desired response key. These declarative representations essentially represent a straightforward explicit encoding of the experimental instructions informing the subjects of the stimulus-response mappings in the experiment. 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 base-level learning mechanism that strengthens a chunk to reflect its frequency of use (see Section 3 of this chapter). Subsymbolic activation processes make a chunk active to the degree that past experience and the present context (as given by the current goal to react to the next stimulus) indicates that it is useful at this particular moment. In the ACT-R sequence learning model, these reinforcements will lead to higher activation levels for the chunks that map stimulus positions to the keys to be pressed, which then results in faster response latencies. This speedup will occur independently of whether the stimulus sequence is systematic or random because it only depends upon the frequency of each retrieval.

The fundamental assumption of the ACT-R 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 a small number of previous stimuli remain in the encoding of the new goal.[3] This assumption has two important implications. First, since every goal contains both the current stimulus and previous one(s), 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 constitutes 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(s) are components of the goal and thus sources of activation. This co-occurrence between previous stimuli (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 stimuli and mapping chunks and thus facilitates further processing. The subsymbolic strengths of associations between consecutive sequence fragments constitute the model's implicit knowledge of the sequence.

In the next section we describe an application of the ACT-R sequence model to an experiment by Destrebecqz and Cleeremans (2001).

## 5.2 Evidence for implicit sequence learning: Destrebecqz and Cleeremans (2001)

Destrebecqz and Cleeremans (2001) recently published an ingenious experiment to explore the contribution of implicit learning in the serial reaction time task. Using an adaptation of the *Process Dissociation Procedure* (Jacoby 1991), Destrebecqz and Cleeremans have provided convincing evidence for unconscious knowledge acquisition by disentangling explicit and implicit learning processes. In their study, subjects who were denied preparation for the next stimulus of a sequence by using a response-to-stimulus interval (RSI) of 0 showed significant knowledge of the sequence as expressed in task performance. Interestingly, participants in the No RSI condition could not refrain from expressing the acquired knowledge even when specifically instructed to do so in an exclusion task. By contrast, subjects who were exposed to a RSI of 250ms were successful in performing the exclusion task.

In their experiment, Destrebecqz and Cleeremans (2001) used two so-called *second-order conditional* (SOC) sequences (Reed & Johnson 1994) in which knowledge of two successive elements is necessary to predict the location of the next stimulus. The two SOC sequences were balanced for stimulus frequency and position. The two sequences comprised the following succession of signal locations: 342312143241 (SOC1) and 341243142132 (SOC2). During the serial reaction time task participants had to react to SOC sequences of length 12 for a total of 15 blocks with each block consisting of 96 trials. As

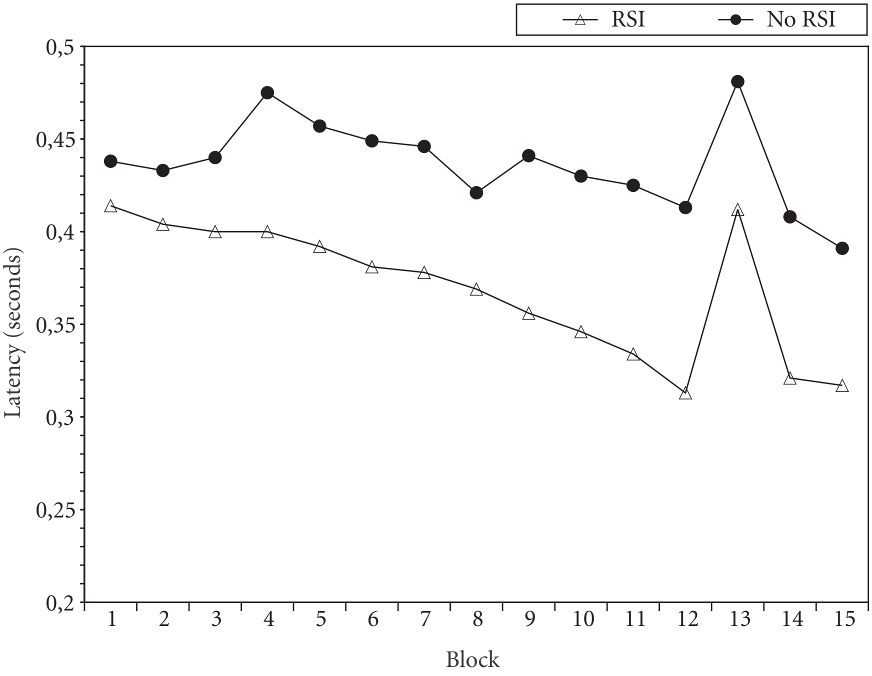


**Figure 7.** Mean reaction times for the training blocks (empirical data)

noted above, a 0ms stimulus-to-response interval was introduced in the No Rsi condition, while a 250ms stimulus-to-response interval was used in the Rsi group. In each condition, half of the participants were trained on Soc1 during the first 12 blocks, then switched to Soc2 in block 13 and finally switched back to Soc1 for blocks 14 and 15. This design was reversed for the second half of the subjects. As can be seen in Figure 7, participants show significantly increased reaction times in block 13, as well as clear effect of the different Rsi conditions (p<.0001). Participants in the Rsi condition are generally faster than those in the No Rsi condition. As Destrebecqz and Cleeremans (2001) note, this may either result from improved learning or from an improvement in expressing available knowledge. Learning was, however, not completely suppressed in the No Rsi condition as indicated by a clear increase in latency in block 13 when participants switched to a new Soc, and in a corresponding decrease of their reaction times when switching back to the original Soc. The authors therefore concluded that the training sequence was learned in both Rsi conditions.

In Figure 8 the performance of the Act-R model in the Destrebecqz and Cleeremans (2001) experiment is plotted.

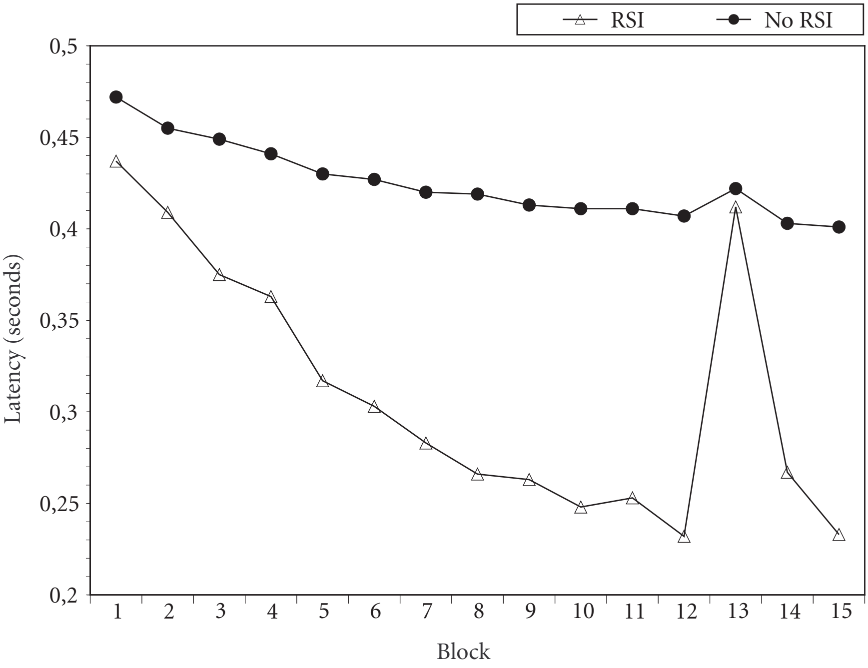


**Figure 8.** Mean reaction times for the training blocks (model data)

All basic characteristics of the data are reproduced. The model is faster in the RSI condition because it sometimes retrieves sequence chunks before the stimulus appears. The anticipation allows it to bypass the mapping of stimulus location to response key after the stimulus appears since the response key is already selected. When switching sequences, this advantage disappears and the performance of the RSI model is degraded to the no-anticipation case. Performance in the No RSI model also improves, not by retrieving sequence chunks but by retrieving mapping chunks faster through the learning of strengths of associations between stimuli and mapping chunks to reflect the structure of the sequence. The degradation of performance when switching sequences is smaller than in the RSI condition.

To measure the ability to project the acquired sequence knowledge in direct memory tests, Destrebecqz and Cleeremans (2001) applied a *free generation task* that participants performed under *inclusion* and *exclusion* instructions. After completion of the training task, participants were informed that the succession of signals followed a systematic repeating pattern. In the *inclusion* instruction, participants were then required to freely generate a total of

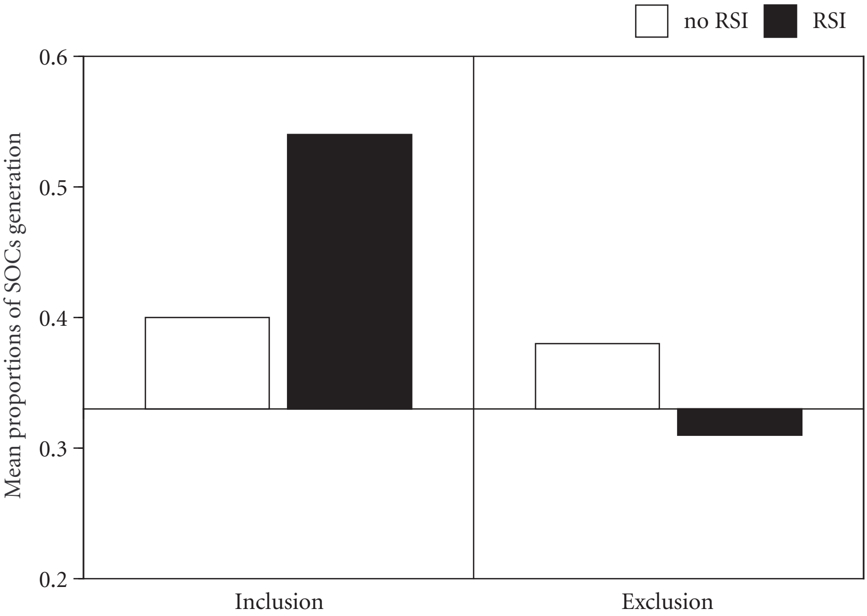


**Figure 9.** Mean proportion of generated second order transitions that were part of the training sequence (inclusion and exclusion condition) from Destrebecqz and Cleeremans (2001)

96 trials that "resembled the training sequence as much as possible". Participants were told to "rely on their intuitions when feeling unable to recollect the location of the next stimulus" (Destrebecqz & Cleeremans 2001). In the *exclusion* instruction they were then asked to generate another sequence of 96 trials. However, this time they were told to *avoid* reproducing the structural pattern previously experienced in the Srt task. Depending on the experimental condition, the stimulus moved when participants pressed a key and appeared with a delay of 0 msec or 250 msec at the corresponding location on the screen.

Figure 9 shows the average inclusion and exclusion scores for the two experimental conditions. The results of the inclusion condition confirm those of the training task: participants from both experimental groups were obviously able to use the acquired knowledge to indicate the sequence structure to which they had to react during the training task. In both conditions, the mean proportions are significantly ($p<.005$) above the chance level of .33 (see Destrebecqz and Cleeremans (2001). The difference in performance between the No Rsi and the Rsi condition is, however, only marginally significant ($p=0.07$). In the *exclusion* instruction, a highly interesting picture emerges: While participants

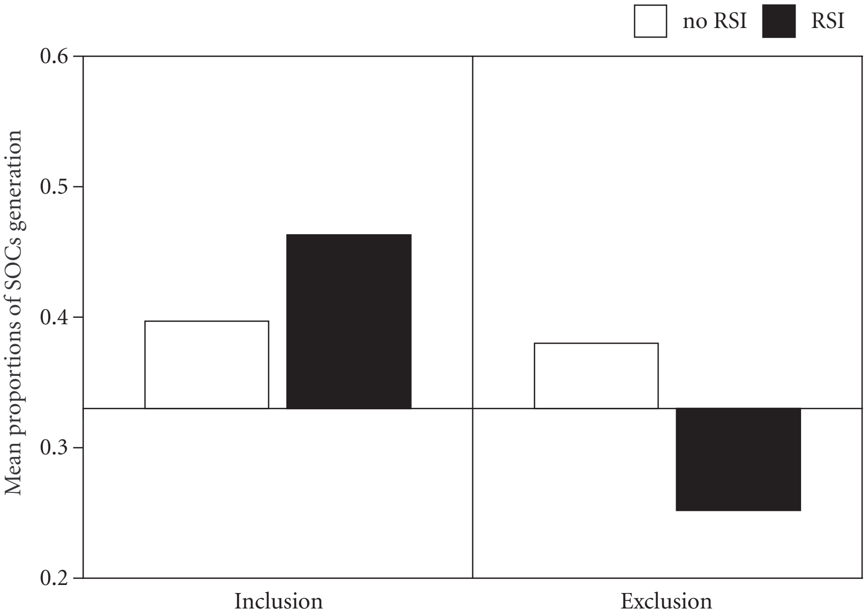


**Figure 10.** Mean proportion of generated second order transitions that were part of the training sequence (inclusion and exclusion condition) as modeled in Act-R

from the Rsi condition performed at slightly worse than chance level, participants from the No Rsi were not able to refrain from generating fragments of the training sequence ($p<.01$). It thus appears that participants that were trained in the No Rsi condition appear to have little control over the knowledge acquired, which Destrebecqz and Cleeremans (2001) interpret as evidence for unconscious learning.

In sum, the experimental data provided by Destrebecqz and Cleeremans (2001) can be interpreted as indicating that sequence learning in the No Rsi condition can at least in part proceed without conscious control and is thus assumed to be implicit. The increased response-to-stimulus interval in the Rsi, on the other hand, seemed to have exclusively improved explicit learning. Figure 10 presents the results of the Act-R model on the generation of the Socs under inclusion and exclusion instructions for the two Rsi conditions.

As with the prediction of the latency data, the Act-R model clearly captures the qualitative data pattern observed in the generation task. In the *inclusion* condition, the Rsi model can generate a significant proportion of the sequence by retrieving sequence fragments. In the *exclusion* condition, it inverts that explicit knowledge by choosing another stimulus than the one specified by the sequence chunk retrieved. The No Rsi model generates a smaller proportion of sequence fragments than the Rsi model because the responses result from the implicit knowledge learned in the strengths of associations, which is less uniquely determinant than direct sequence fragments. However, that knowledge is not subject to reversal in the *exclusion* condition.

Let us describe in more detail the model's procedural knowledge that uses the 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* types the respective key. 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 retrieved key can be typed directly without the need for mapping location to key. When a response has been given, if no stimulus has yet appeared (positive Rsi) the production *New-Goal* creates and focuses on a new goal with the context shifted forward by one stimulus, which will lead to the encoding of the current stimuli subsequence in the current goal. If the stimulus has already appeared however (No Rsi), then the production *Clear-Goal* simply shifts the current context by one stimulus within the current goal without creating a new chunk.

A similar set of productions is used in the generation task. Production *Generate*, similar to *Guess*, attempts to retrieve a sequence fragment given the current context. This fragment represents explicit knowledge of this part of the sequence. If successful and in the exclusion condition, production *Invert* randomly picks a stimulus that is neither the one predicted by the sequence fragment nor the same as the previous one.[4] Otherwise, production *Guess-Location* simply retrieves the most active mapping chunk, which will partly reflect the strengths of associations to the current context. Note that unlike the sequence chunk retrieved by *Generate*, the response generated by this mapping is not subject to inversion but instead represents the implicit knowledge of the sequence that is not subject to cognitive control. Production *Generate-Response* then types the response and updates the goal.

In this section we have shown that the Act-R sequence learning model, which has been successfully applied to a number of different experiments with the serial reaction time task, is also capable of accounting for the empirical data reported by Destrebecqz and Cleeremans (2001). The methodological approach followed in the latter study allowed them to disentangle implicit and explicit learning and resulted in a differentiated picture of the relationship between explicit and implicit learning, offering a challenging and precise benchmark to explore the scope of the model. While a central criticism of previous models of implicit learning has been that they fail to account for explicit learning and how it is different from implicit learning (Stadler & Roediger 1998), ACT-R seems to provide a promising approach to precisely investigate the interplay of both forms of learning.

## 6. Conclusion

In this chapter we have explored the scope of a view on implicit and explicit learning based on the ACT-R cognitive architecture. We have shown that the proposed approach can successfully account for the observed experimental performance in two different domains of research on implicit learning. On a general level, we have attributed explicit learning to ACT-R's declarative chunk learning mechanism, while implicit learning is modeled using subsymbolic activation learning mechanisms. In this sense, the proposed approach can be regarded as a computational instantiation of a view expressed by Cleeremans and Jiménez (1998: 328) in which the authors relate sequence learning to "two different kinds of learning: One process involves memorizing a series of successive events, and another involves developing sensitivity to the statistical structure of the material". The models proposed in this chapter have shown how the ACT-R architecture can be applied in this spirit to investigate how both forms of learning can influence behavior. In our view, computational models based on a hybrid cognitive architecture provide promising steps towards a precise, integrative theory of implicit and explicit learning.

## Notes

* The authors want to thank Dr. Kevin Gluck for his valuable comments on an earlier version of this manuscript.

1. Following Lebiere (1999), the following ratio function is used to calculate the similarity of two values *a* and *b*, representing the sugar production or number of workers in respective instance chunks: $sim(a, b) = \frac{\min(a,b)}{\max(a,b,1)}$. This similarity function scales well to widely different

number values such as those encountered here. Note that the similarity ranges from 0 (no similarity) to 1 (equality).

2. In keeping with the search for constraints across models developed in a common architecture, this value has been used in many other ACT-R models (e.g. Lebiere 1998; Lebiere & West 1999).

3. The number is here fixed at 2, which provides the required power to learn the sequence. Lebiere and Wallach (2000) showed that there is a tradeoff between length of fragments (and hence learning power) and speed of learning. They also discuss a model variant that gradually learns chunks of increasing length.

4. Participants were told not to repeat responses in the generation task.

## References

Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: LEA.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: LEA.

Anderson, J. R. (1978). Arguments concerning representations from mental imagery. *Psychological Review*, *85*, 249–277.

Anderson, J. R. & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: LEA.

Anderson, J. R., Reder, L., & Lebière, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*, 221–256.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396–408.

Berry, D. C. (1997). *How Implicit is Implicit Learning?*. Oxford: Oxford University Press.

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

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

Berry, D. C. & Broadbent, D. E. (1987). The combination of explicit and implicit learning processes in task control. *Psychological Research*, *49*, 7–15.

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

Berry, D. C. & Dienes, Z. (1993). *Implicit Learning: Theoretical and Empirical Issues*. Mahwah, NJ: LEA.

Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology*, *29*, 181–201.

Broadbent, D. E., Fitzgerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, *77*, 33–50.

Broadbent, D. E. & Aston, B. (1978). Human control of a simulated economic system. *Ergonomics*, *21*, 1035–1043.

Buchner, A., Funke, J., & Berry, D. (1995). Negative correlations between control performance and verbalizable knowledge: Indicators for implicit learning in process control tasks? *Quarterly Journal of Experimental Psychology*, *48A*, 166–187.

Byrne, M. D. & Anderson, J. R. (1998). Perception and action. In J. R. Anderson and C. Lebiere (Eds.), *The Atomic Components of Thought* (pp. 167–200). Mahwah, NJ: LEA.

Cleeremans, A. (1993). *Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing*. Cambridge, MA: MIT Press.

Cleeremans, A. & Jiménez, L. (1998). Implicit sequence learning: The truth is in the details. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learning* (pp. 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.

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

Dienes, Z. & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 848–861.

Dienes, Z. & Fahey, R. (1998) The role of implicit memory in controlling a dynamic system. *Quarterly Journal of Experimental Psychology*, *51A*, 593–614.

Dulany, D. E. (1997). Consciousness in the explicit (deliberative) and implicit (evocative) senses. In J. D. Cohen and J. W. Schooler (Eds.), *Scientific Approaches to Consciousness* (pp. 179–212). Hillsdale, NJ: LEA.

Dulany, D. E. (1961). Hypotheses and habits in verbal 'operant conditioning'. *Journal of Abnormal and Social Psychology*, *63*, 251–263.

Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learn*ing (pp. 47–104). Thousand Oaks, CA: Sage.

Gonzalez, C, Lebiere, C., & Lerch, F. J. (1999). ACT-R Learning in a Real-Time Dynamic Decision-Making Task. Presented at the 1999 ACT-R Workshop, George Mason University, Fairfax, VA.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs*, No. 123.

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

James, W. (1990). *Principles of Psychology*. New York: Holt.

Jenkins, J. G. (1933). Instruction as a factor in 'incidental' learning. *American Journal of Psychology*, *45*, 471–477.

Kolers, P. A. & Roediger, H. L. (1984). Procedures of Mind. *Journal of Verbal Learning Behavior*, *23*, 79–91.

Lebiere, C. (1999). The dynamics of cognition: an Act-R model of cognitive arithmetic. *Kognitionswissenschaft, 8*, 5–19.

Lebiere, C. (1998). *The dynamics of cognition: an ACT-R model of cognitive arithmetic*. Ph.D. Dissertation. Technical Report CMU-CS-98-186. Also available on the web at http://reports-archive.adm.cs.cmu.edu/

Lebiere, C. & Wallach, D. (2000). Sequence Learning in the ACT-R Cognitive Architecture: Empirical Analysis of a Hybrid Model. In R. Sun (Ed.), *Sequence Learning: Paradigms, Algorithms, and Applications*. Berlin: Springer Lecture Notes in Computer Science.

Lewicki, P., Czyewska, M., & Hill, T. (1997). Nonconscious information processing and personality. In D. C. Berry (Ed.), *How Implicit is Implicit Learning?* (pp. 48–72). Oxford: Oxford University Press.

Logan, G. D. (1990). Repetitive priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*, 1–35.

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

Manza, L. & Reber, A. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In D. C. Berry (Ed.), *How Implicit is Implicit Learning?* (pp. 73–106). Oxford: Oxford University Press.

Mathews, R. C. & Roussel, L. G. (1997). Abstractness of implicit knowledge: A cognitive evolutionary perspective. In D. C. Berry (Ed.), *How Implicit is Implicit Learning?* (pp. 13–47). Oxford: Oxford University Press.

McGeorge, P. & Burton, A. M. (1990). Semantic processing in an incidental learning task. *Quarterly Journal of Experimental Psychology*, *42A*, 597–609.

Neal, A. & Hesketh, B. (1997). Episodic knowledge and implicit learning. *Psychonomic Bulletin and Review, 15*, 24–37.

Newell, A. & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the power law of practice. In J. R. Anderson (Ed.), *Cognitive Skills and their Acquisition* (pp. 1–56). Hillsdale, NJ: LEA.

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

Partee, B., ter Meulen, A., & Wall, R. E (1990). *Mathematical Methods in Linguistics*. Kluwer Academic Publishers: Dordrecht.

Perruchet, P. & Gallego, J. (1997): A subjective unit formation account of implicit learning. In D. C. Berry (Ed.), *How Implicit is Implicit Learning?* (pp. 124–161). Oxford: Oxford University Press.

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.

Reber, A. S. (1993). *Implicit Learning and Tacit Knowl*edge. New York: Oxford University Press.

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

Reber, A. S. (1969). Transfer of grammatical structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115–119.

Reber, A. S. & Milward, R. B. (1968). Event observation in probability learning. *Journal of Experimental Psychology*, *77*, 317–327.

Reed, J. & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 585–594.

Rubin, D. C. & Wenzel, A. E. (1990). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734–760.

Sejnoski, T. J. & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems, 1* (1), 145–168.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.

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

St John, M. F. & Shanks, D. R. (1997). Implicit learning from an information processing standpoint. In D. C. Berry (Ed.), *How Implicit is Implicit Learning?* (pp. 162–194). Oxford: Oxford University Press.

Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 819–827.

Stadler, M. A. (1989). On learning complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 1061–1069.

Stadler, M. A. & Roediger, H. L. (1998). The question of awareness in research on implicit learning. In M. A. Stadler and P. A. Frensch (Eds.), *Handbook of Implicit Learning* (pp. 105–132). Thousand Oaks, CA: Sage.

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

Thorndike, E. L. & Rock, R. T. (1934). Learning without awareness of what is being learned or intent to learn it. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17*, 1–19.

Wallach, D. & Lebiere, C. (1999). Example-based models of control problems. Paper presented at the 1999 ACT-R Workshop, George-Mason University, Fairfax, Va. See http://hfac.gmu.edu/actr99/.

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.

Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Wilhelm Engelmann.

# Part III

# Reciprocal influences: Implicit learning, attention, and beyond

# Visual orienting, learning and conscious awareness

Tony Lambert*
University of Auckland

## 1. Scope of the chapter

In his celebrated nineteenth century text *Principles of Psychology*, William James (1890/1983) wrote that

> "Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others ..."

In many respects, this frequently quoted passage is admirably clear and appropriate well over a century after it was written. However, one might also be tempted to elaborate on James' introductory phrase, and remark that although everyone *thinks* that they know what attention is, the complex and multi-faceted nature of this concept can easily lead the unwary to confusion and pitfall. In discussing the relation between attention and implicit learning, the topic of this volume, it is especially important to distinguish between two distinct but related meanings of the term attention. A central issue in many chapters of this volume is the relation between implicit learning and attention in the sense of mental resources or effort (e.g. see the chapters by Jiménez and by Shanks). That is, they are concerned with the question of whether processes of implicit learning require mental effort or resources, or whether they are entirely capacity free and can proceed without interfering with other mental processes. This chapter has a very different focus, since it is concerned with the relation between implicit learning and attention in the sense of selective processing or orienting. James was clearly aware that the term attention has at least two distinct meanings. On the one hand, it is often used to refer to the idea of mental resources and their limitations. Phrases such as '*limited attention*' or '*attentional capacity*' imply the notion of attention as a resource; indeed several influential theories of attention have attempted to provide quantitative descriptions of the way in which *attention as resource* can be divided between concurrent mental activities (e.g. see Wickens, Gordon, & Liu 1998). This notion of attention as a limited resource may be seen as a *cause* of selectivity. That is, because resources are limited, processing must be allocated selectively to a sub-set of the available input information. However, when one speaks, for example, of selectively attending to a certain object, it is clear that the term attention is being used here to refer to the *selectivity itself*. Hence, the term attention can be used to refer both to the cause of processing selectivity (i.e. limited resources) and to the selective processes themselves. The present chapter is concerned with attention in this latter sense of selectively orienting towards some sensory objects rather than others. It is concerned with the dynamics of visual orienting and with effects of learning, both implicit and explicit, on the processes involved in shifting attention from one visual object to another.

Consciousness is intimately related both to attention and to implicit learning, the topics of this volume. As William James' definition makes clear, a primary consequence of selectively attending is that the object of selective attention gains clearer representation in consciousness. However, there is a further issue of consciousness which is of particular interest here – that is the relationship of consciousness to attentional learning. The latter refers to learning that occurs when repeated experience with a task affects the deployment of selective attention. As will be seen, this learning results in an improvement in the speed and efficiency of orienting towards objects of interest. Although substantial research efforts have been directed at studying performance improvements that occur as a result of practice (e.g. Shiffrin & Schneider 1977; Phillips & Triggs 2001) and at investigating processes of visual selective attention, current theoretical accounts of visual selective attention (e.g. see Wright 1998) have little to say concerning possible influences of learning on orienting. This is perhaps curious since William James was clearly aware that learning and experience played an important role in the control of selective attention. In addition, a substantial amount of animal research has been concerned with effects of learning, especially habituation, on the orienting response (e.g. Barry & Sokolov 1993). Nevertheless, current theoretical treatments of visual attention are relatively silent with regard to the role of learning and experience in the control of visual orienting. This chapter provides an overview of recent work carried out in our laboratory which has attempted to remedy this theoretical lacuna. The recent and interesting work of Chun & Jiang (1998, 1999; Jiang & Chun, this volume) has also been concerned with this important issue.

## 2. Vision and selectivity

As many authors have remarked, humans are highly visual animals. According to some estimates approximately 60% of brain tissue is concerned with visual processing. The intimate relation between perception and attention is plainly evident in the design of the visual system. Even at the level of gross anatomy it is clearly apparent that an essential design feature of the visual system is that it is *selective* (i.e. attentional) in nature. If the eyes remain stationary we are able to respond to simple visual objects (e.g. luminance increments) within a field that subtends about 200° horizontally and 135° vertically. However, as is well known, visual acuity declines with visual eccentricity. Although one may be aware of a waggling finger in the far periphery, in order to discriminate the fine detail of the finger it must be brought into central vision. It is generally accepted that high acuity colour vision is subserved by the cone receptors of the retina. Anatomical studies of retinal topography have shown that the density of cone receptors undergoes spectacular decline as one moves outwards from the centre of the fovea. Curcio et al. (1990) reported a peak density of 199,000 cones/mm$^2$ at the foveal centre of the average retina. However, at an eccentricity of just half a degree from the foveal centre, this density had declined by 50%. At an eccentricity of 3–4° cone density was about 20,000/mm$^2$ – i.e. about 10% of the density at the foveal centre. Similarly, psychophysical studies have shown that visual acuity also declines sharply as a function of eccentricity. For example, visual acuity has declined by more than 50% at 5° relative to acuity at the foveal centre (Chapanis 1949; Schiffman 1990). Even within central vision there is an eccentricity dependent decline in acuity, so that acuity at 2° is clearly worse than acuity at the centre of the fovea. In the psychophysical function relating acuity to eccentricity there is no clear boundary between foveal, parafoveal, and peripheral vision, since the decline with eccentricity is smooth rather than discontinuous. However, the fovea is usually defined in terms of the area subserved by the central cone rich region. The diameter of this area is 1–2° (Osterberg 1935; Schiffman 1990). The macula is somewhat larger, being about 10° in diameter.

It seems curious that this very basic form of selectivity has not always featured prominently in theories of visual attention. Indeed until quite recently much debate centred on the cognitive issue of whether processing selectivity occurred at an early or late stage of information processing (e.g. see Allport 1989; Lambert 1985). In more recent times it has become apparent that purely cognitive theories of attention, in terms of information processing operations, are sorely incomplete, and that an adequate account of attention must be constrained and informed by knowledge of the neural hardware that underpins selectivity. From this cognitive neuroscience perspective it is obvious that processing selectivity occurs at the earliest possible stage of processing – in the retina.

## 3. Measuring the 'phenomenal fovea'

Although most of us are aware of a difference in visual acuity between central and peripheral vision, the tiny area served by the fovea and macula may be seen as surprising. The small size of these areas of central high acuity vision is certainly a surprise to beginning students and to other individuals untainted by education about visual perception or sensory physiology. If I close my eyes for a few seconds and then open them, my immediate subjective impression is that I see an entire scene laid out with clarity. I am *not* aware of a large expanse of poorly resolved, grainy and relatively colourless vision, which covers most of the visual field, apart from a relatively tiny central area, in which objects are seen with full clarity and colour. This suggests a poor correspondence between the careful and objective measurements made by psychophysicists concerning the relation between visual performance and eccentricity and our phenomenal experience of seeing. That is, the strong degree of selectivity in visual processing, apparent from both anatomical and psychophysical observation, does not appear to be clearly represented in visual awareness. In order to gain some quantitative leverage on this issue, Caroline Heffer and I designed a brief questionnaire which aimed to probe the phenomenology of the distinction between central and peripheral vision, as represented in the visual experience of naïve observers. These observers were equipped with only rudimentary knowledge of the visual system, but had a lifetime of visual experience to draw upon.

The first item of our questionnaire had the following wording.

> "Most people are aware that an object in central vision is seen with better clarity and detail than a similar object in peripheral vision. This area of detailed central vision is known as the fovea. The aim of this exercise is to measure subjective estimates of the size of this central area of clear vision in different people.

> Imagine you are looking at a (very!) large poster which fills your entire field of view. About how much of the poster do you think would fall within the fovea – the area of detailed central vision?
>
> _____ % of the poster would be within the area of detailed central vision"

The correct answer to this question is extremely small – the fovea would occupy less than 0.1% of the total area of the poster. Even the central 10° of vision (the macula) would cover less than 1% of the poster.[1] The mean estimate provided by our participants (N=12) was 34% (S.D.=23) – a dramatic over-estimation, to put it mildly. Part of this inaccuracy might possibly arise from a difficulty in estimating the proportionality of two areas, both of which are roughly circular, such as the fovea and the entire visual field. For example, mathematically naïve respondents may fail to realise that the area of a circle increases in proportion to the *square* of its radius. However, even if one interprets the figure of 34% as referring to relative diameters/radii, rather than relative areas, this would still represent a dramatic overestimation of the extent of the fovea (by a factor of about thirty).

The second item of the questionnaire posed essentially the same question in a slightly more concrete form.

> "Imagine you are looking at a picture situated about 50cms from your eyes. If you kept your eyes perfectly still, details at the edge of the picture would seem somewhat indistinct, while details at the centre would be seen with perfect clarity. I would like you to try and estimate the size of this area of perfectly clear central vision by imagining this situation.
>
> When viewing a picture about 50cms away the size of the area within which I would be able to see details with perfect clarity is about
>
> ____ cms wide by about ____ cms high."

On average, participants judged the dimensions of the fovea as 27.8cms wide (S.D.=22.4cms) by 24.1cms high (S.D. = 17.4cms). Once again, this is an over-estimation. At a viewing distance of 50cms the fovea would occupy an area about 1.7 cms in diameter. The central 10° of vision would cover an area about 8.7cms in diameter.

The third item of the questionnaire introduced an element of actual, rather than imagined visual performance. Participants viewed each of two pictorial scenes, and were provided with moveable cardboard masks. They were then requested to:

> "Please look at the spot in the centre of this picture and try to keep your eyes perfectly still. Now move the cardboard masks inwards from the edge of the picture in order to progressively cover any parts of the picture that appear less

> than perfectly clear due to their distance from the centre. Do this until details at the edge of the visible area appear to be just as clear as details in the centre. Do not worry if after doing this, the masks cover most of the picture. Carry on reducing the size of the picture until you have the impression that **all** of the remaining picture appears in the area of perfectly clear central vision – the fovea."

Participants performed this task while viewing two different pictures. One was an illustration from a child's picture book, and contained a large amount of fine detail; the second was a photograph of dramatic mountain scenery.[2] For the former picture, the average width of the aperture that remained after moving the masks inwards was 8.3°; in the latter case, average aperture width was $11.4^0$. Thus, for this item participants' estimates of foveal extent were still significantly larger than the 1–2° diameter given in most textbook treatments of visual function. Nevertheless, in this case where participants viewed a real visual stimulus estimates of foveal extent were clearly smaller, and closer to correct value, than responses to items (1) and (2) in which participants imagined viewing situations. It might also be noted that although participants adjusted the width of the viewing apertures to be larger than the fovea, there was a broad correspondence between the width of the remaining apertures and the dimensions of the macula (8–10°).

What are we to make of these findings? The question 'How large is the central area where you see things with perfect clarity', can be seen as an issue of criterion placement. Because both acuity and cone density increase continuously with decreasing eccentricity, even within the fovea, the size of the area where acuity reaches an *absolute* peak is miniscule indeed – subtending less than 0.5° of visual angle. All of our respondents felt that an area substantially larger than this satisfied the criterion 'enjoys perfect clarity'. However, it is of interest that participants placed their boundary criterion at an eccentricity that was substantially broader than the 1–2° diameter identified as the cone rich, high acuity fovea by vision scientists. Although it is clear that the hardware of the visual system is only capable of delivering clear and detailed representation over a very small visual area, individuals generally feel *as if* the area of clear vision is far larger than this. There appears to be a very poor correspondence between the extreme selectivity evident from psychophysical and anatomical observation and the texture of everyday visual experience.

This issue of selectivity in the architecture of the visual system prompts a number of interesting questions. To begin with one might ask why the visual system has evolved with this design, in which the fovea is very tiny in comparison with the overall extent of the visual field. One possible answer to this question has come from artificial intelligence work in the area of machine vision. For example, it has been suggested that the raison d'etre for this design feature is essentially *attentional* since it arises from limitations of processing capacity with respect to building perceptual representations of the visual world. When artificial intelligence workers began to grapple with the problem of building machines that could see, it was realised that the business of constructing fully elaborated perceptual representations of multiple visual objects was dauntingly complex. So complex indeed, that to maintain complete on-line representations of *all* the objects within the field of view of a human eye would probably require an unfeasibly large amount of computing power – i.e. a brain the size of a planet. The design solution settled upon in the course of evolution and which may be necessary for seeing machines as well, has been to build a visual system that is highly selective. In such systems fully elaborated perceptual descriptions are constructed and made available to other on-line processes, only for objects that are in the focus of attention (e.g. see Cave 2001). It is often the case, but not always true, that the object of attention and foveal vision will coincide. As is well known, it is also possible to attend *covertly* to objects appearing at non-foveal locations.

Following on from the logically prior issue of the *raison d'etre* of strong selectivity in vision, a second question is to ask why individuals have a tendency to wildly over-estimate the extent of the area which enjoys optimal clarity. I would like to suggest that there are probably two reasons for this. The first arises from the highly dynamic nature of normal visual behaviour – the selectivity implicit in the distinction between foveal and extra-foveal vision is accompanied by an ability to perform rapid movements of attention. This enables currently non-attended objects to be rapidly brought into the focus of attention. When viewing a natural scene several kinds of attention movement may occur. *Overt* attention shifts are likely to occur, in which there is an observable change in the orientation of the eyes and/or head. For example, saccadic eye movements may be observed, which alternate with periods of fixation, in which the eyes remain stationary. In addition there may be rapid movements of *covert* attention between objects in foveal and extra-foveal vision. It is known that allocation of covert attention can enhance the quality of perceptual representation for objects in both foveal and extrafoveal vision (see Wright 1998). This ability to shift the spotlight of focal attention rapidly from one object to another may be at least partly responsible for the subjective impression that the area of clear visual representation is considerably larger than the objectively measured extent of the fovea.

A further reason for this over-estimation may arise from processes of perceptual completion and filling in. The relative importance of top-down and bottom-up processes has of course been a long running subject of theoretical controversy and debate in the psychology of perception. Given that the bottom-up signal for perception is considerably less rich and detailed for extra-foveal compared to foveal stimuli, poverty of the stimulus arguments in favour of the importance of top-down influences (priming, expectations, background knowledge) appear more plausible for this class of stimuli. Indeed, the powerful effect of such influences can easily be experienced by the simple expedient of closing one eye. With monocular viewing we all have an area of approximately 5° by 7° which receives no visual signal whatever. This is of course the blind-spot which arises from the absence of visual receptors within the optic disc, where the optic nerve exits the eye. Interestingly, the dimensions of the blind spot are considerably larger than those of the fovea. However, under normal viewing conditions it is quite impossible to '*see*' this blank area, because the perceptual system automatically constructs a complete visual representation which fills in the area occupied by the optic disc (see also Ramachandran & Gregory 1991).

It is clear then, that in studying the visual system, attentional and perceptual processes are inextricably linked. In an important sense, seeing *is* an attentional process. It is apparent from the gross hardware of the eye that there is strong selectivity in the processing of foveal and non-foveal objects. Such a system requires a means of shifting attention rapidly and effectively, so that objects of interest that initially receive sub-optimal processing from a parafoveal or peripheral location can be brought rapidly and effectively into the focus of attention, which will generally also involve the execution of an eye movement in order to foveate the object.

Some readers might object that although retinal processing is indeed highly selective, this is not *attention*. Attention, it might be argued refers to more centrally located 'processing bottlenecks' or filters (c.f. Broadbent 1958). However, if one views attention (in at least one sense) as referring to processing selectivity that is related to resource limitation, then this writer at least can see no principled reason for not regarding retinal selectivity as an essentially attentional phenomenon.

## 4. Visual orienting

In recent work on attention (e.g. see Wright 1998) processes of visual orienting have been viewed in terms of two orthogonal dichotomies. On the one hand visual orienting may be either *overt* or *covert*. As indicated earlier, overt visual orienting occurs when there is an observable movement of the eyes, head or perhaps of the whole body which results in a reorientation of the eyes towards an object of interest. Covert orienting occurs, in contrast, when there is a change in the focus of attention from one object to another, without any observable change in eye or head direction. Most readers will be familiar with the phenomenon of 'looking out of the corner of your eye', in which an individual appears to be oriented in a particular direction, but is covertly attending to something else. The phenomenon of covert attention was well known to Helmholtz in the nineteenth century and has been studied extensively in recent decades (e.g. see Posner 1980; Posner & Raichle 1994). In addition to being either overt or covert, visual orienting may also be either under voluntary control or elicited reflexively. On the one hand, attention can be drawn in a rapid involuntary way towards a salient change in stimulation occurring away from the current focus of attention. For example, as I write this chapter my attention is occasionally captured involuntarily as a bird flies by the window, which is in my peripheral vision to the right of the computer screen. Alternatively, one can form a voluntary intention to shift attention away from the current focus in order to inspect another part of the visual world. This two-process view of visual orienting is supported by a substantial body of research. Many laboratory investigations of visual orienting have employed tasks in which participants are presented with two different forms of spatial cue (e.g. see Cheal, Lyon, & Gottlob 1994). Voluntary orienting has been studied by presenting participants with a symbolic indicator such as an arrow, in central vision. For example, in an influential study reported by Posner, Nissen and Ogden (1978) the location of a target object was related to the orientation of the central arrow. On 80% of trials with an arrow, the target appeared at the 'valid' location – i.e. the location indicated by the arrow. On 20% of trials the target appeared at the 'invalid' location – i.e. the location opposite to that indicated by the arrow. There was a clear performance advantage, in terms of response latency, for valid relative to invalid trials. In explaining this result Posner et al. proposed that in response to the arrow cue, the focus of visual attention was shifted under voluntary control to the valid location. Subsequent research has explored the nature of attentional orienting produced by central precues in some detail (see Cheal, Lyon, & Gottlob 1994).

Reflexive orienting, in contrast, has been studied using peripherally presented spatial cues. (e.g. see Muller & Rabbitt 1989; Cheal & Lyon 1991). In this case, the location of the target is related to the *location* of the cue, rather than to its visual form. Once again, performance can be compared across valid trials, where the target appears at the location marked by the peripheral cue, and invalid trials, where the target appears at a different location. Jonides (1981) compared the attentional effects of central and peripheral spatial cues. In the former case, a central arrow indicated the likely location of the target. In the latter case, the arrow was presented at a peripheral location, directly adjacent to the likely location of the target. Jonides' study and subsequent research has shown that orienting in response to central and peripheral precues differs in several respects. In particular, orienting in response to peripheral cues appears to have a more rapid time course, is harder to suppress (Cheal & Lyon 1991; Muller & Rabbitt 1989), and is unaffected by performing a secondary task. This evidence has supported the view that orienting in response to peripheral cues involves a rapid, reflexive process, while orienting in response to central cues involves a slower, consciously controlled process.

## 5. William James on attention

Nearly one hundred years prior to the popularity of cueing techniques as a way of studying visual attention, William James also distinguished between reflexive and voluntary orienting. For James, attention could be either "Passive, reflex, non-voluntary, effortless" or "Active and voluntary" (James 1890/1983: 394). However, James also drew two further distinctions. Firstly attention could be either "sensorial" (i.e. attention to incoming sensory information) or "intellectual" (i.e. attention to inner thoughts, images, memories etc). In addition, attention could be either "Immediate; or ... Derived: immediate when the topic or stimulus is interesting in itself, without relation to anything else; derived when it owes its interest to association with some other immediately interesting thing" (James 1890/1983: 393). Hence, for William James learning and experience played an important role in the control of attentional behaviour. Since modern studies of attention have had little to say concerning the role of learning in relation to visual orienting a programme of investigation has been undertaken here at Auckland, examining the Jamesian notion of derived attention. These experiments have incorporated the notion of derived attention into the framework of the spatial cueing technique, which was described above and which has been widely employed in studies of visual spatial attention.

## 6. Derived attention and implicit learning

The basic design framework adopted in all of the experiments to be described here is as follows. Participants performed a simple detection task which involved responding to the onset of a target stimulus, which could appear at either of two locations, on the left and right of the display. Prior to onset of the target, participants were presented with a '*cue*' stimulus. In common with many other investigations of spatial cueing, there was a predictive relationship between the nature of the cue stimulus and the location of the target. However, unlike most other studies of spatial cueing, in a number of our experiments participants were not informed of this predictive relationship. For example, in Experiment 1 of Lambert et al. (1999) the pre-trial cue comprised the letters, W and S, presented 8.5° to left and right of a central fixation cross. On some trials the letters appeared with W on the left and S on the right, while on other randomly chosen trials the letters appeared with the reverse arrangement. Unknown to the participants, the display was programmed so that the target object usually (p=.8) appeared on the same side as one of the letters – this was W for half the participants and S for the other half. The experiment had several aims. We were interested in the ability of participants to learn about the predictive relationship between cues and targets and to show evidence of this learning in orienting behaviour. The stimulus onset asynchrony (SOA) between onset of the cue letters and onset of the targets was varied in order to assess the speed with which attention was oriented in response to the predictive information carried by the letters. As indicated above, previous work had shown that speed of orienting was an important parameter in studies of the distinction between voluntary and reflexive/automatic orienting. In studies of reflexive orienting using peripheral spatial cues, clear differences between valid and invalid trials have been observed even with very brief (100ms or less) SOAs between cue and target onset. In contrast, in studies of voluntary orienting with central, symbolic cues, orienting effects tend to increase more slowly with SOA, attaining a maximum at SOAs of 300-400ms.

In addition, a crucial issue for this experiment was to investigate whether attentional learning could occur at an implicit level. Individuals were exposed to arbitrary relationships between apparently incidental events (the peripheral letters) and target events, and orienting behaviour was observed to see whether participants oriented towards the likely location as predicted by the letters. At the end of the experiment a questionnaire was administered which tested whether participants had gained any awareness of the cue-target relationship in the course of the experiment. The results showed not only that participants had absorbed the cue-target relation, but also that they oriented in response to the letters with great rapidity. A clear difference between valid and invalid trials was observed when the target was presented just 100ms after onset of the peripheral letters. In addition, it appeared that this rapid orienting occurred independently of any awareness of the cue-target relation. Fourteen out of sixteen participants reported that they were unaware of any relationship between the peripherally presented letters and the location of the target. When asked to choose between two statements, one of which described the correct cue-target relationship, while the other described the reverse of the correct relation, about 50% (7/16) chose the *incorrect* statement. The performance of these participants was closely similar to that of the whole group.

It was concluded that these results provide clear support for the Jamesian notion of 'derived attention'. That is, after a relatively brief practice period participants became attuned to the predictive relationships present in the task, and oriented rapidly in response to the letters. Furthermore, our data showed that this learning can proceed at an entirely implicit level. Having established that attentional behaviour is sensitive to predictive relations between target location and letter stimuli, we were interested in exploring this sensitivity in relation to other features of the incidental '*cue*' stimulus. An obvious further candidate is stimulus colour. Lambert and Roser (2001) examined this question using a design that was analogous to the letter cueing experiments just described. In this study the two target locations on the left and right of the display were each surrounded by a square grey frame. The cue stimuli comprised the transient onset of a thin green square and a thin blue-green square which appeared against the grey background of the surrounding frames. The relative location of these colour cues (i.e. $\text{green}_{\text{left}}$ + $\text{blue-green}_{\text{right}}$ vs. $\text{blue-green}_{\text{left}}$ + $\text{green}_{\text{right}}$) varied randomly from trial to trial. Participants performed two 'training blocks' of trials in which the target always appeared on the same side as one of the peripheral cues – for half the participants this was the green cue, for the other half this was the blue-green cue. In the final block of trials there was no relation between target location and the peripheral colour cues. The notion of derived attention was tested by comparing detection performance for trials where the target appeared next to the colour that had been associated with target location during the training blocks, with performance on trials where the target appeared on the opposite side. In this study the proportion of participants who gained explicit awareness of the relation between cue colour and target location was greater (7/30) than had been the case with letter cues. Indeed, in pilot work we had discovered that when the two peripheral cue colours were clearly distinct (e.g. red vs. green), a very high proportion of participants gained explicit awareness of the relation between target location and the cues. Clearly, this would rule out an interpretation of performance purely in terms of implicit learning. In order to render the cue-target relation less salient, and hence less accessible to explicit awareness, we chose cue colours that were subjectively similar – green and blue-green. Under these conditions 23% of participants gained awareness of the cue-target relation, and these participants showed faster response times for valid than for invalid trials. However, 40% of participants failed to gain any awareness of the cue-target, and these participants also showed significantly faster response times for valid compared to invalid trials. For both groups a validity effect was apparent at both short (100ms) and long (600ms) SOAs. Questionnaire responses from the remaining participants (37%) provided equivocal evidence concerning awareness of the cue-target relation. Interestingly, orienting effects were absent for this *Semi-Aware* group. This was interpreted in terms of the formation and testing of *incorrect* hypotheses concerning the cue-target relation, leading to inappropriate orienting towards the invalid location on at least some trials by these participants.

The experiments just described establish that learning takes place when individuals are exposed to associations between target events and apparently incidental 'cue' events. At a theoretical level, it is clear that in order to explain these effects one need only postulate a relatively simple form of associative learning. The learning that is involved may be seen as simple in two respects – it involves sensitivity to simple binary associations between cue attributes and target attributes; and secondly the attributes manipulated in the experiments can be seen as relatively 'simple'. Encoding of the visual form features required to distinguish 'W' from 'S', encoding of stimulus colour and encoding of target location are all thought to occur at a relatively early stage of visual processing. Despite this simplicity, these findings are of theoretical interest for several reasons. Firstly, as already discussed, the fact that *attentional* responses, both covert and overt are open to learning effects is of interest, and is consistent with the Jamesian notion of derived attention. That is, after a brief period of task performance, the cue events come to attract attention by virtue of their association with target location. Secondly, the results show that this form of attentional learning can proceed at an implicit level.

Having established these points, a logical next step would be to explore the limits of derived peripheral cueing by increasing task complexity. One way of doing this is to increase the complexity of the cue attributes that are associated with target location. For example, would implicitly learned derived cueing effects be observed if participants were exposed to an association between target location and *semantic* attributes of cue stimuli? Curiously, we performed a series of experiments testing this hypothesis before undertaking the letter and colour cueing described above. As Antipodeans it might be said that we are used to doing everything upside-down, including experimental research. The real reason was that the rationale for our semantic cueing experiments was derived from thinking about some earlier findings (Lambert, Beard, & Thompson 1988) demonstrating semantic processing of unattended sources of information. In light of these results, Lambert and Sumich (1996) aimed to explore a possible functional role for semantic processing of incidental information, by testing for a link between such processing and attention movements. This was done by using three and four letter words as the 'cue' stimuli. As in the other experiments, participants made a simple detection response to targets that could appear on the left or right of a display screen. Six hundred milliseconds prior to target onset, a single word was presented for 67ms either 5° to the left or 5° to the right of a central fixation cross. Participants were instructed to ignore this word, and to concentrate on the target detection task. Sixty different cue words were employed in the experiments, of which half referred to living things (e.g. DOG) while the remainder referred to non-living objects (e.g. HAT). For half the participants, the display was programmed so that the target usually appeared on the same side as the cue word if this referred to something living, and on the opposite side if it referred to something non-living. The reverse contingency was present for the remaining participants. The hypothesis that visual orienting behaviour would be sensitive to an association between target location and a semantic attribute of the cue word was tested by comparing performance on trials where the target appeared at the likely location, as indicated by 'cue' word category, with performance on trials where the target appeared at the unlikely location. Lambert and Sumich (1996) reported three experiments which all shared this basic design. In all three experiments participants responded more rapidly to targets that appeared at the likely location, as indicated by the semantic category of the ostensibly unattended 'cue' words. These findings suggested that the human attentional system is indeed exquisitely sensitive to contingent relations between the spatial location of target events and other attributes of the visual environment. It is worth noting that the design of these experiments provided an extreme test of the hypothesis that visual attention would be sensitive to associations involving semantic information. As indicated earlier, at the beginning of the experiment participants were instructed to ignore the laterally presented words, which were drawn from a large (120 item) stimulus set. The words themselves varied with respect to a wide variety of stimulus dimensions – number of letters, word frequency, number of syllables, imageability, meaningfulness are just a few of these dimensions. The target location – semantic category relation was of course entirely arbitrary, and participants had no way of knowing that the living – nonliving dimension was at all relevant to the task. Questionnaire responses certainly indicated that only a tiny minority (three out of 120 participants) showed any evidence of explicit awareness regarding the link between target location and word category. Despite the obscurity of this relationship, analyses of response times indicated a significant advantage for trials where the target appeared at the likely location. Hence, the cue-target relation had indeed been learned, resulting in a small but measurable effect on attentional behaviour. This effect appeared entirely independent of conscious awareness concerning the cue-target relation. Given the subtlety of this relationship, it was concluded that this performance represented a remarkable feat of sensitivity by the human perceptual-attentional system.

## 7. Voluntary derived attention

In our initial investigations of derived attention, participants were never informed of the cue-target relation, and our post-experiment questionnaires were designed with care, in order to check for the presence of even fragmentary knowledge of the relevant relationship. This was perceived as important, since we surmised that if participants were aware of the cue-target relation then any attentional effects could be explained in terms of the familiar construct of voluntary orienting. In this case the findings would of course be lacking in theoretical interest. However, an interesting feature of the results reported by Lambert et al. (1999) was the *speed* with which differences between valid and invalid trials emerged following onset of the cue. An advantage for valid over invalid trials was observed when the SOA was very brief – just 100ms. As noted earlier, the time-course of voluntary orienting effects is typically slower than for automatic-reflexive orienting – taking some 300-400ms to fully develop, in comparison with a time-course of 50-100ms for reflexive orienting. This led to the following thought experiment. If the design of the experiment were altered so that participants were rendered aware of the cue-target relation, and if a similar pattern of rapid orienting were then to be observed, it is uncertain whether an explanation in terms of *voluntary orienting* as traditionally conceived (e.g. see Posner & Raichle 1994) would be appropriate. This led to a deeper consideration of the distinction between voluntary and reflexive orienting and a new set of experiments that may be viewed both as a critical exploration between the distinction between voluntary and reflexive orienting *and* as investigations of the Jamesian notion of *voluntary derived attention.*

Lambert and Duddy (2002) noted that the notions of voluntary and reflexive orienting have generally been associated with the effects of central and peripheral cues respectively. That is, the rapid and relatively automatic effects of peripheral cues have been attributed to a hypothetical process termed reflexive orienting, while the slower and ostensibly controlled effects of central, symbolic cues have been attributed to a process referred to as voluntary orienting. However, this analysis is problematic, since overall comparisons of effects of central and peripheral cues have tended to confound three distinct factors. Lambert and Duddy (2002) pointed out that the design of the central cueing task and the peripheral cueing task differ in at least three respects. Firstly, the two designs differ eponymously and by definition with respect to the visual eccentricity of the cue – central and peripheral. A second difference is the need to discriminate between different cue stimuli. In the case of central cues, participants must discriminate between different cues, such as left and right facing arrows. In the case of peripheral cueing, discrimination between different cue stimuli is not required since orienting occurs in response to the mere onset of a cue at a particular spatial location. In addition Lambert and Duddy (2002) identified *spatial correspondence* as a third factor which covaries with cue eccentricity and cue discrimination in commonly employed versions of the central and peripheral cueing task. In the peripheral cueing task spatial correspondence is an important factor in the following sense. On valid trials the location of the cue stimulus and the location of the target stimulus correspond (e.g. both left, both right), whereas on invalid trials cue and target locations do not correspond. This factor of spatial correspondence is absent in the central cueing task, since there is no relationship between the location of the cue stimulus, which is always central, and the location of the target stimulus.

This analysis of course raises the question as to which of the three factors is responsible for the differences in performance observed in earlier studies of orienting in response to central and peripheral cues. Lambert and Duddy (2002) reported a series of five experiments designed to answer this question. In all five experiments participants were presented with bilateral letter cues prior to making a simple detection response to target stimuli that could appear on either the left or right of a display screen. In Experiments 1, 3A, 4 and 5 the letter cues were presented either at a peripheral location, or centrally (immediately to the left and right of a central marker). The cues were always either X on the left and T on the right, or vice versa. Participants were informed that the display was programmed so that target stimuli usually appeared on the same side as one of the letters – for half the participants this was 'X', for the other half it was 'T' – and were instructed to try to use this information in order to pay attention to the likely location, on the basis of the letter cues. Thus, for these participants, a valid trial occurred when the target appeared on the same side as 'X' and opposite to 'T'; an invalid trial occurred when the target appeared on the same side as 'T' and opposite to 'X'. In terms of our three factor analysis of orienting in response to spatial precues, three design features of these experiments are worthy of note: (1) The cue eccentricity factor was manipulated – on some trials the letter cues were central, while on others they were peripheral; (2) The cue discrimination factor was held constant – on all trials participants needed to discriminate which letter was 'X' and which 'T' in order to orient appropriately; (3) The spatial correspondence factor was present and also held constant – valid and invalid trials were defined in terms of a spatial correspondence between the side of presentation of a cue letter (X or T) and likely target location. A consideration of these design features led in turn to three distinct predictions concerning expected patterns of results. All five experiments incorporated a range of SOAs between cue onset and target onset, so that speed of orienting in response to the letter cues could be assessed. If the differences in speed of orienting observed in earlier studies of spatial cueing were driven by the cue eccentricity factor, then one should observe *rapid* orienting in the *peripheral* cue condition and relatively *slow* orienting in the *central* cue condition. That is, clear differences between valid and invalid trials should be apparent even with very brief SOAs in the peripheral condition, while in the central cue condition valid-invalid differences may only become apparent with somewhat longer SOAs. If on the other hand, differences in speed of orienting observed in earlier work have been driven by the need to discriminate between different cue stimuli, then one would expect to observe relatively *slow* orienting in *all conditions* – since participants always needed to discriminate which cue letter was X and which T in order to orient appropriately. And finally, if previously observed differences in speed of orienting have been driven by the spatial correspondence factor, then *rapid orienting* should be observed in *all conditions.* A straw poll of colleagues and other experts in the field of visual attention suggested that smart money was on the second prediction – which entailed the plausible supposition that slow orienting would be observed because in this situation accurate orienting is contingent upon a time consuming letter discrimination process. The results were clear-cut, and showed that in this instance the smart money was mistaken. Rapid orienting was observed in all conditions. Indeed, in Experiment 3A a clear advantage for valid over invalid trials was observed even with extremely brief SOAs between cue onset and target onset (0ms, 33ms, 66ms). Furthermore, the time-course of orienting effects was closely similar for central and peripheral letter cues. These results suggested that the need to discriminate between different cue stimuli is *not* the crucial feature responsible for the relatively slow orienting observed in earlier studies of orienting in response to central, symbolic cues. It also suggested that cue eccentricity was not responsible for the varying time-course of orienting observed in many earlier studies (Cheal & Lyon 1991; Muller & Rabbitt 1989; but see also Warner, Juola, & Koshino 1990 for findings consistent with the Lambert & Duddy 2002 data). Accordingly the results were interpreted as support for the hypothesis that cue-target spatial correspondence is a crucial factor mediating speed of orienting in response to precues. However, in a sense this hypothesis was supported merely by default – since the evidence was inconsistent with the other two hypotheses. The aim of Experiments 2 and 3B was to test the spatial correspondence hypothesis in a more direct and active way. This was done by examining the time-course of orienting in a design that was identical to that employed in Experiments 1, 3A 4 and 5 in all respects, save for the presence of cue-target spatial correspondence. Participants were presented with bilateral letter cues, as before. However, in these experiments the letter cues were always X on the left accompanied by X on the right, or $T_{left}$ accompanied by $T_{right}$. Half the participants were instructed that T cues indicated that the next target would probably appear on the right, while X cues indicated that the next target would probably appear on the left. The other participants were informed that T cues predicted a left target, while X cues predicted a right target. The perfectly symmetrical nature of the cue displays (i.e. X + X ; T + T) ensured that there was no spatial correspondence between the cue and target stimuli. Accordingly, this was termed the *spatial translation* condition, since encoding the cue information required translating a symmetrical stimulus into a representation embodying the spatial information necessary to support orienting towards one side of the display. In agreement with the spatial correspondence hypothesis, orienting effects were observed to be significantly weaker, and to have a slower time course in the spatial translation conditions of Experiments 2 and 3B, in comparison with the spatial correspondence conditions of Experiments 1 and 3A.

Overall, these results suggest two main conclusions. Firstly, they show that spatial correspondence is a crucial factor mediating speed and efficiency of orienting in response to spatial precues. When this factor was present, clear differences between valid and invalid trials were observed, even at the briefest SOAs between cue and target onset. When this factor was absent, orienting effects were significantly weaker, and were present only at somewhat longer SOAs. In addition, the results suggest that the process of derived attention, by which cue stimuli come to attract attention by virtue of their association with target events, can operate at both an implicit level, and via explicit-voluntary control. Indeed, at a gross level the overall magnitude of the orienting effects observed under explicit instructional conditions (Lambert & Duddy 2002) appeared to be stronger (30-40ms) than the effects observed by Lambert et al. (1999) under implicit conditions (10-20ms).

## 8. Concluding discussion

This chapter began by reminding readers of the extreme selectivity embodied in the architecture of the visual system. However, it appears that the strong degree of selectivity, apparent from anatomical observation and measured psychophysical performance is not reflected in the texture of visual awareness. Findings were reported from a questionnaire that attempted to probe the phenomenology of the distinction between central and peripheral vision. Results showed that for most individuals the subjective extent of the "area of perfectly clear central vision" was considerably larger than the measured extent of the fovea. The subjective extent of this 'phenomenal fovea' appeared to be rather larger when participants imagined a typical viewing situation, in comparison with a situation where participants viewed a real stimulus. It was suggested that at least part of this disparity between the relatively tiny psychophysical/anatomical fovea and what I have termed the phenomenal fovea may arise from the fact that vision is not only extremely selective, but also highly dynamic. Under normal viewing conditions visual attention can dance lightly and rapidly from feature to feature – and this may contribute to the impression of an expanse of detailed vision that is much larger than the 1–2° foveal diameter that emerges from psychophysical measurement under closely controlled conditions. The strong selectivity apparent at the earliest stage of visual processing, in the retina, may be viewed as an essentially attentional phenomenon, since it is thought to arise as a result of resource limitations in the computation of high level perceptual descriptions of visible objects. Hence in this very basic sense, seeing (i.e. visual awareness) can be viewed as an attentional process.

It is apparent from this analysis that in order to operate effectively, a strongly selective visual system requires mechanisms that enable attention to be shifted rapidly and accurately between task relevant regions of a visual scene. The studies reviewed in this chapter show that learning and experience play a pivotal role in enabling rapid shifts of attention towards the likely locations of target objects. The notion of derived attention, originally proposed by William James, was used to describe the propensity of cue stimuli to capture attention, by virtue of learned associations between cue attributes and target location. Experimental evidence was reviewed showing that these associations may be learned either at an implicit level, via mere exposure to the cue-target contingencies, or via an explicit instructional set. In the former case, it appears that even highly abstract attributes, such as the semantic category of a cue word can influence attention deployment, following exposure to a specific cue-target contingency. An interesting feature of the derived attentional cueing effects reported by Lambert and Duddy (2002) was the rapidity with which attention appeared to be shifted. Even with extremely brief SOAs between cue onset and target onset (0ms – 66ms), response times were quicker for targets at the likely compared to unlikely location.

As William James realised, attention and conscious awareness are intimately related concepts. A primary consequence of selective attention is that certain objects gain clearer representation in conscious awareness. Hence, the attentional learning process proposed above may be seen as regulating and optimising the sequence of objects and events that enter conscious awareness during performance of a complex task. It is perhaps not surprising then that this learning process can proceed at an implicit (non-conscious) level. If the function of the process is to improve the efficency with which successive perceptual objects enter consciousness and are acted upon, it would appear theoretically incoherent if the process itself were to require representation in consciousness. Nevertheless it is clear that derived attention effects can be observed, and may even be enhanced with an explicit instructional set: i.e. telling participants that targets are likely to occur near a particular form of cue. This is not necessarily inconsistent with the previous point. It is clear that conscious decisions may often result in a cascade of lower level processes that are not themselves represented in consciousness. For example, after making a conscious decision to re-read the previous paragraph, I am aware of successive words and phrases as I read, but not of the lower level processes that control the sequence of saccades that occur as my eyes move across the page. When our participants were informed explicitly that targets are likely to occur near a particular cue letter, attentional performance readily reflected this knowledge, but in debriefing sessions participants often claimed to be unaware of using knowledge of the cues to direct attention – especially on trials where the SOA between the letter cues was very brief.

The processes underpinning derived spatial cueing effects may have broad relevance to perception and action in everyday life. They may play an important role in the performance improvements that accompany repeated practice on almost any perceptuomotor task, from operating a word processor to playing tennis, to monitoring an air traffic control display. With repeated performance individuals will be exposed to a variety of associations between different elements of a task. As a result of exposure to these associations a form of attentional learning may occur, which enables participants to orient rapidly and efficiently to regions of the task scene that are most likely to contain task relevant objects. William James termed this form of learning 'derived attention'.

## Notes

* The author would like to acknowledge the support of Auckland University Research Committee. I would also like to thank Caroline Heffer for assistance with preparing, administering and then analysing the visual awareness questionnaire described in this chapter.

**1.** These figures hold true under either of two assumptions: (a) that the 'very large poster' is a flat square, subtending 135° horizontally x 135° vertically at the eye, or (b) that the poster is a spherical cap with a diameter of 135°.

**2.** The first picture was from the popular '*Where's Wally*' by Martin Handford (this may be known as '*Where's Waldo*' to North American readers) – in which the child searches for Wally in an extremely crowded scene containing many Wally look-alike figures (Handford 1987). The second picture was a photograph of the Southern Alps of New Zealand, and was taken from Forman (1999).

## References

Allport, A. (1989). Visual attention. In M.Posner (Ed.), *Foundations of Cognitive Science* (pp. 631–682). Boston, MA: MIT Press.

Barry, R. J. & Sokolov, E. N. (1993). Habituation of phasic and tonic components of the orienting reflex. *International Journal of Psychophysiology, 15* (1), 39–42.

Broadbent, D. E. (1958). *Perception and Communication.* London: Pergamon Press.

Cave, K. (2001). Selection can be performed effectively without temporal binding, but could be even more effective with it. *Visual Cognition, 8,* 467–487.

Chapanis, A. (1949). How we see: A summary of basic principles. In *Human Factors in Underseas Warfare.* Washington, D.C.: National Research Council.

Cheal, M. & Lyon, D. (1991). Central and peripheral precuing of forced choice discrimination. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 43A,* 859–880.

Cheal, M., Lyon, D., & Gottlob, L. (1994). A framework for understanding the allocation of attention in location precued discrimination. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47A,* 699–739.

Chun, M. & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71.

Chun, M. & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10,* 360–365.

Curcio, C., Sloan, K., Kalina, R., & Hendrickson, A. (1990). Human photoreceptor topography. *Journal of Comparative Neurology, 292,* 497–523.

James, W. (1983). *The Principles of Psychology.* Cambridge, Mass.: Harvard University Press. (Original work published 1890).

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye movement. In J. Long and A. Baddeley (Eds.), *Attention and Performance IX* (pp. 187–203). Hillsdale, N.J.: Erlbaum.

Lambert, A. J. (1985). Selectivity and stages of processing: an enduring controversy in attentional theory. *Current Psychological Research and Reviews, 4*, 239–256.

Lambert, A. J., Beard, C. T., & Thompson, R. J. (1988). Selective attention, visual laterality, awareness and perceiving the meaning of parafoveally presented words. *Quarterly Journal of Experimental Psychology, 40A*, 615–652.

Lambert, A. J. & Duddy, M. (2002). Visual orienting with central and peripheral precues: Deconfounding the contributions of cue eccentricity, cue discrimination and spatial correspondence. *Visual Cognition,* in press.

Lambert, A. & Roser, M. (2001). Effects of bilateral colour cues on visual orienting: Revisiting William James' 'Derived attention'. *New Zealand Journal of Psychology, 30,* 15–22.

Lambert, A. J. & Sumich, A. L. (1996). Spatial orienting controlled without awareness: A semantically based implicit learning effect *Quarterly Journal of Experimental Psychology, 49A,* 490–518.

Lambert, A., Naikar, N., McLachlan, K., & Aitken, V. (1999) A new component of visual orienting: implicit effects of peripheral information and sub-threshold cues on covert attention. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 321–340.

Muller, H. & Rabbitt, P. (1989). Reflexive and voluntary orienting of visual attention: Time course of Activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315–330.

Osterberg, G. A. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica, 13* (Suppl. 6), 1–97.

Phillips, J. G. & Triggs, T. J. (2001). Whither automaticity and human performance? In Columbus, Frank (Ed). (2001). *Advances in psychology research, Vol. 4.* (pp. 151-173). Huntington, NY, US: Nova Science Publishers, Inc.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 32A,* 3–25.

Posner, M., Nissen, M. J., & Ogden, W. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. Pick and I. Saltzman (Eds.), *Modes of Perceiving and Processing Information* (pp. 137–157). Hillsdale, N.J.: Erlbaum.

Posner, M. I & Raichle, M. (1994) *Images of Mind.* New York: Scientific American Library.

Ramachandran, V. S. & Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature, 350,* 699–702.

Schiffman, H. R. (2000). *Sensation and Perception.* New York: John Wiley.

Shiffrin, R. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190.

Wickens, C., Gordon, S., & Liu, Y. (1998). *An introduction to human factors engineering*. New York : Longman Press.

Warner, C. B., Juola, J.F., & Koshino, H. (1990). Voluntary allocation versus automatic capture of visual attention. *Perception & Psychophysics, 48,* 243–251.

Wright, R. (1998). *Visual Attention.* Oxford: Oxford University Press.

# Contextual cueing

## Reciprocal influences between attention and implicit learning

Yuhong Jiang and Marvin M. Chun
Massachusetts Institute of Technology / Vanderbilt University

Consciousness and its counterpart – unconsciousness – are arguably the most fascinating aspects of the human mind. They are also elusive research topics, as determining whether something is or is not within the realm of consciousness typically relies on the subject's verbal report. This makes it difficult to tell whether infants and non-human animals have consciousness, and whether human adults' report of consciousness is reliable. Although consciousness is an intriguing property, research in the past twenty years has accumulated more and more evidence that a large part of human cognition is implicit, occurring outside of awareness (Reber 1989; Stadler & Frensch 1998).

Implicit, unconscious cognition can be dissociated from explicit, conscious cognition at both neural and functional levels. Although it is unlikely that all forms of implicit cognition (e.g., motor sequence learning and artificial grammar learning) rely on the same neural substrates, implicit and explicit processes typically differ in their neural basis. For example, damage to the medial temporal system (including the hippocampus) produces amnesia, an inability to form new long-term conscious memory (Milner, Squire, & Kandel 1998), but the same patient retains his capacity of motor skill learning, a form of implicit learning and memory. As a complementary example, damage to the basal ganglia as in Parkinson's disease impairs the patient's ability to acquire motor skills (e.g., Doyon et al. 1997), yet the same patient typically has largely preserved explicit learning and memory. It is debatable whether implicit learning is a phylogenetically primitive system that relies less on age, IQ, and individual difference (Reber, Walkenfeld, & Hernstadt 1991; but see Fletcher, Maybery, &

Bennett 2000). Nonetheless the existence of a system (or a myriad of systems) of implicit cognition is well established by now.

The concept of attention enters the picture of implicit learning for two closely related reasons. Theoretically, "attention" and "consciousness" are related concepts. They may be superficially interchangeable: not paying attention results in the lack of awareness of ongoing events (Mack & Rock 1998); and attending to something almost immediately brings it to awareness and leads to successful memory recall in the future (Rock & Gutman 1981). Empirically, there are data supporting the notion that whereas explicit learning relies on a controlled, attentive process, implicit learning can largely operate automatically, without any reliance on attentional resources (e.g., Frensch, Lin, & Buchner 1998; Jiménez & Méndez 1999; Stadler 1995). The argument that attention and consciousness are interchangeable, therefore, is not without ground. To accept this argument has far reaching consequences. For instance, one could use an attentional modulation effect as an indicator of awareness. Attentional modulation of neurons at numerous neural stages of visual processing (e.g., V1, V2, V4, MT, IT; see Luck et al. 1997; Moran & Desimone 1985) may be explained as neural correlates of the animal's awareness. In addition, if implicit learning requires no attention, our educational system could in principle be revised to promote implicit learning without overloading the learner's limited attentional capacity.

However, our premise is that to the contrary, attention and consciousness are not one and the same. Implicit learning, devoid of any conscious attempt to learn and ability to recall, may nevertheless penetrate attentional mechanisms by shaping how attention is deployed (see also Lambert, this book; Lambert, Naikaar, McLachlan, & Aitken 1999). In addition, implicit learning, requiring no awareness, may nevertheless depend on selective attention. We will rely on a new implicit learning paradigm, spatial contextual cueing (Chun & Jiang 1998), to make this point. Following the introduction of contextual cueing, we will describe how implicit learning guides attention to behaviorally important aspects of visual input. Then we will focus on how selective attention determines what information can be implicitly learned.

## 1. Contextual cueing

The term contextual cueing defines two important aspects of this paradigm. "Contextual" refers to the impact of other information, typically co-occurring items, on the processing of the target. In our studies of implicit learning, such context acquires predictive power from the statistics of our displays experienced over time. “Cueing” refers to the top-down guidance of attention. Contextual cueing, therefore, refers to the top-down influence of attention by predictive contextual information that was implicitly learned.

In our own research, we have studied the cueing of attention to a target’s spatial location, shape (identity), motion trajectory, or the temporal moment at which it occurs. Across different experiments, we have defined context by the spatial configuration of the target and accompanying distractors, the shape (identity) of the distractors, the distractor motion trajectory, or the temporal sequence (duration, tempo, or order) of the distractors. In all these studies, target processing was facilitated when the target was consistently paired with a certain context of distractors that was repeatedly presented to the observer (Chun & Jiang 1998, 1999; Olson & Chun 2001).

For example, in the spatial contextual cueing paradigm (Chun & Jiang 1998), we co-varied the spatial location of the target and the spatial layout or configuration of the target and distractors. To reduce the likelihood of explicit learning and memory, we disguised the learning component by emphasizing the visual search task only. Our subjects were instructed to search for a target – a *T* rotated either to the right or to the left – among distractors – rotated *L*s. The items were presented at randomly selected positions on the computer screen. Typically, 12 items were on the display. Because the distractors and the target are similar in shape, search relies on focused attention that scans through the items slowly. Any predictive cue to the target location, therefore, can potentially enhance search speed. Our subjects were merely instructed to perform the search task as quickly and as accurately as possible, without any hint that some of the spatial configurations would repeat.

Target locations were consistent within repeated spatial configurations. Thus, spatial context provided a predictive cue to the target location. Below, we will detail the specific procedure because it is similar across many of our contextual cueing experiments. In a typical study, we divided the entire test session into 30 blocks, and each block contained 12 Old and 12 New displays intermixed with each other. The set of Old displays was 12 different spatial configurations of distractors that were generated at the beginning of the session and repeated across blocks throughout the session. Each configuration was associated with a different target location, which was fixed, so that the spatial context (configuration) would be predictive of the target as subjects experienced the repeated displays. The New displays were randomly generated for each block to serve as a baseline. However, the same set of target locations was used for the New displays throughout the entire experiment to equate the

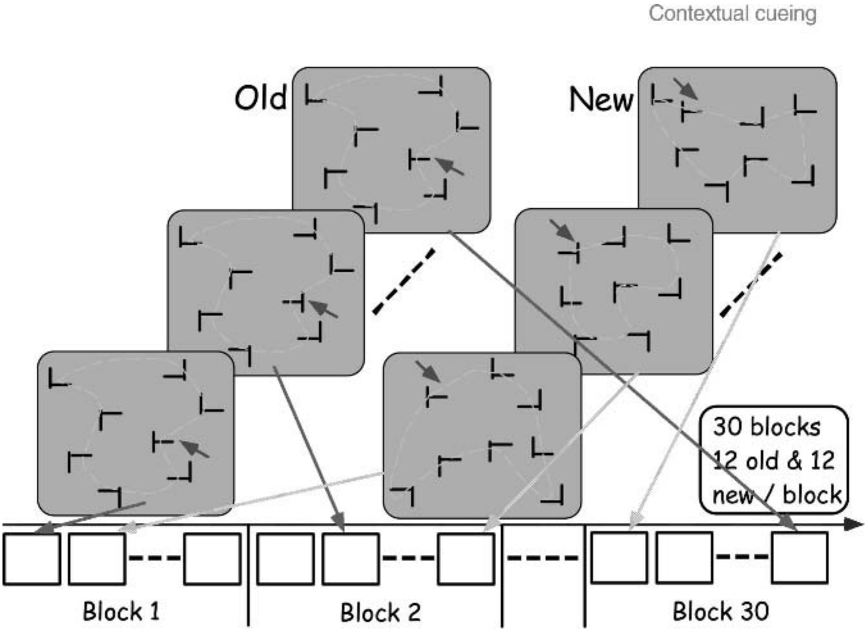


**Figure 1.** A schematic illustration of the contextual cueing paradigm. In the old condition, a configuration was repeated across blocks. The target identity and its associated motor response were chosen at random, but the target location and its associated configuration were repeated. In the new condition, the target location was repeated across blocks, but was presented among a variable configuration for each repetition. The dotted outline of display layout shows the imaginary configuration for the target. It was not actually presented on the display.

target location probabilities between Old and New conditions. To sum, Old displays presented spatial contexts that were predictive of target location, while the New displays were not predictive. Figure 1 shows a schematic illustration of the design.

To search through the Old displays faster than the New displays, observers must first be able to discriminate the old configurations (typically 12 of them) from a total of nearly 400 different new configurations. This is not an easy task, given that virtually an infinite number of different configurations can be generated, and that any given configuration is not highly discriminable from any others. In addition, observers need to access the associated target location within a particular configuration. Finally, the extraction of the configuration and its associated target location has to be fast enough to be useful for visual

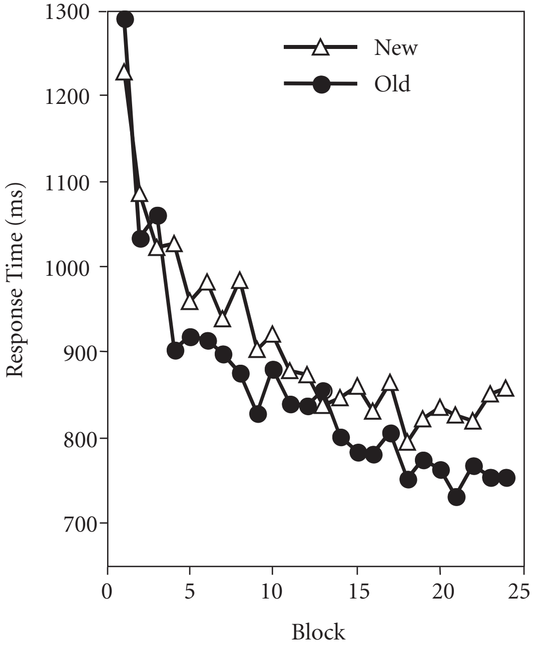


**Figure 2.** Search response time is faster for targets appear in Old vs. New displays. Contextual cueing is measured by the difference between the Old and New conditions.

search. That is, this process needs to occur within a second or so, given that search is completed in about 1 second even for novel displays.

Given the complexity of the nature of learning, it is indeed remarkable that observers searched through the Old displays faster than the New displays (Figure 2). The significant enhancement typically occurred after merely five or six repetitions, revealing an exceptional learning ability to quickly make use of invariant information learned shortly earlier. We believe that observers learned each instance of the repeated configurations and discriminated them from novel layouts (Logan 1988; Palmeri 1997). Once a display was presented, observers quickly extracted the configuration and used the association between a configuration and the target location to guide attention to the target location. Contextual cueing, therefore, reveals how learning can contribute to the top-down guidance of visual attention.

The guidance of top-down attention by contextual cueing has important theoretical implications for studies of attention. Although prior research has shown that many factors affect how we allocate top-down attention (Wolfe 1994), such research has typically relied on explicit instruction to focus on task-relevant aspects of search arrays (e.g., "search through the red items"). Prior work has ignored the role of subjects' past experience. For example, observers may be pre-cued to attend to one part of the visual field more than other parts (Posner, Snyder, & Davidson 1980), or to attend to one object more than another (Pylyshyn & Storm 1988). No doubt, human observers can follow such instructions well, but in our daily visual environment, such top-down instruction is not always available. In contrast, experience is always available. As we navigate through an environment over time, what was important in the past tends to remain important in the present. Learning the invariant features and properties of our visual environment can increase the efficiency of visual processing (Gibson 1966). Contextual cueing reveals the power of past learning to guide top-down attention. Learning is quick; a few repetitions of the display is sufficient to produce a significant speed difference. It is powerful, as complex and highly similar visual displays can be distinguished from one another. It is efficient and robust, as it can speed up search by 100 ms or so and is consistently shown across subjects. Moreover, contextual cueing is implicit. In the following section, we review the implicit nature of contextual cueing and its other properties.

## 2. Implicit learning guides top-down attention

### 2.1 What is learned in contextual cueing?

Before we assess the implicit nature of contextual cueing, we need to characterize what information is learned during contextual cueing. Only then can we probe whether subjects had awareness of such information. Research on other paradigms of implicit learning has taught us that it is not easy to pinpoint exactly what aspect of the task is learned. In the serial reaction time (SRT) task, for instance, observers respond more quickly to a repeated series than to a random series of target positions (Nissen & Bullemer 1987). Such advantage in RT, however, may come from several sources: learning of the visual repetition at the perceptual level, learning of the pattern of repeated movement at the motor level, or learning of some abstract representation of the repetition (Keele, Jennings, Jones, Caulton, & Cohen 1995; Mayr 1996; Willingham 1999). Even in a task that is devoid of a motor component such as the artificial grammar (AG) task, a question remains as to what portion of the grammar is learned. Is it the entire abstract grammar, or is it just particular sets of fragments that are learned? Similar questions can be asked about contextual cueing. In the cascade from perception to motor response, at which stage does learning occur? If it is at the perceptual stage, do observers learn the entire configuration or just portions of it?

In the spatial contextual cueing paradigm described above, we can quickly rule out learning at the motor response stage (left or right key press), and we can rule out learning of the target identity (left or right *T*). This is because the identity of the target and its associated key press was randomized from block to block, for both New and Old conditions. Eye movement learning may be part of what is learned, but eliminating eye movement by using brief displays does not eliminate contextual cueing, suggesting a purely perceptual component of learning (Chun & Jiang 1998). Confining learning to the perceptual end, we can also rule out probability learning of target locations. This is because target locations for both Old and New conditions were repeated the same number of times over the entire experiment.

Why do observers search through the Old displays faster than the New displays, if they did not learn the target identity and its associated motor response, or the target location probability? One hypothesis is that maybe the Old displays become perceptually familiar; and observers simply search through a familiar display faster than through an unfamiliar one. To test this hypothesis, we modified the Old displays such that the target was allowed to move within the configuration across blocks (i.e., the target and a randomly selected distractor swapped positions from repetition to repetition). In this design, the layout through which search was performed was repeated and was thus familiar to subjects, but the context no longer predicted target location. If familiarity with a search configuration underlies the original contextual cueing effect, we should expect the same amount of facilitation whether or not the target location was consistently paired with the configuration. However, our results showed that contextual cueing was abolished when the configuration no longer predicted the target location (Chun & Jiang 1998). Thus, familiarity with a configuration is insufficient for search speed to improve. This result converges nicely with Wolfe and colleagues' finding that search through repeated layouts was not faster than through new layouts (Wolfe, Klempen, & Dahlen 2000). In a sense, the visual system is amnesic to the repetition of the layout. Repetition of the configuration per se is not useful for visual search. Only when a repeated layout is consistently paired with a target location can it be informative of search.

Thus, learning in the spatial contextual cueing paradigm involves not only the discrimination among highly similar configurations, but also the consistent pairing between the target location and a configuration.

## 2.2 What is learned: The entire configuration or portions of it?

During spatial contextual learning, did subjects learn the entire global configuration, or just portions of it? Several experiments have been conducted to characterize what aspects of "spatial configuration" are learned in spatial contextual cueing.

In one experiment, we perturbed the configurations. In our earlier studies, the distractors were repeated in their exact same locations across Old display repetitions. However in a separate study, distractors were allowed to jitter within about 0.5 degrees visual angle, so that each exact location of the item changed and the entire configuration changed somewhat (Chun & Jiang 1998). The resulting distorted configurations were different from repetition to repetition, yet they still resembled the prototype from which the distorted versions were generated (Palmeri 1997). Results showed that observers still searched faster in this perturbed Old condition than the New condition. This finding shows that invariant contextual information can be extracted and learned from noisy input.

In a second study, Olson and Chun (2002) studied whether different aspects of spatial context may be weighted differently. They divided the display into two halves: left and right in one experiment, and up and down in another. Distractors positioned at the same half as the target were on average closer to the target, and thus formed the short-range context. Distractors at the opposite half formed the long-range context. The separation of the short- and long-range contexts permits independent manipulation of these contexts. Observers were tested in four conditions: both short- and long-range contexts were repeated; only short-range context was repeated; only long-range context was repeated; and neither short- nor long-context was repeated. Results showed similar contextual cueing in both-old and short-range old conditions, but no learning in both-new and long-range old conditions. The lack of contextual cueing in the long-range old condition is surprising, showing that the implicit learning mechanism may fail to encode invariant information under some circumstances, even when the repeated information in the long-range context was just as predictive of target location as the short-range context was.

To find out whether the lack of benefit in the long-range condition was due to a failure in learning invariant information positioned far away, or due to disruption from random information close by, Olson and Chun (2002) manipulated the objects that were spatially positioned between the target and the long-range context. Items were inserted between the target and the long-range context in one but not another condition. Olson and Chun then repeated the long-range but not the short-range context. Contextual cueing was found only when no items were positioned in the space between the target and the long-range context. This suggests that the lack of learning found before in the long-range context condition was not because the absolute distance between the target and the long-range context was too far. Instead, random items close to the target interfered with learning of the long-range context. It is unclear, however, whether the short-range items disrupted the learning of long-range items, or whether they simply prevented the learned long-range context from influencing search speed. To find out whether short-range items prevented learning or expression of the long-range context, one needs to remove the short-range context during testing (for an analogous distinction in the SRT task, see Frensch, Wenke, & Runger 1999).

Finally, one can explore whether spatial learning extends to 3-D displays. The real spatial environment that we live in is three-dimensional, but the spatial context tested so far has been two-dimensional. Recently Kawahara asked whether the visual system was sensitive to the depth relations in 3-D spatial context (Kawahara, personal communication). To find out, Kawahara segregated the items into two planes using binocular disparity. Some items were closer to the observer and some were farther in depth. The spatial configuration was defined by the layout of the items; and the depth information was preserved in a particular configuration during training. After observers had acquired contextual cueing, Kawahara reversed the depth plane of the two sets of items on a given visual display. Note that the retinal input across the two eyes remains constant with this manipulation; only the perceived depth relations change. Performance in the Old condition dropped significantly after the depth reversal, suggesting that the visual system is sensitive to the depth information in the configuration.

## 2.3 What is learned in other forms of contextual cueing?

In daily life, visual learning is not confined to spatial properties. Identity of items is an important feature of the visual environment as well; knowing that a refrigerator fits in a kitchen scene rather than a bedroom scene makes it potentially easier to identify within an appropriate context (Biederman, Mezzanotte, & Rabinowitz 1982). Movement patterns provide another important cue to guide vision and action; excellence in team sports relies on a player's ability to recognize and predict the movement trajectories of teammates. Just like the association between target location and spatial configuration, contextual cueing can be quickly acquired when the context is defined by semantics or motion trajectories.

In semantic contextual cueing, we generated novel shapes (Chun & Jiang 1999); some were symmetric along the vertical axis, others were symmetric along another, non-vertical axis. Observers were required to search for a vertically symmetric object among objects that were symmetric around tilted axes. Unlike the spatial contextual cueing paradigm, we randomly positioned the target and distractors on each trial so that spatial information was completely useless for search. We created Old displays by consistently pairing the target shape (a particular instance of vertically symmetric objects) with distractor shapes (a set of tilted objects). The New condition was created by pairing the target shape with a randomly selected set of distractor objects. In both the New and Old conditions, target shapes were repeated the same number of times across the session; and the distractor sets were also similarly repeated. However, target and distractor set pairings were consistent only in the Old condition. Significant benefits were obtained for search through Old displays, which is a notable result given that target locations were randomized across Old display repetitions. In this form of contextual cueing, learning of the target shape, distractor shapes, and their consistent pairing allows the shape context to prime the target shape and make it more detectable.

Even more impressive is the learning of random motion trajectories (Chun & Jiang 1999). We generated items that all started as "+"s. The items started to move in random trajectories, and gradually revealed their identity as either *L*s (distractors) or *T* (target). In the Old condition, the target trajectory was consistently paired with the same set of distractor trajectories across repetitions. In the New condition, the target trajectory was variably paired with several sets of distractors across blocks. We matched the initial displays between New and Old to make sure that learning was not due to repetition of the initial layout in the Old condition. Results showed that observers were faster to spot the target when its trajectory was paired consistently with a particular set of distractor trajectories.

Contextual cueing is not limited to repetition of spatial, identity, or motion trajectory information, it can also be applied to purely temporal information. When a target is presented among a sequence of distractors, it can be detected more rapidly if it is presented at a temporal moment that is predictable from its preceding context, defined by the varying durations of a stream of distractors (rhythm) or by a fixed sequence of object identities (Olson & Chun 2001).

These different forms of contextual cueing indicate that the visual system is equipped with a powerful learning system that can quickly extract important invariant information from the visual environment, be it spatial configuration, shape identity, motion trajectory, or temporal sequence. Learning of such invariance then provides top-down guidance to visual attention mechanisms, so that behaviorally relevant information can be quickly processed.

## 2.4 Contextual cueing is implicit

Contextual cueing would not be as interesting if the repetitions were obvious. If observers were aware of the consistent pairing and could immediately notice the repetition, it would be trivial for them to allocate attention to the target location. We have revealed, instead, that observers in the contextual cueing task were unaware of the critical association. In addition, even when instructed of the nature of repetition, observers simply failed to use such information explicitly.

To assess whether contextual cueing can occur outside of awareness, we tested conscious memory for repeated contexts (Chun & Jiang 1998). Following the spatial contextual cueing task, we asked our observers whether they had noticed any repetitions in the spatial configuration. Typically fewer than 20% of the participants reported noticing the repetitions. We then presented a recognition test, in which Old configurations were intermixed with New configurations. Observers were unable to discriminate whether a display was Old or New at above chance levels, regardless of whether they reported noticing repeated displays or not. Such chance level performance in recognition contrasts sharply with the significant RT benefit in visual search. This suggests that learning and memory of contextual information are implicit. Similarly, we found chance-level recognition performance for other forms of contextual cueing, such as shape learning, motion trajectory learning, and temporal sequence learning (Chun & Jiang 1999; Olson & Chun 2001).

A recognition test, however, is not the most sensitive test of explicit knowledge. In fact, Shanks and St. John (1994) have criticized the assessment of explicit learning using recognition tests. According to their information criterion, a test of explicit learning needs to probe the same information useful for the implicit task. For example, in our spatial contextual cueing task, implicit learning revealed in visual search relies on target location information within a particular configuration. An explicit test should also probe the same information. A recognition test does not fully satisfy the information criterion because it assesses observers' familiarity with a configuration, not the association between a configuration and target location. It is possible that observers may have explicit knowledge of the target location among a particular configuration without being able to explicitly identify that configuration as Old.

To assess the implicit nature of contextual cueing more stringently, we designed a guessing test that explicitly probed the information that was used to facilitate visual search (Chun & Jiang, accepted). Our search task was the same as before, except that we added cross marks to separate the display into four quadrants. In the explicit test, we presented observers with Old and New configurations but substituted the target with a distractor. We then asked observers to guess the quadrant that was most likely to have contained the target given a particular configuration. This test procedure was thus more similar to the visual search task than the Old-New recognition test used before. Unlike the Old-New recognition test, this guessing task removed the requirement to make verbal report of which display was old and which was new. Instead, observers were probed to explicitly determine the target location within a particular configuration, the same information that was presumably useful for the visual search task as well. Even in this procedure, however, our observers were unable to guess the target's location at above chance levels, indicating that they truly lacked explicit knowledge of what was learned. This guessing test strengthened our claim that learning during the visual search procedure could occur without explicit knowledge.

To further test the implicit nature of contextual cueing, we explicitly instructed another group of observers that displays would be repeated throughout the experiment and that attending to the repetitions would facilitate search. We encouraged them to notice, memorize, and use such repetitions during the search task. Even with such explicit instruction, subjects did not show a larger contextual cueing benefit than subjects who did not receive instruction. This suggests that explicit learning does not (always) contribute to contextual cueing. In addition, guessing task performance did not improve for the instructed group as well. Although a small portion (25%) of observers reported that they recognized repetitions during visual search, neither these aware nor other unaware observers guessed the target locations for Old displays at above chance levels. Incidentally, those who reported recognizing repetitions during search showed, on average, no contextual cueing during visual search, suggesting that their effort in encoding the displays may have hindered how they approached the search task. Like other prototypical paradigms of implicit learning (e.g., artificial grammar learning), contextual cueing is somewhat impenetrable by explicit instructions (Reber 1989).

The guidance of attention by implicit learning suggests that attention and consciousness are not one and the same. Attention can be shaped by some processes occurring unconsciously, outside of awareness. Attending to something is not merely "willfully" devoting conscious effort to some noticeable aspect of the display. Even when observers are unaware of the critical association between implicitly learned contexts and target locations, attention is guided to targets by the same information.

## 2.5 Neural basis of contextual cueing

Implicit and explicit systems can be dissociated not only at the behavioral level but also at the neurophysiological level. Although the exact neural substrate for implicit learning may vary depending on the particular learning task, implicit learning typically relies on non-hippocampal brain structures. For example, SRT learning is intact in amnesia (Nissen & Bullemer 1987), and neuroimaging studies have shown that visual-motor regions are involved (Grafton, Hazeltine, & Ivry 1998; Rauch et al. 1997). Implicit learning is not affected by damage to the hippocampus, which is critical for explicit, declarative learning and memory (Squire 1992).

The spatial contextual cueing paradigm provides an interesting exception to this rule. Because the task is implicit, conventional wisdom suggests that contextual cueing should also be preserved in amnesia. However, amnesic patients with hippocampal and neighboring medial temporal lobe damage do not exhibit contextual cueing effects (Chun & Phelps 1999). Although this appears to be a puzzling result, it is consistent with animal research that proposes the hippocampus is important for relational, contextual learning (Cohen & Eichenbaum 1993; O'Keefe & Nadel 1978), independent of whether the information is consciously accessible or not. Damage to the hippocampus produced impaired spatial learning and memory in rats. Because our task requires relational, contextual learning, it depends on the integrity of the hippocampus and neighboring structures. This finding does not imply that the hippocampus is not important for explicit learning; it clearly is. However, our result does suggest that the implicit/explicit learning distinction may not always map onto dissociable memory systems in the brain. Future studies need to examine what makes the spatial contextual cueing task hippocampal-dependent: is it because amnesic patients failed to extract the repeated spatial layouts, or because they failed to acquire the association between layout and target location.

## 3. Contextual cueing depends on selective attention

Focused attention is clearly needed for explicit learning and memory. What about implicit learning? Some have argued that it, too, depends on attention (Nissen & Bullemer 1987); but many others believe that implicit learning can proceed (somewhat) automatically, independent of attention (Frensch et al. 1998; Jiménez & Méndez 1999, 2001; Stadler 1995).

The Serial Reaction Time task (SRT) has been commonly used to study this question (Nissen & Bullemer 1987). Observers were presented with four spatial positions, a target (e.g., the symbol "*") was presented at one of the four positions on a trial, and observers pressed one of four keys whose position corresponded to the target position. Observers performed hundreds of trials, and unknown to them, a sequence (of 10 trial positions) was repeatedly presented. As the training progressed, observers got faster in the task. When the repeated sequence was replaced by a random sequence at the end of the session, RT slowed down. Restoring the original repeated sequences immediately brought RT to a faster level. It has been demonstrated that the SRT task could be learned implicitly. Explicit learning may occur particularly when simple, repeated sequences were used; although it is easily prevented if the sequence is probabilistic rather than deterministic (Jiménez, Méndez, & Cleeremans 1996)

To determine whether implicit learning is affected by attention, Nissen and Bullemer (1987) tested subjects under either single or dual task conditions. In the dual-task condition, subjects were presented with the SRT task along with tones of high or low frequencies. They needed to perform the SRT task, and to keep a mental count of the number of high frequency tones. Nissen and Bullemer found that learning in the SRT task was impaired in the dual-task condition, suggesting that the withdrawal of attentional resources by the secondary tone-counting task was detrimental to implicit learning.

This conclusion has been subsequently challenged on several grounds. First, the tone-counting task produced an effect on the expression of what was learned, but not on the learning per se (Frensch et al. 1998). If during the test phase only SRT was performed, then a significant learning effect kicked in immediately, as if subjects had learned the sequence during the dual-task session but failed to express their learning. Second, the impairment shown in the dual-task relied on the structure of the SRT series. The tone-counting task had no effect if the sequence was unique (i.e., each location is uniquely associated with the next possible location). The counting task impaired learning only for ambiguous or hybrid sequences (Cohen, Ivry, & Keele 1990). Third, the detrimental effect of the tone-counting task was not because it grabbed attention away from the SRT task, but because the tones themselves created a sequence that interfered with the acquisition of the primary visuo-motor sequence (Stadler 1995). SRT learning was impaired in a single task when random blank intervals were inserted in the spatial-motor sequence; and it was preserved in a dual-task when the secondary task was letter memory. Although the dissociation between unique and ambiguous sequences in the Cohen et al.'s study has not always been replicated, Stadler's (1995) data provide the strongest evidence that SRT may not depend on attentional resources, at least when the secondary task was tone-counting or letter memory load.

Thus, it may seem that implicit learning, as reflected by the SRT task, can proceed automatically without relying on attention. However, research on attention has taught us that the concept of attention is over-inclusive. Instead of asking whether implicit learning relies on attention, one needs to scrutinize specifically what attention is. Let's for a moment consider attention as mental effort or resources (Kahneman 1973). Attention studies have shown that instead of one pool of resources upon which all mental processes rely, attention is better considered as multiple pools of resources (Navon & Gopher 1979). Auditory and visual tasks, for instance, possibly rely on separate attentional resources (Duncan, Martens, & Ward 1997; Potter, Chun, Banks, & Muckenhoupt 1998; Treisman & Davies 1973). Even though the tone-counting task was extremely effortful, it may not have used any resources needed by the visual-motor SRT task. It is debatable whether the letter load task used by Stadler (1995) relied on the same resources as the SRT task as well. One may object that our argument is untestable, as it is impossible to determine how many different types of separate resources exist. This is precisely the problem with the multiple resources concept (Navon & Gopher 1979). Consequently, it is not easy to answer affirmatively whether implicit learning depends on attentional resources. Instead, it is more feasible to characterize what resources are or are not needed by implicit learning. The answer to this question is likely to depend on the nature of the particular implicit learning task. A visual-motor task (SRT) may give a different answer than a spatial layout task.

In addition to "efforts" and "resources", attention can also be considered as a selective process. Portions of the input are selectively attended while others are ignored. Viewed as a selective process, is attention necessary for implicit learning?

Using the SRT paradigm and probabilistic sequences, Jiménez and Méndez (1999) examined how selectivity affects implicit learning. Subjects in their study always had to respond and attend to the location at which the target appeared. In addition, the shape of the target could be one of four, and it predicted the next location. The association between shape and target location was learned only when observers attended to the shape of items. In contrast, learning of the location series was not affected by a secondary tone-counting task, suggesting a possible dissociation between selection and resources (Kahneman 1973).

We have recently addressed the effect of selective attention using the spatial contextual cueing paradigm (Jiang & Chun 2001). To achieve selective attending, we presented observers with items in two sets of colors: red and green, and instructed observers to attend to one color set (e.g., all the red items). This manipulation allowed us to orthogonally manipulate the repetition of the attended set and the ignored set of items, which were interleaved with each other on the screen. Unlike Jiménez and Méndez's study which varied whether one dimension (shape) of an object was attended or not, our study allowed attention to selectively pick out a group of objects from other objects.

When the attended configuration of items was repeated but the ignored set was random, observers showed a significant contextual cueing effect. This suggested that the attended set could have an effect, even when this set only constituted half of the entire display. In contrast, when the ignored set was repeated but the attended set was random, no contextual cueing was observed (see Figure 3). Interestingly, the impact of the ignored set was modulated by the difficulty of the task. When the target was highly similar to the distractors in shape, repeating the ignored set had no effect on RT. But when the target was dissimilar to the distractors and the task was easy, repeating the ignored set produced a small but significant contextual cueing. We believe that this pattern of results can be explained by the perceptual load theory (Lavie 1995; Lavie & Tsal 1994). When selecting the target from the distractors was easy, some attentional resources were available to process the ignored set, resulting in imperfect selectivity. As the task difficulty increased, observers attended harder to the relevant set and the ignored set was more efficiently filtered. In this framework, attentional load (resources) and selectivity are closely related. The load affects the efficiency of selectivity, which in turn affects the amount of implicit learning of "unattended" events.

One may wonder, if the ignored set failed to support contextual cueing, why was there contextual cueing in the standard paradigm? After all, wasn't the configuration created by distractors, and weren't all distractors rejected in the experiment? The answer lies in that the distractors in the standard paradigm were similar to the target and were thus rejected by an attentive process (Duncan & Humphreys 1989; Treisman & Sato 1990). In contrast, the ignored set tested in Jiang and Chun (2001) differed from the target by a salient single fea-

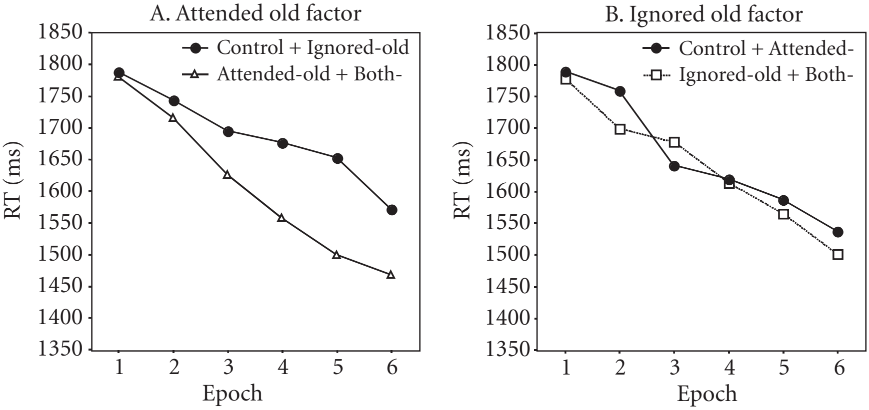


**Figure 3.** The effect of the attended set factor (A) and the effect of the ignored set factor (B). Adopted from Jiang and Chun 2001.

ture – color, and was thus rejected by a pre-attentive process. Only in the latter case was the ignored set efficiently ignored. In contrast, the distractors tested in the standard paradigm were initially attended and then rejected, sufficient to produce contextual cueing.

We conclude, therefore, that spatial contextual cueing depends on selective attention. This is our second piece of evidence against the idea that attention and consciousness are one and the same. Implicit learning acquired outside awareness may nevertheless rely on selective attention. This also serves to underscore the notion (Wright & Whittlesea 1998) that implicit learning is not a passive system that is entirely stimulus driven.

Do all forms of implicit learning depend on selective attention? So far there is not enough data to answer this question affirmatively. We know from Jiménez and Méndez's (1999) study that in SRT, learning of the shape sequence relies on selective attention. So SRT and contextual cueing paradigms converge on this point. It is possible, however, that different implicit learning tasks may rely on selective attention to different degrees. The effect of attention may also differ depending on the role of selection: whether it is to select some objects from others, one dimension of the same object from other dimensions, or one spatial location from several locations. What is clear is that at least some implicit learning tasks critically rely on selective attention; whether all forms of implicit learning behave in the same way requires further investigation.

To conclude, studies on contextual cueing revealed that top-down attention can be efficiently deployed to relevant information extracted from past experience. Such deployment of attention occurs largely in an implicit manner. In addition, what can be learned through past experience depends on what information is selectively attended. Implicit learning shapes visual attention, and visual selective attention modulates what can be implicitly learned.

## References

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Chun, M. M. & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71.

Chun M. M. & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10,* 360–365.

Chun, M. M. & Jiang, Y. (accepted). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

Chun, M. M. & Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience, 2*, 844–847.

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

Cohen, N. J. & Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System.* Cambridge, MA: MIT Press.

Doyon, J., Gaudreau, D., Laforce, R. Jr., Castonguay, M., Bedard, P. J., Bedard, F., & Bouchard, J.-P. (1997). Role of the Striatum, Cerebellum, and Frontal lobes n the learning of a visuomotor sequence. *Brain and Cognition, 34,* 218–245.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature, 387,* 808–810.

Fletcher, J., Maybery, M. T., & Bennett, S. (2000). Implicit learning differences: A question of developmental level? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26,* 246–252.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological Research, 61,* 83–98.

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

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1998). Abstract and effector-specific representations of motor sequences identified with PET. *Journal of Neuroscience, 18,* 9420–9428.

Jiang, Y. & Chun, M. M. (2001). Selective attention modulates implicit learning. *Quarterly Journal of experimental Psychology, 54A,* 1105–1124.

Jiménez, L. & Méndez, C. (1999). Which attention is needed for implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25,* 236–259.

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

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 948–969.

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall.

Keele, S. W., Jennings, P., Jones, S., Caulton, D., & Cohen, A. (1995). On the modularity of sequence representation. *Journal of Motor Behavior, 27,* 17–30.

Lambert, A., Naikar, N., McLachlan, K., & Aitken, V. (1999). A new component of visual orienting: Implicit effects of peripheral information and subthreshold cues on covert attention. *Journal of Experimental Psychology: Human Perception & Performance, 25,* 321–340.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception & Performance, 21,* 451–468.

Lavie, N. & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics, 56,* 183–197.

Logan, G. D. (1988). Towards an instance theory of automaticity. *Psychological Review, 95,* 492–527.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77,* 24–42.

Mack, A. & Rock, I. (1998). *Inattentional Blindness.* Cambridge, MA: MIT Press.

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 350–364.

Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron, 20,* 445–468.

Moran, J. & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229,* 782–784.

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

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

O'Keefe, J. & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.

Olson, I. R. & Chun, M. M. (2001). Temporal contextual cueing of visual attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27,* 1299–1313.

Olson, I. R. & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition, 9*, 273–302.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23,* 324–354.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174

Potter, M. C., Chun, M. M., Banks, B., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24,* 979–992.

Pylyshyn, Z. W. & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197.

Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., Bush, G., Breiter, H. C., & Rosen, B. R. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping, 5,* 124–132.

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

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17,* 888–896.

Rock, I. & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 275–285.

Shanks, D. R. & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367–447.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99,* 195–231.

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

Stadler, M. A. & Frensch, P. A. (Eds.) (1998). *Handbook of Implicit Learning.* Thousand Oaks, CA: Sage Publications.

Treisman, A. & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance, 16,* 459–478.

Treisman, A. M. & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), *Attention and Performance IV* (pp. 101–117). New York: Academic Press.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition, 27,* 561–572.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of guided search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception & Performance, 26,* 693–716.

Wright, R. L. & Whittlesea, B. W. A. (1998). Implicit learning of complex structures: Active adaptation and selective processing in acquisition and application. *Memory & Cognition, 26,* 402–420.

# Attention and implicit memory

Neil W. Mulligan* and Alan S. Brown
Southern Methodist University

> Memory . . . is the knowledge of a former state of mind after it has already once dropped from consciousness; or rather *it is the knowledge of an event*, or fact, of which meantime we have not been thinking, *with the additional consciousness that we have thought or experienced it before.* [italics in original]
>
> (James 1890: 648)

> Whatever future conclusions we may reach as to this, we cannot deny that *an object once attended to will remain in the memory*, whilst one inattentively allowed to pass will leave no traces behind. [italics in original]
>
> (James 1890: 427)

This chapter investigates the role of attention during encoding in implicit memory. The above quotes provide a useful starting point because they present a view of memory and attention that is traditional in two ways. First, James focuses on memory as the faculty of conscious recollection; the recognition of a current mental state as the product of past experience. Second is the assertion that attention plays a critical role during memory encoding, that attending to an object enhances later memory for its occurrence. The present paper examines these traditional views in light of recent research on attention and implicit memory.

The first of the James' quotes reflects the common usage of the term "memory" to imply conscious recollection. However, researchers have long supposed that memory for prior events can affect behavior when people are not trying to remember, and indeed, when people are not aware that memory for prior events is operative (Ebbinghaus, 1885; James, 1890, in his chapter on Habit, was not insensitive to this point). In modern psychology, a confluence of results from cognitive psychology, neuropsychology, and neuroscience has focused interest onto these same issues, centering around the distinction between explicit and implicit memory (e.g., Roediger & McDermott 1993; Schacter 1987, Tulving & Craik 2000; Squire 1992). "Explicit memory refers to intentional or conscious recollection of prior experiences, .. .. implicit memory, by contrast, refers to changes in performance or behavior that are produced by prior experience on tests that do no require any intentional or conscious recollection" (Schacter 1992: 244).

Explicit memory is typically measured with traditional memory tests, such as recognition and recall, in which participants are directed to think back about some prior (usually experimenter-provided) event and report information about it. On implicit memory tests, participants are simply asked to perform a task (e.g., completing word fragments; generating category examples), without reference to any prior experience. Memory for prior events is inferred from the increased ability to complete, generate, identify, or otherwise process the present stimuli. The enhanced processing is called priming.

The principles that govern implicit and explicit memory appear to differ in important ways (for reviews see Roediger & McDermott 1993; Schacter 1987; Toth 2000). Particularly striking are population dissociations between implicit and explicit memory. For example, compared to healthy control subjects, amnesics are profoundly impaired on explicit memory tests but often not on implicit tests (Shimamura 1993). Several other populations (such as people with depression, or schizophrenia, and older adults) show a similar pattern: deficient explicit memory coupled with normal, or near-normal, levels of priming on implicit memory tests (e.g., Denny & Hunt 1992; Elliott & Greene 1992; Light 1991; Mitchell, Brown, & Murphy 1990; Schwartz, Rosse, & Deutsch 1993). Pharmacological treatments have produced similar dissociations. For example, several studies have reported that administering benzodiazepine (a class of drugs including alprazolam, triazolam, diazepam, and midazolam) prior to a study session produces poor explicit memory but equivalent performance on implicit tests compared to a placebo-control group. Thus, pharmacological amnesia produces the same type of dissociation as produced by organic amnesia: it affects conscious recollection but appears to have no affect on unconscious influences of memory (see Curran 2000, for a review). In addition to population and pharmacological dissociations, a number of experimental manipulations, such as levels-of-processing, the read/generate manipulation, and study modality, have also produced dissociations between priming and performance on explicit tests (Roediger & McDermott 1993). Such dissociations, coupled with recent neuroimaging evidence (e.g., Gabrieli 1998; Schacter & Badgaiyan 2001), argue for the existence of separable components of memory underlying implicit and explicit memory phenomena.

The second James' quote embodies the traditional view that attention during encoding plays a critical role in later memory performance. Consistent with this view, there is overwhelming evidence that dividing attention during encoding reduces performance on many explicit measures of memory, including free and cued recall and recognition (e.g., Baddeley, Lewis, Eldridge, & Thompson 1984; Broadbent 1958; Cherry 1953; Craik, Govoni, Naveh-Benjamin, & Anderson 1996; Fisk & Schneider 1984; Moray 1959; Murdock 1965; Norman 1969). These findings encouraged early information-processing psychologists to concur with James, and posit that attention is required for the formation of durable memory traces (e.g., Broadbent 1958; Cherry 1953; Moray 1959; Norman 1969). For instance, Norman (1969) in his book *Memory & Attention*, states that attention is required for the analysis and organization of incoming stimulus information. It is only with such attention-demanding encoding processes that new information could be integrated into existing knowledge structures in ways that support later retrieval (Norman 1969, especially pp. 177–181). A similar, central role for attention in encoding has also been espoused in more recent work (e.g., Bentin 1994; Craik 1989; Craik et al. 1996; Cowan 1995; Fisk & Schneider 1984).

Given the centrality of attention in theories of memory encoding, and the rather uniform effects of divided attention on explicit memory, it is important to evaluate the role of attention in implicit memory, especially because different researchers have come to very different conclusions about the reliance of implicit memory on attention. Some researchers have concluded that attention during encoding is crucial for implicit as well as explicit memory (e.g., Pashler 1998; Wood, Stadler, & Cowan 1997). Others have concluded that implicit memory, in contrast to explicit memory, has little reliance on attention and largely reflects automatic encoding processes (e.g., Bentin et al. 1995; Besson, Fischler, Boaz, & Raney 1992; Graf & Mandler 1984; Isingrini et al. 1995; Jacoby, Toth, & Yonelinas 1993; Jacoby et al. 1989; Jelicic, Bonke, Wolters, & Phaf 1992; Parkin et al. 1990; Parkin & Russo 1990; Shallice et al. 1994; Szymanski & MacLeod 1996). Still others have suggested that attention is important for some forms of implicit memory but not others (e.g., Gabrieli et al. 1999; Mulligan & Hartman 1996; Wolters & Prinsen 1997).

## 1. Initial studies: Attentional state dissociates implicit and explicit memory

Some of the initial studies of attention and implicit memory revealed dissociations between implicit and explicit tests. Consider the results of Parkin and Russo (1990). During the encoding phase of this experiment, participants named fragmented pictures of everyday objects. Each study trial began with the most fragmented version of the picture and progressed through increasingly less fragmented versions of the picture, until the participant succeeded in identifying the object. The level of fragmentation was recorded and the participant proceeded to the next trial. In the full attention condition, this was the sole task. In the divided attention condition, this study task was performed simultaneously with a tone-monitoring task, in which a series of tones were presented over an audio tape at random intervals every 2 to 5 seconds. Participants categorized the tones as high, medium, or low, as they tried to identify the fragmented pictures. Twenty-four hours later, participants were either given an explicit test, in which they recalled the names of the pictures, or an implicit test, in which they identified fragmented pictures using the same procedure as in the study session. The picture-fragment test included pictures from the study episode as well as counterbalanced new pictures. Priming was indicated by identification of studied (or old) pictures at lower levels of clarification than the new pictures. For the recall test, the full attention group recalled an average of 8.5 of the 15 studied pictures whereas the divided attention group recalled only 5.9. As expected, dividing attention had a substantial effect on explicit memory, reducing recall by 30%. However, the amount of priming on the picture-fragment task was essentially the same for the full and divided attention conditions (with averages of 2.19 and 2.06, respectively on Parkin & Russo's picture fragmentation scale).

The results of Parkin and Russo (1990) indicate that dividing attention during encoding, which virtually always reduces explicit memory, may have little impact on implicit memory, at least for some divided-attention tasks and for some implicit memory tests (see Jacoby et al. 1989; Mulligan & Hartman 1996; Parkin, Reid, & Russo 1990; Russo & Parkin 1993; Schmitter-Edgecombe 1996a, 1996b for similar results). Likewise, when attention is diverted from an information source by a very distracting task, explicit memory may be eliminated entirely (Cherry 1953; Moray 1959) while implicit memory tests still demonstrate evidence of retention (e.g., Eich 1984; Jelicic, Bonke, Wolters, & Phaf 1992; Mandler, Nakamura, & Van Zandt 1987; Merikle & Reingold 1991; Seamon, Marsh, & Brody 1984). For instance, Eich (1984) used a dichotic listening task in which the unattended channel included a series of homophones together with words that biased their less frequent meaning (e.g., taxi-FARE). Subjects showed no memory for the unattended homophones on an explicit test of recognition. However, when the subjects were simply asked to spell the target homophones, with no overt reference to the listening task, there was a greater likelihood of choosing the biased rather than the unbiased spelling. Thus, an implicit test of memory indicated that information about the unattended information was retained (cf. Wood et al. 1997).

These types of dissociations gave impetus to the claim that implicit memory relies largely on automatic encoding processes, which we refer to as the attentional view of implicit memory. However, subsequent research demonstrates that this view is no longer tenable.

## 2. The distinction between perceptual and conceptual implicit memory

A thorough consideration of recent research on attention and implicit memory begins with the distinction between perceptual and conceptual implicit tests as described in the transfer-appropriate processing (TAP) account. This is a useful starting point because the TAP view is a prominent account of implicit memory, rendering the most successful account of functional dissociations (Roediger 1990; Roediger & McDermott 1993; Roediger, Buckner, & McDermott 1999). In addition, much of the recent research on attention and implicit memory has been framed around the perceptual-conceptual distinction (e.g., Gabrieli et al. 1999; Light, Prull, & Kennison 2000; Mulligan 1997, 1998; Mulligan & Hartman 1996; Mulligan & Stone 1999; Schmitter-Edgecombe 1996a, 1996b, 1999). Within the TAP framework, performance on a memory test is enhanced to the extent that encoding processes are re-engaged at the time of retrieval (Kolers & Roediger 1984; Morris, Bransford, & Franks 1977). More specifically, the TAP framework proposes that memory tests should be characterized primarily by the types of memory processes that they engage, rather than by test instructions (i.e., implicit vs. explicit). Towards this end, the TAP framework differentiates between two broad classes of memory process, perceptual processes, defined as the analysis of perceptual or surface-level features, and conceptual processes, defined as the analysis of meaning or semantic information.

The most widely used implicit memory tests involve the identification or completion of degraded or ambiguous perceptual cues, such as word fragments or stems, briefly presented words, or fragmented pictures. The TAP framework posits that these tests are primarily dependent on perceptual retrieval processes. As a consequence, these tests are sensitive to variation in perceptual processes at encoding but relatively insensitive to variation in conceptual encoding. In contrast, commonly-used explicit tests (such as free recall and recognition) are assumed to primarily utilize conceptual processes during retrieval (Craik, Moscovitch, & McDowd 1994; Roediger, Weldon, Stadler, & Reigler 1992). The TAP framework accounts for dissociations between implicit and explicit tests in terms of the differences between underlying processes rather than in terms of task instructions. Specifically, because the most common implicit and explicit memory tests rely differentially on perceptual and conceptual processes, performance on these two types of tests benefit from different types of encoding procedures and are thus dissociable.

It is important to note, however, that not all implicit tests are perceptual and not all explicit tests are conceptual. With the appropriate combinations of memory cues and task instructions, conceptual implicit tests and perceptual explicit tests may be constructed. The TAP framework proposes that conceptual and perceptual retrieval tasks are dissociable regardless of their implicit–explicit status. This parsimonious approach has had a good deal of success in accounting for and predicting dissociations among memory tests (see Roediger & McDermott 1993, for a review).

With regard to the effects of attention, the TAP framework produces two implications. If, as is traditionally assumed, dividing attention at study reduces the amount of conceptual or semantic processing (e.g., Broadbent 1971; Craik 1983; Craik & Byrd 1982; Craik et al. 1996; Norman 1969) but has little or no effect on perceptual identification processes, then the TAP framework predicts that: 1) divided attention should have little or no effect on perceptual tests; and 2) divided attention should reduce priming on conceptual tests. Consequently, the TAP view predicts that divided attention will produce a functional dissociation not only between perceptual implicit tests and conceptual explicit tests, but also between perceptual and conceptual implicit tests.

Most of the initial studies on attention and implicit memory focused on perceptual implicit memory and produced results consistent with the TAP view (some exceptions will be discussed later). For example, Parkin and Russo (1990) found that dividing attention disrupted free recall but not perceptual priming in picture-fragment completion. Parkin et al. (1990) used a similar tone-monitoring to divide attention at encoding and found that it reduced recognition memory but not perceptual priming in the word-fragment completion test (see also Russo & Parkin 1993). Smith and Oscar-Berman (1990) reported that dividing attention during encoding left priming in lexical decision unaffected when the studied materials were words (though interestingly, not when the studied items were non-words). Bentin et al. (1995) reported similar results; dividing attention during encoding with dichotic listening impaired recognition memory but not perceptual priming in lexical decision (this experiment only used words as studied materials). In addition, Jacoby et al. (1989) found that dividing attention during encoding reduced recognition but had no significant effect on priming in the fame judgement task (although it is unclear that this task would be categorized as perceptual according to the TAP framework; see Mulligan & Hartman 1996, for discussion).

There were two aspects of this research that prompted the study by Mulligan and Hartman (1996). First, the earlier research had focused on perceptual implicit memory, and Mulligan and Hartman wanted to determine if the dissociation between explicit and implicit memory would generalize to conceptual tasks. The TAP view suggests it would not, whereas the attentional view suggests that it would. Second, the prior studies in which divided attention dissociated (perceptual) implicit and explicit memory had used memory tests that differed in several ways besides task instructions (the defining, or operational, characteristic of implicit and explicit memory tests) (Schacter, Bowers, & Booker 1989). A typical example is Parkin and Russo (1990) who contrasted free recall and picture fragment completion, tasks that differ in terms of overt retrieval cues and response requirements, as well as task instructions. Thus, it was unclear if these dissociations were actually due to type of test (i.e., implicit vs. explicit) or to one of the other differences between the implicit and explicit tests used. In order to determine if the differentiating effects of attention are due to the implicit/explicit distinction, one needs to compare performance on memory tests that are equivalent in all regards except test instructions.

Mulligan and Hartman (1996) conducted a pair of experiments to examine these issues. In each experiment, participants studied words under a full or divided-attention condition. In the full attention condition, participants' sole task was to read aloud a series of words presented on a computer screen. In the divided attention condition, participants read aloud the words and simultaneously monitored a series of aurally-presented digits, signaling whenever they heard three odd numbers in a row. After a brief distractor task, participants were given one of a number of memory tests. In one experiment, the test was either word-fragment completion, a perceptual implicit test, or its explicit counterpart, word-fragment cued recall. The latter task used the same word-fragment cues but participants were instructed to use the fragments to try to remember completions from the study list. The other experiment examined conceptual implicit memory, using the category-exemplar production task, a commonly-used conceptual implicit test in which participants are presented with the names of categories and asked to rapidly produce examples. Although the participants are not informed, some of the categories correspond to examples presented in the study list and some do not. Conceptual priming is computed as the difference between the proportion of the studied exemplars produced and an appropriate baseline measure. A separate group of participants was given a category-cued recall test, the explicit counterpart of category-exemplar production, in which participants are presented with the same category names and asked to use them to recall studied examples.[1]

As noted above, the TAP view predicts that dividing attention should impact the conceptual implicit test, as well as its explicit counterpart. In addition, the TAP view predicts that there should be no effect on the perceptual test, to the extent that the full and divided attention conditions produced complete perceptual analysis of the stimulus. With regard to the word-fragment cued recall test, the TAP view suggests that this task should be affected by divided attention because conceptual memory processes appear to play a larger role in explicit tests (Craik et al. 1994; Graf & Mandler 1984; Roediger et al. 1992). Thus, changing the instructions of a perceptual implicit test to render a comparable explicit test should also have the effect of making the test sensitive to the effects of divided attention.

The results of this study confirmed these expectations and are presented in Figure 1. First, as expected, explicit memory was reduced by divided attention in both the word-fragment and category cued recall tests. Second, perceptual priming in the word-fragment completion task was unaffected by this divided-attention manipulation. Coupled with the results of word-fragment cued recall, this result indicates that dividing attention can dissociate perceptual priming from explicit memory when matched implicit and explicit tests, ruling out alternative explanations based on differences in retrieval cues or response requirements. Third, dividing attention had a substantial effect on the category-exemplar production test, not only reducing conceptual priming, but rendering it non-significantly different from zero. Figure 1 represents a dissociation not only between perceptual implicit memory and explicit memory but also between perceptual and conceptual implicit memory. Thus, conceptual implicit memory appears to be quite reliant on attention during encoding.

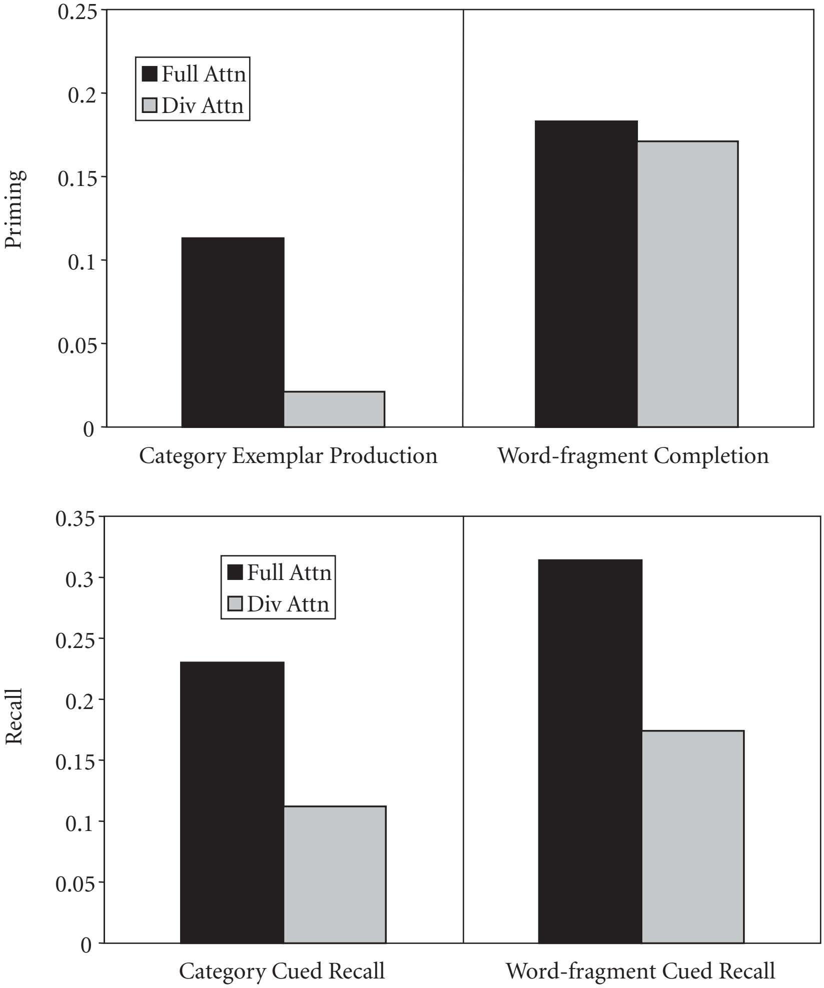


**Figure 1.**

## 3. Attention and conceptual implicit memory

The results of Mulligan and Hartman (1996) indicated that dividing attention can have substantial impact on conceptual priming as measured by the category-exemplar production test. This result has been replicated a number of times. Schmitter-Edgecombe (1996b) used a tone-monitoring task (similar to that used by Parkin & Russo 1990) to divide attention and found that it reduced both category-cued recall and category-exemplar priming, a pattern of results replicated under somewhat different conditions in Schmitter-Edgecombe (1999). Gabrieli et al. (1999) used the same digit monitoring task as Mulligan and Hartman (1996) and likewise found significant reductions in conceptual priming in category-exemplar production. Finally, Light et al. (2000) report a consistent result using a somewhat different divided-attention task. In this study, younger and older adults encoded words in either a full or divided-attention condition. In the full attention condition, each word was presented on a computer screen and judged as pleasant or unpleasant (by pressing an appropriately labeled key). In the divided-attention condition, each word was presented with a single-digit number on either side (e.g., 5 BRONZE 7). Participants were to judge the word as in the full attention condition and also to add the two numbers together and report whether the sum was an even or odd number. This divided-attention task produced sharp reductions in category-cued recall and in category-production priming for both the older and younger adults.

Category-exemplar production task appears to be sensitive to divided attention. Mulligan (1998) reported that this effect generalized to two other conceptual priming tasks, general knowledge questions and word association. The former test consists of a set of questions presented as a general knowledge, or trivia test. The answers to some of the questions are presented in the study portion of the experiment. An increased tendency to answer these questions correctly (relative to control questions) represents priming in this task. In the word association test, participants are presented with cue words (e.g., DOCTOR) and asked to produce the first word that comes to mind. Associates of some of the cue words (e.g., NURSE) were presented at study. Priming is evident to the extent that studied associates are more likely to be produced than associates not presented at study. Within the TAP framework, these tests are considered conceptual because they are affected by conceptual encoding manipulations, such as the generation manipulation (Blaxton 1989, 1992) and levels-of-processing (Hamann 1990), and are unaffected by perceptual encoding manipulations, such as varying study modality (Challis et al. 1993; Vaidya et al. 1997) (see Roediger & McDermott 1993, for a review). The TAP framework implies that these implicit tests should be affected by divided attention during encoding. Mulligan (1998) manipulated attention at encoding with the digit-monitoring task and found greater priming in the full than divided attention condition for both the general knowledge and the word-association tests.

### 3.1 Some limitations on the effects of divided attention on conceptual priming

The results reviewed above imply that conceptual priming is generally affected by attentional manipulations. However, other studies have produced conflicting results, raising several issues about the relationship between attention and conceptual priming, and attention and implicit memory more generally. First, divided attention tasks may differ in the extent to which they distract participants from the primary (encoding) task, a difference we refer to as the strength of the divided-attention manipulation. Such differences may account for divergent results (Mulligan 1997; Wolters & Prinsen 1997). For example, although Mulligan and Hartman (1996) and others found that divided attention reduced priming in the category-exemplar production task, a study published at about the same time (Isingrini et al. 1995) produced opposing results. In Isingrini et al.'s, participants studied category examples under conditions of full or divided attention and subsequently were tested with category-exemplar production or its explicit counterpart, category-cued recall. The divided attention manipulation reduced recall but not, in contrast to Mulligan and Hartman (1996), conceptual priming. This led Isingrini et al. to conclude that conceptual implicit memory reflects automatic encoding processes.

The type of distractor task differed between these two studies, leading Mulligan (1997) to hypothesize that the strength of the attention manipulation might account for the divergent results. Isingrini et al. (1995) used a letter detection task to divide attention, in which participants monitored a tape recording of a series of consonants and signaled if they detected a "B" or a "G". Mulligan and Hartman (1996) divided attention with the digit-monitoring task. It can be argued that Isingrini et al.'s was the less demanding of the two because participants only had to evaluate one stimulus at a time; they did not have to maintain any preceding consonants in memory. In contrast, participants in the Mulligan and Hartman study had to evaluate each stimulus, plus maintain information about immediately preceding stimuli. In addition, Mulligan and Hartman kept their participants focused on the detection task by prompting subjects if they missed a target sequence, while Isingrini et al. did not, which may also have increased the effectiveness of the Mulligan and Hartman's divided-attention task.[2]

To examine this issue, Mulligan (1997) varied attention at encoding by manipulating short-term memory loads. Prior to the presentation of each study word, the subject was presented with an attentional load of 0 (a full attention condition), 1, 3 or 5 digits and letters. The digits and letters were to

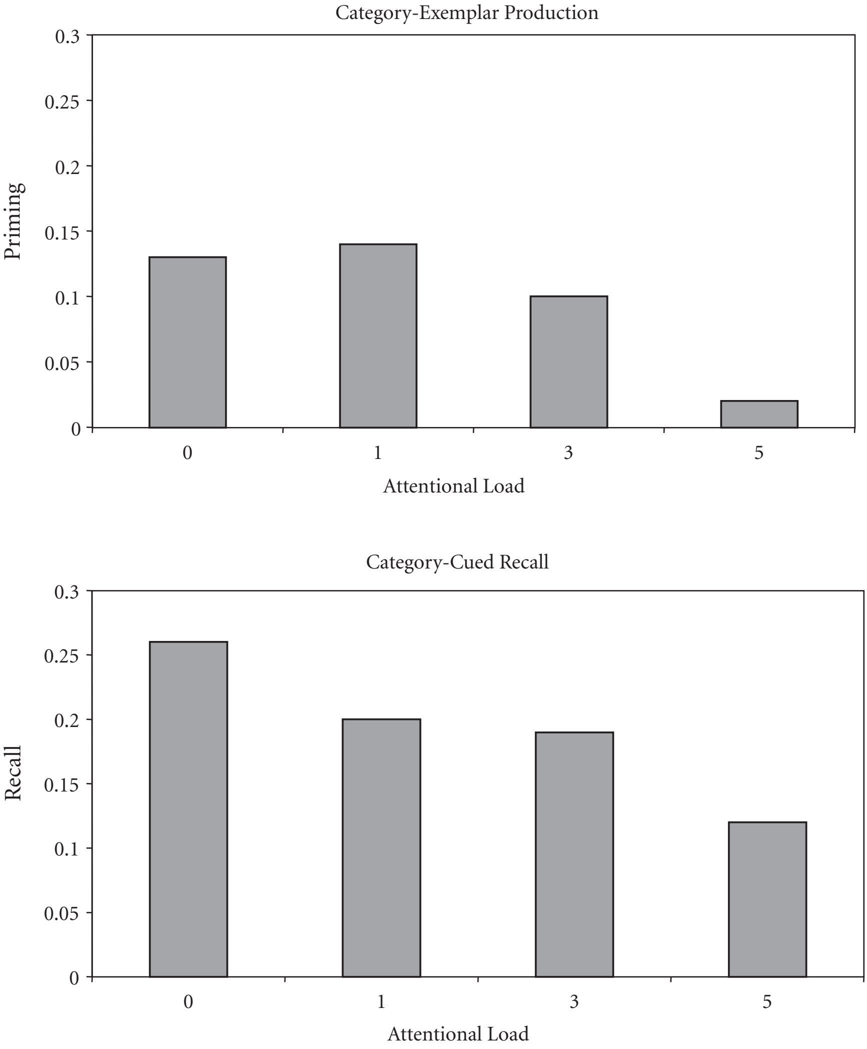


**Figure 2.**

be maintained in memory until a RECALL signal was presented a few seconds later. This technique allows a controlled division of attention over several levels within a single experiment (e.g., Baddeley & Hitch 1974; Engle, Conway, Tuholski, & Shisler 1995). Subsequently, memory for the study words was assessed with either the category-exemplar production test or category-cued recall.

The results of Mulligan (1997) are presented in Figure 2. Consider first the recall results. Recall in the 0-load condition was significantly greater than in any other condition, performance in the 5-load condition was significantly lower than in any other condition, and performance in the 1- and 3-load conditions did not significantly differ. For category-exemplar production, priming in the 5-load condition was significantly lower than in any of the other conditions and no other pair of conditions differed significantly. In addition, t-tests revealed significant priming in the 0-, 1- and 3-load conditions but not in the 5-load condition.

These results may resolve the conflicting results of Isingrini et al. (1995) and Mulligan and Hartman (1996). Considering the 0-load condition as a full attention condition and the 3-load condition as a moderate divided-attention condition, the results replicate Isingrini et al. Over this range, dividing attention produced a small and non-significant decrease in category-exemplar production and a larger, significant decrease in category-cued recall. If we contrast the 0-load (full attention) condition with the 5-load condition (considering it to be a strong divided-attention condition), the results replicate Mulligan and Hartman: Dividing attention produced large, significant decreases in performance on both memory tests. Furthermore, as in Mulligan and Hartman, strong division of attention not only decreased priming, it eliminated it. This appears to resolve the discrepancy between Mulligan and Hartman's results and those of Isingrini, et al. Conceptual priming in the category-exemplar production task is affected by divided attention, but the division of attention in Isingrini et al. was too mild to reveal this.

Mulligan and Stone (1999), using the attention load paradigm of Mulligan (1997), documented a different limitation of the effect of divided attention on category-exemplar production. First, it should be noted that in experiments using the category-exemplar production task, several examples from each category are typically presented in the study list. In Mulligan (1997), the study lists were randomly organized such that no two examples from a category appeared in sequence, and all of the examples from a category were in the same attention condition (i.e., 0-, 1-, 3-, or 5-load condition). As noted above, Mulligan found that higher attentional loads reduce conceptual priming. Mulligan and Stone obtained this same result when attentional load was manipulated within categories, such that examples from the same category were presented with different sized attentional loads, as well as between categories. Mulligan and Stone also presented study lists blocked by category, such that all the examples from the same category appeared in sequence, rendering the categorical structure of the list quite salient. When the study list was presented in a blocked fashion, and attention was manipulated between categories, then conceptual priming was reduced. However, if the study list was blocked (rendering category-level information readily accessible during encoding) and attention was varied within categories, then priming in this task was unaffected, even in the 6-load condition, a more intensive attentional load than those used by Mulligan (1997). Under these same conditions, category-cued recall is substantially affected. Thus, Mulligan and Stone demonstrate that as attentional demands increase, conceptual priming in the category-exemplar production task is reduced when the categorical structure of the list is not obvious, but not if the categorical information is quite salient (i.e., if category level information is available in working memory at the time of encoding).

The strength of the attentional manipulation or the structure of the encoding list may ameliorate the normally deleterious effects of divided attention on category-exemplar production. In fact, the strength of the attentional manipulation may impact the amount of priming for other conceptual implicit tests (i.e., word association, general knowledge questions) that have otherwise exhibited robust effects of divided attention. Another limiting condition of the effects of divided attention on conceptual priming was raised by Gabrieli et al. (1999), who argued that divided-attention effects may not generalize to all conceptual priming tasks. During the study portion of Gabrieli et al.'s experiments, participants were presented with a number of category examples in either a full or divided attention condition. In the full attention condition, each word was judged as manmade or natural. In the divided-attention condition, the participant made the manmade-natural decisions while simultaneously carrying out the digit-monitoring task. Subsequently, participants were given either the category-exemplar production task, or a category-verification task. In the latter task, participants were presented with a category-verification question (e.g., Is this a type of furniture?) followed by a test item (e.g., DESK) which may or may not be a member of the category. Some of the test items were from the study list and others were not (the counterbalanced, new items). Participants answered each question as quickly as possible and priming was measured as decreased reaction time to old compared to new items. Gabrieli et al. argue that this task is conceptual because the relevant information to-be-retrieved is conceptual rather than perceptual in nature.

Gabrieli et al. (1999) found that dividing attention reduced priming in the category-exemplar production task (replicating Mulligan & Hartman 1996, as noted above), but not in the category-verification task. Light et al. (2000) replicated this result with older and younger adults using their sum task (i.e., odd-even judgement of the sum of two digits, described above) to divide attention. While this task significantly reduced cued recall, recognition memory, and category-production priming, it did not significantly reduce priming in the category-verification task for either younger or older adults (Light et al. 2000). These results are problematic for the TAP view because the divided-attention tasks appear to be robust manipulations of attention and yet they do not affect an implicit task that relies on the retrieval of conceptual information.

In contrast to Gabrieli et al. (1999) and Light et al. (2000), a study by Bentin et al. (1998) produced the opposite conclusion. The differences in procedures between Bentin et al. on the one hand, and Gabrieli et al. and Light et al. on the other, may shed light on the type of processing required by this task. In each study trial of Bentin et al.'s Experiment 3, two words were presented, one above and one below a fixation point. One of the words was printed in red and the other in blue. Participants were told to attend to one of the words and ignore the other based on color. Participants decided whether the attended word referred to a living or a non-living object, and were later tested using the same speeded living – non-living judgements. The test items consisted of words attended at study, those unattended, and new items. Reaction times were significantly faster to the attended than either unattended or new words, and the latter two did not differ. Thus, diverting attention from a word not only reduced but eliminated priming in the category verification test. Given the similarity between Bentin et al.'s implicit test and that used by Gabrieli et al. (1999), Bentin et al.'s result imply that there are attentional requirements for the category verification task. The differences in the selective attention manipulation of Bentin et al. and the divided attention manipulations of Gabrieli et al. (1999) and Light et al. (2000) foreshadow issues that arise as we consider the role of attention in perceptual implicit memory.

## 4. Attention and perceptual implicit memory

Except for the conflicting results from category verification, conceptual priming appears dependent on attention during encoding in the category-exemplar production, general knowledge questions, and word association tasks. Studies using perceptual implicit tests have produced even more mixed results. As noted earlier, several studies found that perceptual priming was unaffected by manipulations of attention (e.g., Clarys et al. 2000; Gabrieli et al. 1999; Mulligan 1998; Mulligan & Hartman 1996; Kellogg et al. 1996; Parkin & Russo 1990; Parkin et al. 1990; Russo & Parkin 1993; Schmitter-Edgecombe 1996a, 1996b; Szymanski & MacLeod 1996). In contrast, a number of studies have found that directing attention away from the identity of the target stimuli can reduce later perceptual priming (e.g., Bentin et al. 1998; Crabb & Dark 1999; Hawley & Johnston 1991; Kinoshita 1995; Light & Prull 1995; MacDonald & MacLeod 1998; Stone, Ladd, & Gabrieli 2000; Stone, Ladd, Vaidya, & Gabrieli 1998).[3]

## 4.1 Hypotheses on the relationship between attention and perceptual priming

Experiments on attention and perceptual priming have used a variety of attentional manipulations, materials and memory tests, giving rise to a number of factors that may account for the divergent results. In our recent research (Mulligan 2001; Mulligan & Hornstein 2000), we have focused on three factors that seemed especially promising: (a) stimulus identification during study; (b) intra-modal versus cross-modal manipulations of attention; and (c) the type of response to distracting stimuli.

### 4.1.1 *Stimulus identification*

First, attentional effects in perceptual priming may be mediated by stimulus identification. Studies reporting divided-attention effects typically use attentional manipulations that render overt (or covert) identification of the study items unlikely or nearly impossible (e.g., Bentin et al. 1998; Crabb & Dark 1999; Eich 1984; Hawley & Johnston 1991; Kinoshita 1995; MacDonald & MacLeod 1998). These studies typically use selective attention tasks, in which participants' attention is directed away from the identity of the study stimuli. In contrast, those studies finding no effect of attention used divided-attention manipulations that rendered overt (or covert) identification likely, and in most of these studies, overt identification of the study items was required (e.g., Clarys et al. 2000; Gabrieli et al. 1999; Mulligan 1998; Mulligan & Hartman 1996; Parkin et al. 1990; Russo & Parkin 1993; Smith & Oscar-Berman 1990; Schmitter-Edgecombe 1996a, 1996b). Thus, the stimulus-identification hypothesis, suggests that dividing attention impairs perceptual priming *only* if the attentional manipulation disrupts identification of the study items.

The stimulus-identification account is consistent with the results of studies utilizing very brief presentations at encoding (e.g., Gellatly, Parker, Blurton, & Woods 1994; Hawley & Johnston 1991). Priming increases directly with identification rates at encoding, indicating that disrupting identification has deleterious effects on perceptual priming. Likewise, directing attention away from the lexical properties of the study item reduces perceptual priming (Hayman & Jacoby 1989). In addition, this account is consistent with theoretical views of perceptual priming. The TAP account predicts that truncation of perceptual or lexical analysis would curtail later perceptual priming (Roediger & McDermott 1993). A variant of the TAP view suggests that perceptual priming for verbal materials requires lexical access (i.e., perception of the word as a lexical entity) (e.g., Bentin et al. 1998; Hayman & Jacoby 1989; Weldon 1991; Weldon & Jackson-Barrett 1993). If an attentional manipulation precludes lexical access on at least some study trials, then a concomitant reduction in priming is expected.

### 4.1.2 *Intra- vs. cross-modal manipulation of attention*

The second factor involves the distinction between intra-modal and cross-modal manipulations of attention. Most studies reporting null effects used cross-modal division of attention (Mulligan & Hartman 1996; Parkin et al. 1990; Parkin & Russo 1990; Russo & Parkin 1993; Schmitter-Edgecombe 1996a, 1996b). For example, Mulligan and Hartman (1996) divided attention across visual study words and an auditory (digit-monitoring) distractor task. Likewise, Parkin and Russo (1990) paired visual presentation of target pictures with an auditory tone-monitoring task. This contrasts with studies reporting effects of attention on perceptual priming, which manipulated attention intra-modally (Bentin et al. 1998; Crabb & Dark 1999; Eich 1984; Hawley & Johnston, 1991; Kinoshita 1995; Light & Prull 1995; Mulligan & Hornstein 2000; Stone et al. 1998; Stone et al. 2000). For example, Eich (1984) used a dichotic listening paradigm in which both the to-be-remembered words and the distractor task (prose in the shadowed channel) were presented aurally. In Kinoshita (1995), target words flanked by digits were briefly visually presented (100 ms). In the attended condition, participants named the word and ignored the digits. In the unattended condition, participants judged whether the digits were both odd or even, and ignored the word. In this study, attention was manipulated within the visual modality. Both Eich and Kinoshita found less perceptual priming in the unattended than attended condition.

The difference between intra- and cross-modal division of attention may be important, especially in light of research on the componential nature of attention. Componential models of attention were motivated by evidence against unitary views of attention (e.g., single resource models) and differentiate among multiple aspects or components of attention (for reviews see Heuer 1996; Koelega 1996). While the proposed components differ somewhat across theories, one of the basic distinctions is between central (or modality-independent) and peripheral (or modality-specific) aspects of attention. The most intensively studied modality-specific aspect of attention is visual-spatial (hereafter referred to as visual attention; e.g., Allport 1989; Treisman 1988; Duncan 1999). For example, Johnston, McCann, and Remington (1995) proposed a distinction between two limited-capacity mechanisms of attention: one responsible for selective aspects of attention (input attention) and the other (central attention) involved in higher-level mental functions (decision making, response selection, etc.). Distinctions along these lines are found in other componential analyses (e.g., Duncan 1999; Duncan, Martens, & Ward 1997; Posner & Boies 1971; Posner & Petersen 1990; Wickens 1984). Central attention has also been associated with Baddeley's (1986) model of working memory, especially its central executive component (e.g., Rosen & Engle 1997).

The componential view raises the possibility that the encoding processes tapped by perceptual priming tasks depend on input, or modality-specific, aspects of attention but not on central aspects of attention. In contrast, the encoding processes subserving conceptual priming (such as elaboration) rely on both central and input attention. Results from studies using concurrent memory load are consistent with this proposal. In this paradigm, a short-term memory load is presented prior to the presentation of a target stimulus and recalled after the offset of the target stimulus. This manipulation is designed to occupy working memory (i.e., central attention) to various degrees (e.g., Baddeley & Hitch 1974), and presumably does not impact input attention because the target stimulus is presented in isolation, without competing within-modality distractors. Consistent with the componential view, concurrent memory load reduces later priming in the conceptual, category-exemplar production task (Mulligan 1997) but not in the perceptual tests of word-fragment completion (Clarys et al. 2000; Mulligan 1998) or picture naming (Gabrieli et al. 1999). In addition, performance on a variety of explicit tests is reduced by concurrent memory load (e.g., Baddeley & Hitch 1974; Clarys et al. 2000; Mulligan 1997, 1998). In sum, the componential analysis of attention suggests that perceptual priming may be dependent on modality-specific aspects of attention. Because the relevant input modality is visual in the present set of studies, we refer to this as the visual-attention hypothesis.

### 4.1.3 *Distractor selection*

A third set of issues centers on the role of responding to the distractors during encoding and on target-distractor synchrony. As noted above, studies reporting attention effects in perceptual priming typically used selective attention tasks (e.g., Bentin et al. 1998; Crabb & Dark 1999; Eich 1984; Hawley & Johnston 1991; Kinoshita 1995; Light & Prull 1995; Stone et al. 1998; Stone et al. 2000). In the reduced attention condition of these studies, attention was directed away from the identity of the study stimuli by requiring participants to respond to a distractor rather than to the study item. In the studies reporting null effects of attention, dual tasks were typically used to divide attention, in which participants responded to both the distractor and study stimuli (e.g., Clarys et al. 2000; Gabrieli et al. 1999; Mulligan 1998; Mulligan & Hartman 1996; Parkin et al. 1990; Russo & Parkin 1993; Schmitter-Edgecombe 1996a, 1996b). It should be noted that in most studies these three factors (stimulus identification, intra- vs. cross-modal manipulations, and selective vs. dual-task manipulations) are intertwined. Specifically, selective attention manipulations typically present targets and distractors in the same modality, and the reduced attention conditions of these studies typically render stimulus identification less likely. In contrast, the dual-task manipulations typically presented the distractor task in a different modality than the target items and render identification of the stimulus likely (or required it) in the reduced (i.e., divided) attention condition.

Considering some of the differences between dual-task and selective attention tasks prompted Mulligan and Hornstein (2000) to propose a distractor-selection hypothesis, which suggests that selecting a distractor for response (or a stimulus dimension other than identity in the case of a single stimulus) disrupts encoding of the target stimulus (or its identity).[4] This hypothesis was motivated by the central-bottleneck model of Pashler (1994, 1998), a model based on evidence from the psychological-refractory-period (PRP) methodology. This theory proposes that memory encoding requires a central (amodal) bottleneck process, a bottleneck that also subserves response selection and memory retrieval (see Pashler 1994; 1998, for reviews). Consistent evidence comes from Jolicoeur and colleagues' recent research on the attentional blink and on visual encoding tasks (e.g., Arnell & Jolicoeur 1999; Dell'Acqua & Jolicoeur 2000; Jolicoeur 1999). This view argues that selecting a response to a distractor is a source of disrupted memory encoding, and divided-attention effects should be most evident when the response to the distractor is contemporaneous with target encoding. Alternatively, if responses to the distractors are infrequent or do not occur during target encoding, little effect of the divided attention manipulation may be observed.

In contrast to the visual attention hypothesis, the distractor selection viewpoint proposes that there is no fundamental difference between intra- and cross-modal manipulations of attention. Rather, the differential effects are attributed to differences in response requirements and target-distractor synchrony that are typically confounded with cross-modal versus intra-modal manipulations (see Mulligan & Hornstein 2000, for discussion). Commonly-used cross-modal distractor tasks, such as digit monitoring, require infrequent responses and distractors are not presented in synch with target items. This combination would minimize the extent to which the cross-modal manipulations occupy the central processing bottleneck and disrupt target encoding. In commonly-used intra-modal manipulations, response to distractors is much more frequent and distractors are typically presented at the same time as targets. Under the central-bottleneck view, these tasks are more likely to occupy the central bottleneck during target encoding. Thus, this account implies that there is no difference between intra- and cross-modal manipulations of attention when they are equated on frequency of response selection and target-distractor synchrony, making the modality of the secondary task irrelevant to divided-attention effects in perceptual priming.

## 4.2 Recent experiments on attention and perceptual implicit memory

In recent research, we investigated the role of attention in perceptual priming in light of the issues discussed above. Earlier studies have produced conflicting results, but these studies employed disparate materials, implicit tests, and procedures. It is important to investigate these issues with a common set of procedures and a constant implicit testing procedure. The present experiments used perceptual identification as the measure of perceptual priming for three reasons. First, it exemplifies the class of perceptual implicit tests (see Roediger & McDermott 1993, for a review). Second, it has produced conflicting results, in some cases being affected by divided attention and in others not. Third, this task has higher reliability than other implicit tests (Buchner & Wippich 2000), indicating that it is more likely to be sensitive to encoding manipulations.

### 4.2.1 *Intra-modal division of attention with Stroop-like materials (Mulligan & Hornstein 2000)*

The first series of experiments (Mulligan & Hornstein 2000) focused on the stimulus identification and distractor-selection accounts. In these experiments, a Stroop-like manipulation was used to manipulate attention to the identity of study words. Specifically, in Mulligan and Hornstein's (2000) Experiment 1, study words were presented in colored print under one of four encoding conditions. In the *read* (or full attention) condition, participants attended to the word's identity (reading each word aloud). In the *color* condition, participants attended to print color rather than identity (naming the color print for each word). In the *both* condition, participants' attention was divided between the word identity and color: participants read the word *and* pressed a key to in-

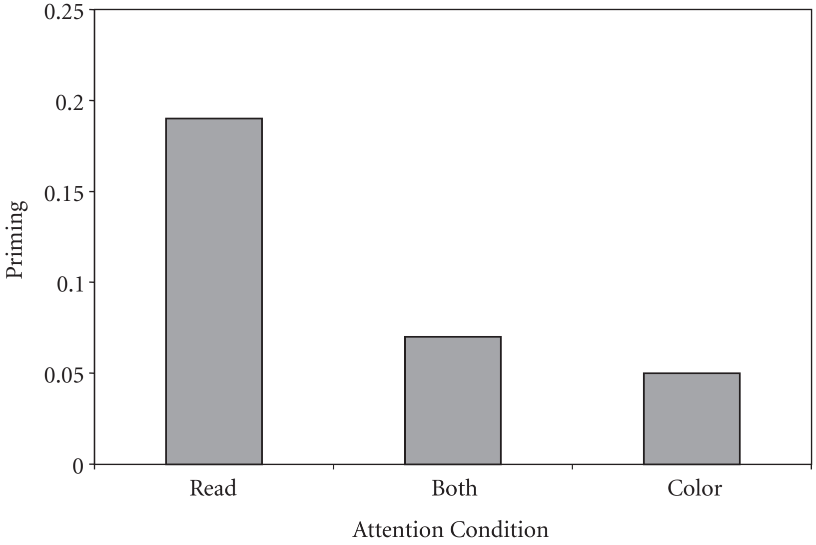


**Figure 3.**

dicate color. Thus, in the *both* condition, overt identification is coupled with intra-modal division of attention. Based on prior research (e.g., Stone et al. 1998), the *read* condition should lead to greater priming than the *color* condition because the selective attention manipulation in the *color* condition typically reduces perceptual priming. The critical comparison involved the *both* condition. If the only requirement for intact perceptual priming is complete stimulus identification, then priming should be equal in the *read* and *both* conditions. Alternatively, if perceptual priming is disrupted by responding to a distractor dimension of the stimulus, then priming should be lower in the *both* than the *read* condition.

Figure 3 presents the results of this experiment. The *read* condition produced significantly more priming that the *color* or *both* conditions, and the latter conditions did not differ from one another. The difference between the *read* and *color* conditions replicates prior findings that manipulations of selective attention reduce perceptual priming. The critical contrast reveals less priming in the *both* than *read* condition, showing that priming is reduced in the *both* condition even though the identity of each study word was fully processed (i.e., the words were successfully read aloud). This argues against the stimulus identification account, although it is consistent with either the visual-attention or distractor-selection hypotheses.

Two additional experiments replicated this central finding in somewhat different ways. First, there was one ambiguous aspect of the above results. In the *both* condition, participants typically read the word aloud before identifying the color with a keypress. Since it is possible that effects of distractor selection are eliminated if identity is the last, rather than first, stimulus attribute processed, Mulligan and Hornstein (2000, Experiment 3) attempted to determine if the reduced perceptual priming in the *both* condition was caused by the sequencing of responses. This experiment consisted of three encoding conditions: the *read* condition plus two variants of the *both* condition. In one of the *both* conditions, participants were required to read the word first before identifying its color; in the other *both* condition, participants identified the color before reading the word. The results indicated that the order of responding was not important: the two divided attention conditions produced significantly less priming than the read condition and they did not differ from each other.

A second follow-up experiment revealed that the same results obtained when attention was divided across objects rather than across dimensions of a single object. As an attentional manipulation, the Stroop-like task discussed above is unusual because it manipulates attention across dimensions of a single perceptual object. Studies of attention and memory typically divide attention across distinct perceptual objects (e.g., shadowing, digit monitoring, tone monitoring, etc.). In line with the more traditional attentional manipulation, Mulligan and Hornstein (2000, Experiment 4) presented study words flanked by two colored blocks. At study, participants either read the word, identified the color of the blocks, or did both. The *read* condition yielded greater priming than the other two conditions, which did not differ from one another. In summary, dividing attention with this Stroop-like manipulation reduced later perceptual priming even when the word was overtly identified, a result that held whether word identity was processed first or last, and whether attention was divided across different perceptual objects or across dimensions of a single object.

### 4.2.2 *Intra-modal vs. cross-modal division of attention*

The foregoing results indicate that effects of attention on perceptual priming are not simply tied to word identification. In these experiments, the distractor (or distractor dimension) and the targets were presented simultaneously and in the same modality. In addition, the distractors required overt responses on each trial. Consequently, both the visual-attention and distractor-selection hypotheses remained tenable and subsequent experiments were designed to contrast these hypotheses (Mulligan 2001). In these experiments, matched intra-modal and cross-modal manipulations of attention were implemented and perceptual implicit memory was again assessed with perceptual identification. In the divided attention conditions, overt identification of the study words was required to isolate potential effects of the attentional manipulations from the effects of disrupted stimulus identification.

First, a preliminary experiment was conducted, using the digit-monitoring task to divide attention. This is a standard attention manipulation: cross-modal (auditory distractor; visual target), with targets and distractors presented asynchronously and the overt response to distractors being relatively infrequent (e.g., Craik 1982; Gabrieli et al. 1999; Jacoby 1991; Jacoby et al. 1989; Fernandes & Moscovitch 2000; Mulligan & Hartman 1996; Schmitter-Edgecombe 1999). As noted above, other cross-modal divided-attention tasks leave perceptual priming unaffected, and such was expected here. Half the participants encoded words under full attention conditions, and half under divided-attention conditions (simultaneously carrying out the digit monitoring task). The mean perceptual priming did not differ between the full (.17) and divided (.16) attention groups. Separate groups of participants given a recognition test confirmed that the present divided attention task reduced explicit memory. This demonstrates that with the present materials, procedures and implicit test, the typical dissociation is obtained: cross-modal division of attention dramatically affects an explicit memory but not a perceptual implicit test.[5]

In the next experiment, different versions of the digit-monitoring task were developed so that the digits were presented either aurally (the typical, cross-modal division) or visually (atypical, intra-modal division). In addition, the presentation of the distractor digits was synchronized with the onset of the study word. In the visual version, each digit was presented in four locations surrounding the study word: above, below, to the left, and to the right. In the auditory version, each digit was presented over the computer's speakers. In both cases, the study word was synchronized with the distracting digit (see Mulligan 2001, for details). According to the distractor selection hypothesis, target-distractor synchrony should produce divided-attention effects regardless of the intra- or cross-modal format. According to the visual attention hypothesis, any divided-attention effect should only occur in the intra-modal condition.

Both divided attention conditions produced significant decrements in perceptual priming (auditory =.06; visual = .07) relative to full attention (.17), and the two divided attention conditions did not differ from each other. These results imply two things. First, the standard digit-monitoring task can disrupt perceptual priming when targets and distractors are presented synchronously. Second, cross-modal division of attention can disrupt perceptual priming, even when all study words are overtly identified. Given that this disruption is equivalent in the auditory and visual divided-attention tasks, the visual attention hypothesis is not supported. Rather the results support the distractor-selection hypothesis, and its corollary that the type of memory encoding that supports perceptual priming relies on a central (amodal) bottleneck process.

If responses to the distractor task are infrequent or do not occur during target encoding, then, according to the distractor-selection hypothesis, divided attention is less likely to compete for the central-process bottleneck during memory encoding and is thus less likely to produce an effect. The prior experiment examined this issue by modifying the traditional digit-monitoring task to produce target-distractor synchrony. However, the traditional digit-monitoring task requires relatively infrequent response selection. An overt response need only be made after a string of three odd digits is detected, which may only occur every third or fourth trial (see Mulligan 2001, for details).

The next experiment assessed the effects of intra- and cross-modal division of attention, and target-distractor synchrony, when distractors require a response on each trial. In the divided-attention conditions of this experiment, a digit was presented on each trial and an odd-even decision was required for each digit. In this experiment, each study word was presented for 1000 ms and the distracting digit was either presented synchronously (as in the prior experiment) or consistently offset in time by 500 ms. The goal of this manipulation was to determine if synchrony is critical under conditions of frequent response selection. Thus, the four divided attention conditions constituted a 2 x 2 design, with modality of the distractor task (visual vs. auditory) and target-distractor synchrony (synchronous vs. offset) as between-subjects manipulations. A fifth group encoded the words under full attention conditions. A perceptual identification test revealed that priming was significantly higher in the full attention condition (.16) than in any of the divided-attention conditions, and that these latter conditions did not differ (mean priming between .05 and .07). Thus, all forms of divided-attention reduced priming. This is again inconsistent with the visual-attention hypothesis, because auditory division of attention produced as large an effect on perceptual priming as visual division of attention. The results also imply that when the distractor task requires frequent response selection, the deleterious effects of divided attention do not depend on exact synchrony between presentation of targets and distractors.

A final experiment demonstrates that the prior results generalize to visual and auditory distractor tasks employing different classes of distracting stimuli. Divided attention effects sometimes depend on the relationship between the specific type of distractor and target items (e.g., Fernandes & Moscovitch

2000), and to draw stronger inferences about cross-modal attentional effects, it is critical to demonstrate that the cross-modal effects are due to central processes (e.g., response selection) (e.g., Arnell & Jolicoeur 1999; Pashler 1994, 1998) and not to peripheral interference or access to common representational systems (Fernandes & Moscovitch 2000).

To evaluate the possibility that the disrupted memory encoding revealed in the prior experiments is produced by the particular combination of targets (words) and distractors (digits), new visual and auditory distractor tasks of comparable difficulty were used (see Mulligan 2001, for a complete description of these tasks and evidence on their comparability). In the auditory version, each study word was accompanied by two successive tones that either went up, went down, or stayed the same frequency. In the visual version, each word was accompanied by flanking triangles (e.g., ì train ó). An initial pair of the flankers appeared (for 200 ms), was removed (for 100 ms), and then a second pair appeared (for 200 ms) either one line above, one line below, or on the same line as the initial pair of flankers. During this entire time, the word remained on the screen in the same position. On each trial in the divided attention conditions, participants named the word aloud and indicated (by pressing an appropriately-labeled key) whether the tones/flankers went up, went down, or stayed the same. Consistent with the earlier experiments, priming in the full-attention condition (.18) was significantly greater than in either of the divided-attention conditions which did not differ from each other (.05 in both conditions).[6]

There are several aspects of the present results to consider. First, dividing attention either within or across modalities reduced perceptual priming. Auditory and visual distractor tasks of equal difficulty produced comparable effects on perceptual identification, a result consistent with the prior experiments in which distractors presented in sync with targets reduced later priming. Second, the present results indicate that cross-modal divided attention effects generalize across different distractor stimuli. Since it is unlikely that tones produce peripheral interference with visually-presented words or compete for the same representational system (e.g., Arnell & Jolicoeur 1999), these results argue against domain-specific explanations of the cross-modal divided-attention effect found here and in the earlier experiments.

### 4.3 Discussion of experiments on attention and perceptual implicit memory

The foregoing experiments provide important constraints on hypotheses concerning the effects of attentional manipulations on perceptual priming. Many of the prior selective attention manipulations have rendered study word identification unlikely (e.g., Bentin et al. 1998; Crabb & Dark 1999; Eich 1984; Hawley & Johnston 1991; Kinoshita 1995; MacDonald & MacLeod 1998), thus raising the possibility that attentional manipulations disrupted perceptual priming to the extent that they disrupted study word identification at encoding. This possibility fits comfortably with the TAP account of perceptual priming (Roediger 1990; Roediger & McDermott 1993) as well as the view that priming requires lexical access of the study words (Bentin et al. 1998; Hayman & Jacoby 1989; Weldon 1991; Weldon & Jackson-Barrett 1993). Both views suggest that truncated perceptual analyses of the study words decrease later priming and that complete lexical analysis of the words is sufficient for intact priming. However, the present set of results are inconsistent with these views. Specifically, in the Stroop manipulation, the *both* condition required study word identification and yet reduced priming relative to the *read* condition. Likewise, the divided attention conditions of the last set of experiments all required that study words be read aloud, guaranteeing lexical access, and in all cases but one (the standard digit-monitoring task, with target-distractor asynchrony) these conditions reduced priming relative to the full attention condition. Thus, disrupted stimulus identification cannot account for the present effects of divided attention (cf. Mulligan & Hornstein 2000).

Related to the stimulus-identification hypothesis, Stone et al. (1998) argued that some amount of automatic identity processing occurs even in the *color* condition of the Stroop task (see also MacLeod 1991). Because color naming produces less priming than reading, Stone et al. suggest that perceptual priming requires an awareness of the study stimulus during encoding, over and above any automatic identity processing that occurs in the *color* condition. If this is the case, then requiring participants to overtly identify study words (and hence become more fully aware of them) should eliminate the effects of divided attention in the Stroop manipulation. This awareness-at-encoding view predicts that the *both* condition should lead to greater priming than the *color* condition. In addition (and like the stimulus-identification view), this view suggests that the *both* and *read* conditions will lead to comparable levels of perceptual priming to the extent that participants are equally aware of the identity of the stimulus in these two conditions. These results were not obtained, and thus it appears that mere awareness of the word at study is not sufficient to produce full levels of priming.

The present results are also inconsistent with the visual attention hypothesis. This hypothesis suggests that perceptual priming relies on modality-specific components of attention and not on general or amodal attentional resources. This hypothesis is consistent with much of the earlier research using intra-modal selective attention manipulations and with the Stroop manipulation which divides attention within the visual modality. In these cases, presenting target words and distractors within the visual modality resulted in reductions in perceptual priming. However, this hypothesis also predicts that cross-modal division of attention should have minimal effect on perceptual priming, or at least less of an effect than comparable intra-modal manipulations. This expectation is contradicted by the second set of experiments. Specifically, comparable intra-modal and cross-modal division of attention produced comparable reductions in perceptual priming relative to full attention conditions. These results, which generalized over different distractor tasks and different classes of distractors, are inconsistent with the visual attention hypothesis and suggest a central rather than modality-specific locus for these effects.

The results are consistent with the distractor-selection hypothesis, a hypothesis motivated by the central-bottleneck model of Pashler (1994, 1998; see also, Arnell & Jolicoeur 1999; Dell'Acqua & Jolicoeur 2000; Jolicoeur 1999). This view argues that response selection to a distractor disrupts memory encoding. When distractors require frequent responses and when distractors and targets are presented simultaneously, then the process of selecting a response to the distractor is most likely to disrupt memory encoding of the target. Alternatively, if responses to the distractors are infrequent and do not occur during target encoding, little effect of the divided attention manipulation may be observed. This perspective proposes that there is no fundamental difference between intra- and cross-modal manipulations of attention. Rather, any differences in their effects are attributed to differences in response requirements that are typically confounded with cross-modal versus intra-modal manipulations. Commonly-used cross-modal distractor tasks often require less frequent responses and distractors are not presented in synch with target items, thus minimizing disruption of target encoding. In commonly-used intra-modal manipulations, response to distractors is much more frequent and distractors are typically presented at the same time as targets. According to the central-bottleneck view, these tasks are more likely to occupy the central bottleneck during target encoding.

The present series of studies clearly demonstrates that when comparable auditory and visual distractor tasks were implemented, they produced the same reduction in perceptual priming relative to full attention conditions. This replicated across different types of distractor tasks, indicating that the results are not domain-specific and strengthening the case that a central (amodal) bottleneck is the basis of the divided-attention effect.

## 5. Summary and general implications

The research reviewed in this chapter has several important implications for theoretical views about the relationship between attention and implicit memory. First, attentional manipulations can affect both conceptual and perceptual implicit memory, contrary to the notion that implicit memory reflects automatic encoding processes (e.g., Bentin et al. 1995; Graf & Mandler 1984; Isingrini et al. 1995; Jacoby et al. 1993; Jacoby et al. 1989; Jelicic et al. 1992; Parkin et al. 1990; Parkin & Russo 1990; Shallice et al. 1994; Szymanski & MacLeod 1996; Toth 2000). With regard to conceptual priming, it is now clear that sufficiently difficult distractor tasks reduce conceptual priming, although there is still some question about whether this result generalizes to all conceptual tasks, such as category verification (Gabrieli et al. 1999). In addition, the present research joins a number of other studies in demonstrating that perceptual implicit tests are affected by attentional manipulations (e.g., Bentin et al. 1998; Crabb & Dark 1999; Stone et al. 2000) and that cross-modal division of attention can reduce perceptual priming even when the study words are overtly identified.

The TAP view, which framed many of the present inquires, suggests that conceptual priming should be highly reliant on attention because conceptual and elaborative processes require attention. On the other hand, to the extent that the attentional manipulation allows perceptual processing of the study items, perceptual priming should be relatively unaffected. The latter view gave rise to the stimulus identification hypothesis, which was shown to be inadequate. With regard to perceptual priming, attention manipulations (both intra- and cross-modal) that do not prevent stimulus identification may still reduce perceptual priming. With respect to conceptual priming, much of the research comports with the TAP expectation, with the exception of the category-verification test.

As an alternative to the TAP view, Gabrieli et al. (1999) argued for a distinction between production and identification priming tasks, a distinction based on neurological and functional dissociations (see also, Vaidya et al. 1997). In production tasks (e.g., word-stem completion, word association, category-exemplar production), test cues do not define a unique target but merely delimit a class of possible targets, thus inducing retrieval competition among multiple possible responses. In identification tasks (e.g., word or picture naming, lexical decision, category verification, perceptual identification), participants merely identify a test stimulus and are assessed on speed or accuracy. In these tasks, the retrieval cue uniquely determines retrieval of a single appropriate response. Because there is a single correct answer for each retrieval cue, these tests are characterized as non-competitive.

Gabrieli et al. (1999; Vaidya et al. 1997) argued that production but not identification retrieval tasks are sensitive to division of attention during encoding. Vaidya et al. (1997) suggested that simply accessing the target item during encoding would result in full priming in identification tasks (similar to the lexical-access view). On the other hand, production priming tasks benefit from elaboration beyond stimulus identification at encoding. According to this view, division of attention disrupts elaboration, which reduces subsequent production but not identification priming.

This view accounts for the effects of divided attention on conceptual priming tasks (reviewed earlier). Category-exemplar production, word association and general-knowledge questions are production tasks, whereas category verification is an identification task. Fittingly, divided attention affects conceptual priming in the former tests but not the latter. However, the present experiments on attention and perceptual priming are not consistent with this retrieval distinction. Because perceptual identification is an identification task, with a single correct response for each test cue, this task should be insensitive to divided attention manipulations, especially cross-modal division of attention (Gabrieli et al. 1999). However, the Stroop manipulations and auditory and visual distractor tasks reduced priming in this task, even though the representations of the study words were fully accessed during encoding (as evidenced by successful identification) in all cases. This raises the possibility that other identification priming tests that heretofore have been unaffected by cross-modal division of attention, such as the category verification task (Gabrieli et al. 1999), would be affected by manipulations that require response selection to a distractor at the same time as study word encoding.

The distractor-selection hypothesis, based on the notion of a central processing bottleneck (Pashler 1994, 1998), might form the basis of an account of the present research. This view proposes that all forms of memory encoding are subject to the central-processing bottleneck. In contrast to the TAP view and the production-identification distinction, there should be no implicit test immune to divided-attention effects, provided that the divided-attention task requires frequent response selection close in time to memory encoding. However, this account requires further specification to be useful and a potential direction, which borrows some elements from the production-identification distinction, is proposed below.

As detailed thus far, the distractor selection hypothesis accounts for the effects of different attentional manipulations on a perceptual priming task. However, further detail is necessary to account for the observation that a single attentional manipulation may produce different effects on different types of memory tests. For instance, why does the standard (i.e., cross-modal, asynchronous) digit-monitoring task reduce performance on explicit and conceptual implicit tests but not on perceptual implicit tests such as word-fragment completion or perceptual identification (Mulligan 1998)? Perhaps, memory tests that rely more heavily on strategic encoding processes (e.g., elaborative and organization) are more sensitive to manipulations of attention because such strategic encoding processes require frequent memory retrieval and re-encoding, often in a structured sequence. In terms of the central-bottleneck hypothesis, these strategies rely so heavily on the bottleneck that even intermittent competition from infrequent response requirements of the digit-monitoring task, may be sufficient to disrupt these encoding strategies. While speculative, this account is consistent with the finding that memory tests that rely heavily on elaborative encoding processes make greater use of central resources during encoding (e.g., Craik et al. 1996). This would fit with the present results because the tests that are least sensitive to divided-attention effects (category-verification, lexical decision, perceptual identification) are also least sensitive to variations in elaboration, such as the levels-of-processing manipulation (e.g., Roediger & McDermott 1993; Vaidya et al. 1997).

## Notes

* This research was supported in part by grant 1-R03-MH61324-01 from NIMH. Address correspondence to either Neil Mulligan or Alan Brown, Department of Psychology, Southern Methodist University, Dallas, TX 75275-0442, or via email to either mulligan@mail.smu.edu or abrown@mail.smu.edu

[1] The retrieval requirements of the category-exemplar production and category-cued recall tests are typically not equivalent. In the former test, participants are instructed to produce a number of responses to each category (eight, in Mulligan & Hartman 1996), whereas in

the latter test the number of responses is under the control of the participant and is usually relatively small (perhaps as few as one of two examples per category). To determine whether the difference in response requirements is important, Mulligan and Hartman (1996) also used a modified-category-cued recall test which required an equivalent number of responses as the conceptual implicit test. In particular, participants were asked to use each category cue to remember as many studied words as possible. Then, when they could no longer remember words from the study list, they were to freely generate examples from the category until they reached a total of eight examples either recalled or generated. The results from the modified recall task were identical to the typical recall task, and only the latter are discussed here.

**2.** If Isingrini et al.'s (1995) was a weaker manipulation of attention and if attention is necessary for later conceptual priming, we might expect a small (though possibly non-significant) effect of this manipulation. In fact, the divided attention condition produced somewhat less priming than the full attention condition. Considering all four of the age groups in Isingrini et al. each age group exhibited a small (but non-significant) effect of divided attention. By a sign test, the Null Hypothesis of no effect of attention yields $p = .0625$ (one-tailed), suggesting that differences in the strength of the attention manipulation may account for the conflicting results.

**3.** There is also conflicting evidence regarding the word-stem completion test, with some studies finding divided-attention effects (Clarys et al. 2000; Gabrieli et al. 1999) and others not (Wippich, Melzer, & Mecklenbrauker 1998). However, the nature of this implicit test is in dispute; some researchers argue that this test is primarily perceptual and others, primarily conceptual (cf. Gabrieli et al. 1994; Keane, Gabrieli, Fennema, Growdon, & Corkin 1991; Roediger et al. 1992).

**4.** A second possible role for selection focuses on potential positive effects of target selection rather than on negative effects of distractor selection. Specifically, selecting the target stimulus (or, more precisely, the identity of the target stimulus) for response may be required for full priming (Wippich 1995). This hypothesis implies that passive perception of the stimulus is not sufficient for maximal priming. In its most straightforward form, this hypothesis amounts to the stimulus identification hypothesis in that successful selection and response typically implies stimulus identification. The only question is whether the response must be overt (e.g., naming). The stimulus identification hypothesis, relying on the notion of lexical access, states that identification may be covert. Prior research argues against a role for overt naming (versus silent reading) in later perceptual priming (Downes et al. 1996; MacDonald & MacLeod 1998). Consequently, the most tenable target-selection hypothesis appears equivalent to the stimulus identification hypothesis.

**5.** It is preferable to use implicit and explicit tests matched on retrieval cues and response requirements when the researching the relative effects of divided attention on implicit and explicit memory (Mulligan 1998). This is feasible for implicit tests such as word-fragment completion or category-exemplar production. Explicit cued-recall tests can be created by pairing the retrieval cues and response requirements (i.e., production of target items) with explicit test instructions. This renders a design capable of satisfying the retrieval-intentionality criterion (see Figure 1 for an example; Mulligan & Hartman 1996; Schacter, Booker, & Bowers 1989). However, the perceptual identification task does not naturally lend itself to an explicit counterpart. Recognition memory is the most reasonable comparison because it uses the same retrieval cues (i.e., whole words) even though it does not require the same type of response. Recognition memory is also a good choice because it is less sensitive to divided attention than many other explicit tests, such as free or cued recall (e.g., Craik et al. 1996). Consequently, recognition produces a conservative measure of the impact of divided attention on explicit memory, and dissociations between recognition and implicit tests are less likely to be produced by quantitative differences in sensitivity.

**6.** This experiment also included full-attention control conditions, in which the distractors were presented but did not require response. In these conditions, participants were told to read the words aloud and ignore the accompanying tones or flankers. These full attention conditions also led to significantly greater priming than the divided attention conditions, implying that the mere presence of the distractors does not produce the divided-attention effect in perceptual priming.

## References

Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 631–682). Cambridge, MA: MIT Press.

Arnell, K. M. & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 630–648.

Baddeley, A. (1986). *Working Memory*. Oxford: Clarendon Press.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113*, 518–540.

Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Bentin, S. (1994). Semantic processing of ignored stimuli: The role of attention in memory. In C. Umilta and M. Moscovitch (Eds.), *Attention and Performance* (Vol. 15, pp. 551–569). Cambridge, MA: MIT Press.

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 54–67.

Bentin, S., Moscovitch, M., & Nirhod, O. (1998). Levels of processing and selective attention effects in encoding in memory. *Acta Psychologica, 98*, 311–341.

Besson, M., Fischler, I., Boaz, T., & Raney, G. (1992). Effects of automatic associative activation on explicit and implicit memory test. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 89–105.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 657–668.

Blaxton, T. A. (1992). Dissociations among memory measures in memory-impaired subjects: Evidence for a processing account of memory. *Memory and Cognition, 20*, 549–562.

Broadbent, D. E. (1958). *Perception and Communication*. London: Pergamon Press.

Broadbent, D. E. (1971). *Decision and Stress*. San Diego, CA: Academic Press.

Buchner, A. & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40*, 227–259.

Challis, B. H., Chiu, C.-Y., Kerr, S. A., Law, J., Schneider, L., Yonelinas, A., & Tulving, E. (1993). Perceptual and conceptual cueing in implicit and explicit retrieval. *Memory, 1*, 127–151.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America, 26*, 554–559.

Clarys, D., Isingrini, M., & Haerty, A. (2000). Effects of attentional load and ageing on word-stem and word-fragment implicit memory tests. *European Journal of Cognitive Psychology, 12*, 395–412.

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. New York: Oxford University Press.

Crabb, B. & Dark, V. (1999). Perceptual implicit memory requires attentional encoding. *Memory & Cognition, 27*, 267–275.

Craik, F. I. M. (1982). Selective changes in encoding as a function of reduced processing capacity. In F. Klix, J. Hoffman, and E. van der Meer (Eds.), *Cognitive Research in Psychology* (pp. 152–161). Berlin: Verlag.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society, Series B 302*, 341–359.

Craik, F. I. M. (1989). On the making of episodes. In H. L. Roediger and F. I. M. Craik (Eds.), *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving* (pp. 43–57). Hillsdale, New Jersey: Erlbaum.

Craik, F. I. M. & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik and S. E. Trehub (Eds.), *Aging and Cognitive Processes* (pp. 191–211). New York: Plenum Press.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*, 159–180.

Craik, F. I. M., Moscovitch, M., & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 864–875.

Curran, H. V. (2000). Psychopharmacological perspectives on memory. In E. Tulving and F. I. M. Craik (Eds.), *The Oxford Handbook on Memory* (pp. 539–554). New York: Oxford University Press.

Dell'Acqua, R. & Jolicoeur, P. (2000). Visual encoding of patterns is subject to dual-task interference. *Memory & Cognition, 28*, 184–191.

Denny, E. B. & Hunt, R. R. (1992). Affective valence and memory in depression: Dissociation of recall and fragment completion. *Journal of Abnormal Psychology, 101*, 575–580.

Downes, J. J., Davies, E. J., Davies, P., Perfect, T. J., Wilson, K., Mayes, A. R., & Sagar, H. J. (1996). Stem-completion priming in Alzheimer's disease: The importance of target word articulation. *Neuropsychologia, 34*, 63–75.

Duncan, J. (1999). Converging levels of analysis in the cognitive neuroscience of visual attention. In G. W. Humphreys, J. Duncan, and A. Treisman (Eds.), *Attention, Space, and Action: Studies in Cognitive Neuroscience* (pp. 112–129). New York: Oxford.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature, 379*, 808–810.

Ebbinghaus, H. (1964/1885). *Memory: A Contribution to Experimental Psychology*. New York: Dover.

Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory and Cognition, 12*, 105–111.

Elliott, C. L. & Greene, R. L. (1992). Clinical depression and implicit memory. *Journal of Abnormal Psychology, 101*, 572–574.

Engle, R. W., Conway, A. R. A., Tuholski, S. W., & Shisler, R. J. (1995). A resource account of inhibition. *Psychological Science, 6*, 122–125.

Fernandes, M. A. & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects both at retrieval and encoding. *Journal of Experimental Psychology: General, 129*, 155–176.

Fisk, A. D. & Schneider, W. (1984). Memory as a function of attention, level of processing, and automatization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 181–197.

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49*, 87–115.

Gabrieli, J. D. E., Keane, M. H., Stange, B. Z., Kjelgaard, M. M., Corkin, S., & Growdon, J. H. (1994). Dissociations among structural-perceptual, lexical-semantic, and event-fact memory systems in Alzheimer, amnesic, and normal subjects. *Cortex, 30*, 75–103.

Gabrieli, J. D. E., Vaidya, C. J., Stone, M., Francis, W. S., Thompson-Schill, S. L., Fleischman, D. A., Tinklenberg, J. R., Yesavage, J. A., & Wilson, R. S. (1999). Convergent behavioral and neuropsychological evidence for a distinction between identification and production forms of repetition priming. *Journal of Experimental Psychology: General, 128*, 479–498.

Gellatly, A., Parker, A., Blurton, A., & Woods, C. (1994). Word stem and word fragment completion following semantic activation and elaboration. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1099–1107.

Graf, P. & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23*, 553–568.

Hamann, S. B. (1990). Level-of-processing effects in conceptually driven implicit tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 970–977.

Hawley, K. J. & Johnston, W. A. (1991). Long-term perceptual memory for briefly exposed words as a function of awareness and attention. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 807–815.

Hayman, C. A. G. & Jacoby, L. L. (1989). Specific word transfer as a measure of processing in the word- superiority paradigm. *Memory & Cognition, 17*, 125–133.

Heuer, H. (1996). Dual-task performance. In O. Neumann and A. F. Sanders (Eds.), *Handbook of Perception and Action, Vol 3: Attention* (pp. 113–153). London: Academic Press.

Isingrini, M., Vazou, F., & Leroy, P. (1995). Dissociation between implicit and explicit memory tests: Effects of age and divided attention on category exemplar generation and cued recall. *Memory and Cognition, 23*, 462–467.

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

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122*, 1–16.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by divided attention. *Journal of Experimental Psychology: General, 118*, 115–125.

James, W. (1890). *The Principles of Psychology*. London: MacMillan.

Jelicic, M., Bonke, B., Wolters, G., & Phaf, R. H. (1992). Implicit memory for words presented during anaesthesia. *European Journal of Cognitive Psychology, 4*, 71–80.

Johnston, J. C., McCann, R. S., & Remington, R. W. (1995). Chronometric evidence for two types of attention. *Psychological Science, 6*, 365–369.

Jolicoeur, P. (1999). Dual-task interference and visual encoding. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 596–616.

Keane, M. H., Gabrieli, J. D. E., Fennema, A. C., Growdon, J. H., & Corkin, S. (1991). Evidence for a dissociation between perceptual and conceptual priming in Alzheimer's disease. *Behavioral Neuroscience, 105*, 326–342.

Kellogg, R. T., Newcombe, C., Kammer, C., & Schmitt, K. (1996). Attention and direct and indirect memory tasks with short- and long-term probes. *American Journal of Psychology, 109*, 205–217.

Kinoshita, S. (1995). The word frequency effect in recognition memory versus repetition priming. *Memory & Cognition, 23*, 569–580.

Koelega, H. S. (1996). Sustained Attention. In O. Neumann and A. F. Sanders (Eds.), *Handbook of Perception and Action, Vol 3: Attention* (pp. 277–331). London: Academic Press.

Kolers, P. A. & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior, 23*, 425–449.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42*, 333–376.

Light, L. L. & Prull, M. W. (1995). Aging, divided attention, and repetition priming. *Swiss Journal of Psychology, 54*, 87–101.

Light, L. L., Prull, M. W., & Kennison, R. F. (2000). Divided attention, aging, and priming in exemplar generation and category verification. *Memory & Cognition, 28*, 856–872.

MacDonald, P. A. & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98*, 291–310.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

Mandler, G., Nakamura, Y., & Zandt, B. J. S. V. (1987). Nonspecific effects of exposure on stimuli that cannot be recognized. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 646–648.

Merikle, P. & Reingold, E. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 224–233.

Mitchell, D. B., Brown, A. S., & Murphy, D. R.. (1990). Dissociations between procedural and episodic memory: Effects of time and aging. *Psychology and Aging, 5*, 264–276.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*, 56–60.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.

Mulligan, N. W. (1997). Attention and implicit memory: The effects of varying attentional load on conceptual priming. *Memory and Cognition, 25*, 11–17.

Mulligan, N. W. (1998). The role of attention during encoding on implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 27–47.

Mulligan, N. W. (2001). *Effects of cross-modal and intra-modal division of attention on perceptual priming*. Manuscript submitted for publication.

Mulligan, N. W. & Hartman, M. (1996). Divided attention and indirect memory tests. *Memory and Cognition, 24*, 453–465.

Mulligan, N. W. & Hornstein, S. L. (2000). Attention and perceptual priming in the perceptual identification task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 27–47.

Mulligan, N. W. & Stone, M. (1999). Attention and conceptual priming: Limits on the effects of divided attention in the category-exemplar production task. *Journal of Memory and Language, 41*, 253–280.

Murdock, B. B. (1965). Effects of subsidiary task on short-term memory. *British Journal of Psychology, 56*, 413–419.

Norman, D. A. (1969). *Memory & Attention: An Introduction to Human Information Processing*. New York: Wiley and Sons, Inc.

Parkin, A. J., Reid, T. K., & Russo, R. (1990). On the differential nature of implicit and explicit memory. *Memory and Cognition, 18*, 507–514.

Parkin, A. J. & Russo, R. (1990). Implicit and explicit memory and the automatic/effortful distinction. *European Journal of Cognitive Psychology, 2*, 71–80.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220–244.

Pashler, H. E. (1998). *The Psychology of Attention*. Cambridge, MA: MIT Press.

Posner, M. I. & Boies, S. J. (1971). Components of attention. *Psychological Review, 78*, 391–408.

Posner, M. I. & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*, 1043–1056.

Roediger, H. L., Buckner, R. L., & McDermott, K. B. (1999). Components of processing. In J. K. Foster and M. Jelicic (Eds.), *Memory: Systems, Process, or Function* (pp. 31–65). New York: Oxford University Press.

Roediger, H. L. & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller and J. Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 8, pp. 63–131). Amsterdam: Elsevier.

Roediger, H. L., Weldon, M. S., Stadler, M. A., & Reigler, G. H. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1251–1269.

Rosen, V. M. & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211–227.

Russo, R. & Parkin, A. J. (1993). Age differences in implicit memory: More apparent than real. *Memory and Cognition, 21*, 73–80.

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

Schacter, D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience, 4*, 244–256.

Schacter, D. L. & Badgaiyan, R. D. (2001). Neuroimaging of priming: New perspectives on implicit and explicit memory. *Current Directions in Psychological Science, 10*, 1–4.

Schacter, D. L., Bowers, J., & Booker, J. (1989). Intention, awareness, and implicit memory: The retrieval intentionality criterion. In S. Lewandowsky, J. C. Dunn, and K. Kirsner (Eds.), *Implicit Memory: Theoretical Issues* (pp. 47–65). Hillsdale, NJ: Erlbaum.

Schmitter-Edgecombe, M. (1996). The effects of divided attention on implicit and explicit memory performance. *Journal of the International Neuropsychological Society, 2*, 111–125.

Schmitter-Edgecombe, M. (1996). Effects of divided attention on implicit and explicit memory performance following severe closed head injury. *Neuropsychology, 10*, 155–167.

Schmitter-Edgecombe, M. (1999). Effects of divided attention on perceptual and conceptual memory tests: An analysis using a process-dissociation approach. *Memory & Cognition, 27*, 512–525.

Schwartz, B. L., Rosse, R. B., & Deutsch, S. I. (1993). Limits of the processing view in accounting for dissociations among memory measures in a clinical population. *Memory and Cognition, 21*, 63–72.

Seamon, J., Marsh, R., & Brody, N. (1984). Critical importance of exposure duration for attentive discrimination of stimuli that are not recognized. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 465–469.

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S. J., & Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature, 368*(April), 633–635.

Shimamura, A. P. (1993). Neuropsychological analyses of implicit memory: History, methodology, and theoretical interpretations. In P. Graf and M. E. J. Masson (Eds.), *Implicit Memory: New Directions in Cognition, Development, and Neuropsychology* (pp. 265–285). Hillside, NJ: Erlbaum.

Smith, M. E. & Oscar-Berman, M. (1990). Repetition priming of words and psuedowords in divided attention and in amnesia. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 1033–1042.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195–231.

Stone, M., Ladd, S. L., & Gabrieli, J. D. E. (2000). The role of selective attention in perceptual and affective priming. *American Journal of Psychology, 113*, 341–358.

Stone, M., Ladd, S. L., Vaidya, C. J., & Gabrieli, J. D. E. (1998). Word identification priming for ignored and attended words. *Consciousness and Cognition, 7*, 238–258.

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

Toth, J. P. (2000). Nonconscious forms of human memory. In E. Tulving and F. I. M. Craik (Eds.), *The Oxford Handbook on Memory* (pp. 245–261). New York: Oxford University Press.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 40A*, 201–237.

Tulving, E. & Craik, F. I. M. (Eds.). (2000). *The Oxford Handbook on Memory*. New York: Oxford University Press.

Vaidya, C. J., Gabrieli, J. D. E., Keane, M. M., Monti, L. A., Gutierrez-Rivas, H., & Zarella, M. M. (1997). Evidence for multiple mechanisms of conceptual priming on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 1324–1343.

Weldon, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 526–541.

Weldon, M. S. & Jackson-Barrett, J. L. (1993). Why do pictures produce priming on the word-fragment completion test? A study of encoding and retrieval factors. *Memory and Cognition, 21*, 519–528.

Wickens, C. D. (1984). *Engineering Psychology and Human Performance*. Columbus, OH: Merrill.

Wippich, W. (1995). Priming on verbal perceptual tests: Roles of lexical, surface, and conceptual processes. *Psychological Research, 57*, 250–259.

Wippich, W., Melzer, A., & Mecklenbraeuker, S. (1998). Picture or word superiority effects in implicit memory: Levels of processing, attention and retrieval constraints. *Swiss Journal of Psychology, 57*, 33–46.

Wolters, G. & Prinsen, A. (1997). Full versus divided attention and implicit memory performance. *Memory & Cognition, 25*, 764–771.

Wood, N. L., Stadler, M. A., & Cowan, N. (1997). Is there implicit memory without attention? A re-examination of task demands in Eich's (1984) procedure. *Memory and Cognition, 25*, 772–779.

# The route from implicit learning to verbal expression of what has been learned

## Verbal report of incidentally experienced environmental regularity

Peter A. Frensch*, Hilde Haider, Dennis Rünger,
Uwe Neugebauer, Sabine Voigt and Jana Werg

## 1. Introduction

Two students, Rico and Dieter, participate in the same, rather boring, experiment. They are seated in front of a computer screen on which four screen locations are marked. The four locations are mapped individually to four response keys on the computer keyboard. On any trial, a symbol appears at one of the marked locations and the two students' task is simply to depress, as fast as they can, the response key that corresponds to the location at which the symbol appears.

Virtually any reader who is the least bit familiar with the concept of "implicit learning" has recognized by now that Rico and Dieter perform the so-called Serial Reaction Time Task (SRTT), a task originally introduced by Nissen and Bullemer (1987). In this task, unbeknownst to the two students, the sequence of spatial positions at which the symbols appear is pre-determined and fixed (e.g., Stadler & Frensch 1998).

Two important results are typically observed with the SRTT. First, participants, such as Rico and Dieter, learn, in some way, something about the deterministic character of the sequence of spatial locations. This can be demonstrated easily by introducing a random sequence of spatial positions at some point during task practice. The amount by which the response times of participants to this random sequence increases, relative to the response times to the deterministic sequence, can then be taken as a measure of how much of the deterministic sequence has been learned.

Second and more important in the context of the present chapter, between 10% and 70% of all participants in this task, with the exact percentage varying with experimental condition, are able to verbally describe the deterministic regularity built into the task when asked to do so in a post-experimental interview. Thus, when asked, Rico might be able to perfectly describe the sequence of spatial positions that makes up the deterministic regularity in the SRTT. Dieter, in contrast, might not be able to describe anything; in fact, he might maintain that no such regularity even existed.

The difference between Rico and Dieter in terms of their ability to verbally report the characteristics of the environmental regularity they were exposed to is the starting point for the theoretical ideas and empirical research discussed in this chapter. The main goal of the chapter is to present and discuss a recent theoretical framework that we advance, the Unexpected-Event Hypothesis, in order to explain how the ability to verbally report a regularity that is encountered in the context of an incidental learning situation develops.

More specifically, our first main objective is to describe (at least some of) the mechanisms that link the experience of an environmental regularity in an individual to the individual's ability to verbally report the regularity. The reader should notice that we use the SRTT situation described above merely as an example; the Unexpected-Event Hypothesis is not limited to the experimental tasks that are typically encountered in the implicit learning literature. More generally, the situation we are striving to explain is characterized by two features: (1) an individual experiences an environmental regularity in the absence of any explicit instruction to discover the regularity; and (2) the individual correctly and verbally reports the characteristics of the regularity when asked to do so.

Our second main objective is to explain individual differences in, and situational influences on, the ability to verbally report the characteristics of an implicitly experienced environmental regularity. In this regard, we are trying to provide first and tentative answers to questions such as (1) why is it that some participants, as for instance Rico, in any of the many experiments that have been performed using the SRTT, are able to correctly verbalize the deterministic sequence they experienced whereas others, such as Dieter, are not able to do so?; and (2) why is it that in some experimental conditions the number of participants who are able to correctly report the deterministic sequence is very high whereas in others it is quite low?

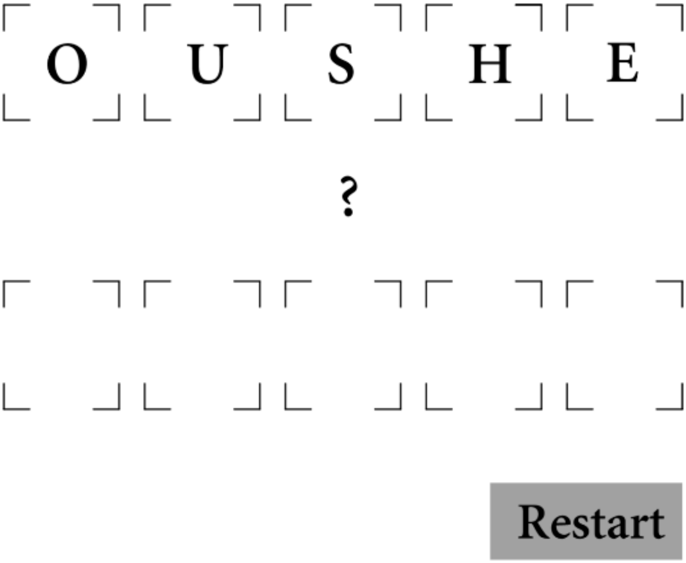


**Figure 1.** Set-up of the anagram experiment.

To provide an example of what kind of experimental situation and what kind of empirical findings is part of our second main focus, consider the set-up of a recent experiment (Haider & Frensch, in preparation). In this experiment, participants were asked to solve anagrams, such as the one depicted in Figure 1. We were primarily interested in the extent to which participants were able to verbally report, by the end of the experiment, a regularity that was built into the task. The regularity was that for all anagrams participants were asked to solve, the first letter of the correct solution word always appeared at Screen Position 4. Thus, the word "House" as the correct solution word to the anagram given in Figure 1, for instance, begins with an "H" and the "H" is located at the 4th spatial screen position from the left.

What we manipulated in this experiment was whether participants solved the anagrams by using a keyboard or by moving a mouse. In the keyboard condition, participants typed in the five spatial positions in the sequence of the solution word; in the mouse condition, they dragged the letters over the screen into the five boxes shown below the question mark in Figure 1. The main question of interest was whether the response manipulation (i.e., solving anagrams by using a keyboard versus by using a mouse) would affect the likelihood of participants' being able to correctly report the rule built into the task, that is, the rule that the first letter of the solution word always appeared at the 4th spatial screen position from the left.

The results of this little experiment were stunning: More than 60% of the participants in the keyboard condition were able to verbally describe the regularity built into the task when asked to do so at the end of the experimental session. By comparison, only slightly more than 10% of the participants in the mouse condition were able to do so. Later in the chapter we return to this particular experiment and offer an explanation for the surprising result.

The chapter is divided into four main parts. First, we briefly describe some existing theoretical accounts of how an individual might develop the ability to verbally report a task regularity that was encountered in the context of an incidental learning situation. Second, we present our own, still rather tentative, theoretical framework, the Unexpected-Event Hypothesis. Third, we describe some results of our recent empirical work that are consistent with the main assumptions of the hypothesis. Fourth, the chapter closes with a brief discussion of some of the broader implications of our empirical findings and theoretical framework.

## 2. Theoretical perspectives on the link between an incidental experience of an environmental regularity and the ability to verbally report the regularity

### 2.1 General views from outside the implicit learning domain

The classic explanation for why and when individuals are able to verbally report an experienced environmental regularity goes back to at least Wilhelm Wundt (e.g., Wundt, 1896). According to the classic explanation, verbal report of an experienced regularity is possible when the content of the memory trace representing the experienced regularity enters phenomenal awareness. The question of how and when the content of a memory trace enters phenomenal awareness has been traced from at least three different perspectives, neuropsychological, cognitive, and philosophical.

Farah (1994), for example, has argued that existing neuropsychological explanations of how phenomenal awareness comes about can be grouped into three distinct theoretical categories. First, in "Quality-of -Representation" explanations (e.g., O'Brien & Opie 1999), it is assumed that phenomenal awareness is a causal consequence of particular properties of neural representations, such as their strength or stability over time. In "Privileged-Role" explanations, by comparison, it is assumed that awareness depends on the activity of specific brain systems whose function it is to produce subjective phenomenal awareness. Third, in “Integration“ explanations, awareness is the result of specific processes, namely processes of integration through which the activity of different brain regions can be synchronized or made coherent so as to form the contents of subjective awareness.[1] (Some recent neuropsychological accounts of phenomenal consciousness, e.g., Dehaene & Naccache 2001, attempt to integrate features from the different categories suggested by Farah.)

Perhaps surprisingly, Farah’s (1994) three categories of neuropsychological explanations reflect cognitive views on awareness as well. For instance, the “Quality-of-Representation” account is arguably the most prevalent cognitive explanation of how phenomenal awareness is generated. Norman (1968), for example, argues that the strength of a memory representation determines whether or not the content of the representation enters awareness. Long-term memory representations that are highly activated, that is, that are very strong, are transferred into Working Memory. The content of Working Memory always enters awareness; thus, the strength of a memory representation, in essence, determines whether or not individuals are aware of its contents.

A similar view has been expressed by, among many others, Cowan (1995). According to Cowan’s model, Short-Term Memory consists of activated long-term memory representations. The content of Short-Term Memory does not necessarily enter awareness, however. Rather, awareness is the consequence of memory representations being in the focus of attention. Deployment of attention, then, is a necessary prerequisite for a memory representation to enter awareness.

## 2.2 Specific views from inside the implicit learning domain

In the recent past, several theoretical ideas have been proposed in the implicit learning literature that focus, often more indirectly than directly, on the link between an individual’s experience of an environmental regularity and the ability to verbally describe the regularity. At least three different views have been offered, two of which are reminiscent of two of the three neuropsychological accounts of awareness that have been discussed by Farah (1994). The first view encountered in the implicit learning literature holds that the acquisition of reportable knowledge is based on a learning mechanism that is devoted solely to the generation of reportable knowledge (this is essentially the “Privileged-Role” account described by Farah 1994). According to this view, participants in a typical implicit learning experiment such as the SRTT acquire two independent forms of knowledge about the sequential structure of stimuli: implicit and explicit, that is, reportable knowledge.

According to the second view, implicit knowledge can be transformed into reportable knowledge. Put differently, it is assumed that individuals generate only one form of knowledge during implicit learning episodes that can either remain implicit or become available to verbal report (this is essentially Farah's "Quality-of-Representation" account) . The challenge thus becomes to describe the mechanisms by which verbal reports are generated from implicit knowledge.

A third alternative is to deny the existence of implicit, yet causally efficacious representations altogether. Consequently, any performance gains in an implicit learning task such as the SRTT are attributed to an increase of explicit, reportable knowledge about the task structure. Such an "explicit knowledge" view of implicit learning phenomena offers a natural account for participants' verbal reports and avoids the problem of specifying the relation between separate implicit and explicit knowledge bases. In the following, we describe the proposals in some more depth.

#### 2.2.1 *Reportable knowledge and implicit knowledge are two kinds of knowledge that are generated by separate learning systems*

Dienes and Berry (1997), for instance, as proponents of the first view, argue that implicit and explicit (i.e., reportable) knowledge bases can be distinguished because they differ in terms of three characteristics, the amount of transfer to structurally similar tasks, the robustness of learning, and importantly, qualitatively different modes of learning. Reportable knowledge is assumed to be generated through deliberate hypothesis testing, whereas implicit learning is associated with a focus on particular items.

Willingham and Goedert-Eschmann (1999) focus on the interaction between explicit and implicit sequence learning in the context of the SRTT introduced earlier. Willingham (1998) proposes two modes of processing and learning that rely on different representational formats and brain structures. In the conscious mode, an attention-demanding strategic process contributes to improved performance in the SRTT through target selection and sequencing. It operates on the basis of explicit, conscious knowledge about the SRTT. Again, high-level problem-solving processes are believed to be responsible for the generation of such knowledge.

In the unconscious mode, sequence learning occurs through the tuning of a sequencing mechanism whenever the same sequence of targets is repeated. Implicit sequence learning can occur in parallel with explicit learning, at least so long as overt responses to the stimuli are carried out because proprioceptive feedback is essential for sequence learning in the motor-skill domain. The interaction between the learning systems is assumed to be unidirectional: Explicit sequence knowledge can be used to guide movements, and implicit learning proceeds on the basis of these movements.

In a similar fashion, Cleeremans (1993) suggests that explicit knowledge about the regularity in the SRTT can serve as input for the implicit learning mechanism. He further hypothesizes that the extent to which research participants develop reportable knowledge in the SRTT is influenced by individual differences in the allocation of attention and short-term memory capacity.

Thus, a theoretical position frequently encountered in the implicit learning literature is that the generation of reportable sequence knowledge is dependent on the availability of attentional resources. To a large degree, these conclusions are based on comparisons between groups of participants that are instructed to search for a regularity built into a task and incidental learning groups that do not receive a hint as to the existence of a regularity (e.g., Curran & Keele 1993; Willingham & Goedert-Eschmann 1999). Of course, this comparison tells us little about the difference between participants within the incidental learning group that can report the regularity by the end of training and those who cannot. Although conscious problem-solving type processes often are viewed as the source for reportable sequence knowledge, the theories described above provide no account of the processes themselves. Why, for instance, would a person engage in deliberate hypothesis testing without being explicitly instructed to do so?

Dienes and Perner (1999) provide a tentative answer to this question in the context of artificial grammar learning (e.g., Reber 1967). Endorsing the representational theory of mind, the authors explicate the distinction between explicit and implicit knowledge in terms of the functional use and semantic properties of mental representations. Knowledge is defined as a propositional attitude that has three constituents: First, a proposition that gets represented and serves as the content of the knowledge, second, an attitude toward that content, and third, the self as the holder of the attitude. The attitude of "knowing" (rather than, e.g., an attitude of "desiring" or "imagining") is determined by the way a person uses the representation of a proposition. The representation has to function as a reflection of the state of the world and not, e.g., as a reflection of a nonexistent goal state. Put differently, for a representation like "this is a fence" to constitute knowledge, it has to be put in a "knowledge box" and not, e.g., in a "goal box" which would be appropriate if the person merely intended to put up a fence in the backyard. The central idea now is that knowledge can vary in the degree of explicitness depending on which constituents of the propositional attitude actually get represented (made explicit) and which merely are implied by either the functional use of the representations or their conceptual structure.

How do these distinctions relate to an individual's ability to verbalize the content of a memory representation? Dienes and Perner claim that two necessary conditions must be met in order to achieve phenomenal awareness and thereby verbal expressability of some state of affairs. First, the propositional content needs to be represented explicitly, especially its factivity. Second, this mental state needs to be represented in a higher mental state or at least have the potential for such recursive representation. Put differently, in order to be able to report a fact, one needs to represent that one knows the fact. Thus, a second-order thought represents the holder and the attitude of the first-order state explicitly. This is the case of a fully explicit representation that includes all constituents of the propositional attitude. For instance, the representation "I know (that it is a fact) that the fence in my backyard is green" would enable a person to give a verbal report of the color of the fence in the backyard.

Although Dienes and Perner (1999) provide an insightful analysis of artificial grammar learning, their theory has not been applied to other implicit sequence learning paradigms yet. However, by analogy, the former analysis allows us to draw inferences about the acquisition of reportable sequence knowledge in the SRTT and other tasks. A participant engaging in the SRTT, for instance, might implicitly learn the rule that "the asterisk appears at the second screen position after it was shown at the first position" without representing an attitude of knowing toward it or representing the factivity of the rule. Therefore, the knowledge does not qualify for verbal report. Nonetheless, a person may perceive the successive appearance of the asterisk at Positions 1 and 2 as familiar or the succession of corresponding key presses as a fluent movement. Dienes and Perner's key suggestion is that this altered experience may cause a person to infer explicit knowledge about this regularity. That is, a person analyzes her own responses and experiences and draws inferences about the knowledge she must possess (see Siegler & Stern 1998, for a similar idea). Thus, reportable knowledge is generated by explicit learning mechanisms that analyze organismic changes brought about by implicit learning episodes. Interestingly, Dienes and Perner assume that this reportable sequence knowledge coexists in the knowledge box with the implicit representation of the regularity. A recent framework of implicit learning by Cleeremans and Jiménez (2002) attempts to overcome this representational dichotomy.

### 2.2.2 *Implicit knowledge can be transformed into reportable knowledge*

In their connectionist framework, Cleeremans and Jiménez identify representations with transient patterns of activation in distributed memory systems. Representations are assumed to be graded in that they can be characterized along dimensions such as activation strength, stability in time, and distinctiveness. Information that is encoded in the connection weights between units of a module or between different modules lacks representational quality. It can only be expressed and made available to the system through the activation of corresponding processing units. Therefore, weight-based knowledge is considered one possible form of implicit knowledge.

A related assumption is that learning can have indirect effects that do not need to be accompanied by phenomenal awareness. Knowledge representation in connectionist networks is superpositional in nature, i.e., an individual network is capable of generating a variety of stable activation patterns over its units. Therefore, if learning alters a particular representation through changes in connection weights, all those representations that rely on a shared set of processing units are affected as well. The effects on these related representations are indirect because the alteration of connection weights does not involve a direct updating of the corresponding representations.

A further possibility for knowledge to be implicit depends on the quality of a representation. Only high-quality representations that are sufficiently strong, stable, or distinct qualify for conscious awareness and, hence, can be accessed verbally. Importantly, although weak, low-quality representations are likely to remain implicit, they nevertheless are believed to exert a causal influence on processing, e.g., through associative priming mechanisms.

High-quality, adapted representations require time to develop, both at the level of individual practice trials and over the course of skill acquisition. During an implicit learning episode, the cognitive system constantly attempts to capture the correlational structure of the task, a process that is viewed essentially as a continuous adjustment of weights between units in a connectionist network. In this way, representations are shaped that influence behavior, but, early in training, are likely to lack the necessary strength, stability, or distinctness to enter conscious awareness or to be reported verbally. However, continual exposure to the task will increase the quality of the representations and at the same time their availability to verbal report.

In contrast to the theoretical position, frequently encountered in the literature, that access to phenomenal awareness and verbal expressability is determined by the properties of a representation alone (e.g., Norman 1968; O'Brien & Opie 1999, who equate conscious awareness with stable patterns of activation across the brain's neural networks), Cleeremans and Jiménez consider quality of representation a necessary, but not sufficient condition. They surmise that processes of attention and processes of integration are required for a representation to be expressed verbally. In this respect, the authors find themselves in agreement with recent developments in cognitive neuroscience (c.f., the "Integration" accounts described earlier).

Dehaene and Naccache (2001), for instance, suggest that for a representation to become conscious, it needs to be made available to a variety of processes that are realized in multiple specialized brain regions. Such global accessibility is achieved through a distributed system of neurons with long distance connectivity, the "global neuronal workspace." Top-down attentional amplification is seen as the primary mechanism of recruitment into the global workspace (see Dennett 2001, for a brief description of earlier formulations of the same concept in psychology and philosophy).

Similarly, Kanwisher (2001) assumes that perceptual awareness entails not only a sufficiently strong neural representation, but also access to that representation by most other parts of the brain including the speech system. Interestingly, Kanwisher suggests that perceptual learning does not only affect the quality of a representation, but also access to that representation. Finally, she proposes the "token individuation hypothesis:" A perceptual attribute (a "type") must be associated with the representation of a specific event or object (a "token") in order to enter phenomenal awareness. Kanwisher also notes the strong conceptual link between the binding of types to tokens and visual attention. Token individuation closely matches Dienes and Perner's (1999) requirements for phenomenal awareness, namely explicit representation of the factivity of the propositional content which entails explicit representation of a property, an individual, and the predication of the property to that individual. However, as noted above, additional representation of self and attitude may be necessary for phenomenal awareness in Dienes and Perner's framework. Thus, it remains a debatable question to what degree knowledge needs to be accompanied by relevant metaknowledge in order to qualify for phenomenal awareness and verbal report.

### 2.2.3 *Verbal reports are derived from implicit knowledge*

We now turn to a different view of implicit learning that questions the widely accepted distinction between implicit and explicit, that is, reportable knowledge. Just like Cleeremans and Jiménez (2002), Perruchet and collaborators (Perruchet, Vinter, & Gallego 1997; Perruchet & Vinter, in press) assume that learning shapes both individuals' experience of the world and their internal representations. However, the authors propose that the modification of conscious representations through learning is all that it takes to account for the vast array of implicit learning phenomena. Implicit representations that influence individuals' behavior without changing subjective experience simply do not exist in this framework. Therefore, "implicit learning may be thought of as allowing participants to pass from conscious perceptions and representations to other, generally better structured, conscious perceptions and representations, through the action of intrinsically unconscious mechanisms" (Perruchet et al. 1997: 44). More specifically, Perruchet and Vinter (in press) propose that automatic associative processes operate on the current content of the focus of attention and thus form new units or chunks that capture the structure of the task. However, these chunks can only persist and evolve into the new processing primitives of the system if they are repeated within an appropriate time frame. Thus, learning is viewed as an inevitable consequence of the self-organizing nature of attentional processing.

How do participants arrive at a verbal description of an experienced environmental regularity, such as the sequential regularity governing the SRTT? Perruchet and Vinter (in press) point out that the content of the attentional focus that is subject to processes of self-organization is not identical with the explicit, reportable knowledge about the regularity we are concerned with in this chapter. Several solutions are conceivable: Perhaps a simple "read-out" or description of the representation that has been built through automatic associative processes is sufficient to give a verbal account of at least some fragments of the sequence in the SRTT. This account becomes even more tenable when the learning mechanisms are powerful enough to directly represent structural qualities of the task such as abstract relations (Perruchet & Vinter, in press). Alternatively, additional inferential processes could be necessary that are triggered by participants' altered conscious experience. Thus, Dienes and Perner's (1999) suggestions about the development of explicit knowledge in implicit learning tasks may be viewed as a reasonable extension of the framework proposed by Perruchet and collaborators.

To summarize our discussion of views on the development of reportable knowledge that have been offered in the implicit learning literature: At present there exist at least three qualitatively different accounts in the implicit learning literature of how verbally expressible knowledge is generated. According to the first account, reportable knowledge and implicit knowledge are two kinds of knowledge that are generated by separate learning systems. According to the second proposal, verbal reports can be derived from implicit knowledge. In the third and final proposal, not discussed in depth above, it is assumed that all relevant knowledge acquired in implicit learning tasks is, in principle, available to verbal report.

## 3. The unexpected-event hypothesis

Our own present thinking owes much to the ideas proposed by Dienes and Perner (1999) and by Cleeremans and Jiménez (2002). However, although Dienes and Perner (1999) and Cleeremans and Jiménez (2002) have begun to tackle the mechanisms that are involved in the acquisition of reportable knowledge, much remains unclear about the exact nature of these processes: For example, how and when are the processes triggered in incidental learning situations? How do participants who are able to report a sequential regularity differ from participants who are not able to do so with respect to these mechanisms? Which situational characteristics have an effect on the mechanisms and which do not? In addition, there exists virtually no empirical evidence at all that would support the ideas proposed by Dienes and Perner (1999) and by Cleeremans and Jiménez (2002).

With our theoretical framework, we are trying to address at least some of the unsolved questions posed above. The basic framework is formulated as a stage model and is summarized in Figure 2. We believe that phenomenal awareness of an environmental regularity leads to participants' ability to verbally describe the regularity, and is, itself, the result of explicit, intentional hypothesis testing. Explicit hypothesis testing, in turn, is triggered by the observation of an unexpected event, and focuses on a search for an explanation to the unexpected event (e.g., Frijda 1994; Mandler 1984, 1992; Scherer 2001). In summary, in-

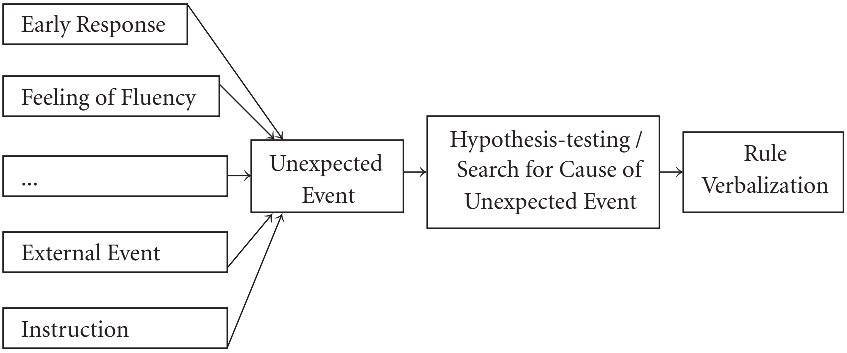


**Figure 2.** The Unexpected-Event Hypothesis.

dividuals are able to verbally describe an experienced environmental regularity because they actively and intentionally search for an explanation to an observed unexpected event.

Perhaps the key assumption in the framework concerns an individual's observation of an unexpected event. An event is unexpected when it is not consistent with what the individual has experienced in this particular situation in the past and therefore does not expect. A sudden sound occurring during the course of an experiment, for example, a freezing of the computer screen, a feeling of familiarity, a rapid motor response that occurs before the next stimulus is even shown – all of these are unexpected events that, when observed, trigger an intentional search for an explanation of why they occurred.

Unexpected events can be of two different kinds. First, they may be events that have no relation at all to the environmental regularity encountered resp. the regularity built into an experimental task. For example, the sudden and temporary freezing of a screen during the course of an experiment may be caused by an internal PC problem that has nothing at all to do with the task regularity constructed by the experimenter. Second, however, the unexpected event may be a direct consequence of the task regularity. This would be the case, for instance, when a participant in the SRTT notices that her finger movements have become much smoother over the course of the experiment.

The presence or absence of a relation between the observed unexpected event and the incidentally experienced regularity has, of course, direct implications for the likelihood with which an individual will become aware of the regularity. The observed unexpected event triggers an intentional search for its cause, typically taking the form of hypothesis testing, and eventually may, or may not, lead to discovery and verbal report of the regularity built into the task. The search will likely lead to discovery and verbal report of the regularity when the observed unexpected event has a close a priori relation to the regularity; it will likely not lead to discovery when this is not the case. In short, we assume that the ability to verbally report an experienced environmental regularity is the result of a successful explicit intentional search for an explanation to an observed unexpected event (see Whittlesea & Williams 2000, 2001a, 2001b, for a similar idea in a different context).

## 4. Empirical support for the unexpected-event hypothesis

Below we describe two types of empirical support for our theoretical framework, descriptive and experimental support. The descriptive support consists of analyses in which we, first, identify the time point during task practice at which individual participants become able to verbally report a regularity built into a task. In a second analysis step, we then obtain data, by working backwards from the identified time points, that addresses our claim that reportable knowledge is the result of an explicit search for the cause of an observed unexpected event. By comparison, the focus of our second type of support, experimental support, is on experimentally manipulating variables that presumably affect one or more of the theoretical stages we propose.

## 4.1 Descriptive support

Much of our descriptive empirical support comes from analyses of individual learning data with the so-called Alphabet Verification Task, AVT (e.g., Haider & Frensch 1996). In the AVT, participants are shown, on a computer screen, alphabetic strings of varying lengths that consist of a letter-digit-letter triplet at the end of the sequence and additional letters up front (e.g., "A B C D E (4) J"). Participants' task is simply to decide whether or not the strings that are shown follow the alphabet. To do so, they process the strings from left to right, interpreting the number in brackets as number of letters that are to be left out. Thus, the string "E (4) J," for example, is processed as E, leave out the next four letters F G H I, and then continue with J K L M N. The correct answer to this string would be that it follows the alphabet. The string "E (4) K," by contrast, would not follow the alphabet because when one leaves out four letters after E, one needs to continue with J, not with K.

The regularity in this task that participants eventually may or may not come to verbally report consists of the fact that errors in incorrect strings of this task occur only at a particular designated serial position, namely at the letter immediately to the right of the bracket. Thus, in order to decide whether or not an entire sequence follows the alphabet, participants need to process only the letter-digit-letter triplet; for the task at hand, processing of the remaining letters is irrelevant. Participants are not informed of this regularity.

### 4.1.1 *Identifying the time point at which verbal report becomes possible*

How can we hope to identify the time point during task practice at which individuals become able to verbally report the regularity built into the AVT? Our argument is as follows: we assume that as soon as participants can verbalize the regularity, they will use this information in order to speed up their task processing. That is, we assume that participants will not process the task-irrelevant information any longer (see Haider & Frensch 1999a, 1999b, for empirical support of this assumption). This, in turn, should lead to a sudden reduction of participants' overall response times for trials, that is, to a discontinuity in trial response times. The main empirical questions thus become: can we empirically identify RT-discontinuities in participants' individual practice trials and if so, are the observed RT-discontinuities indicative of participants' ability to verbally report the regularity built into the task?

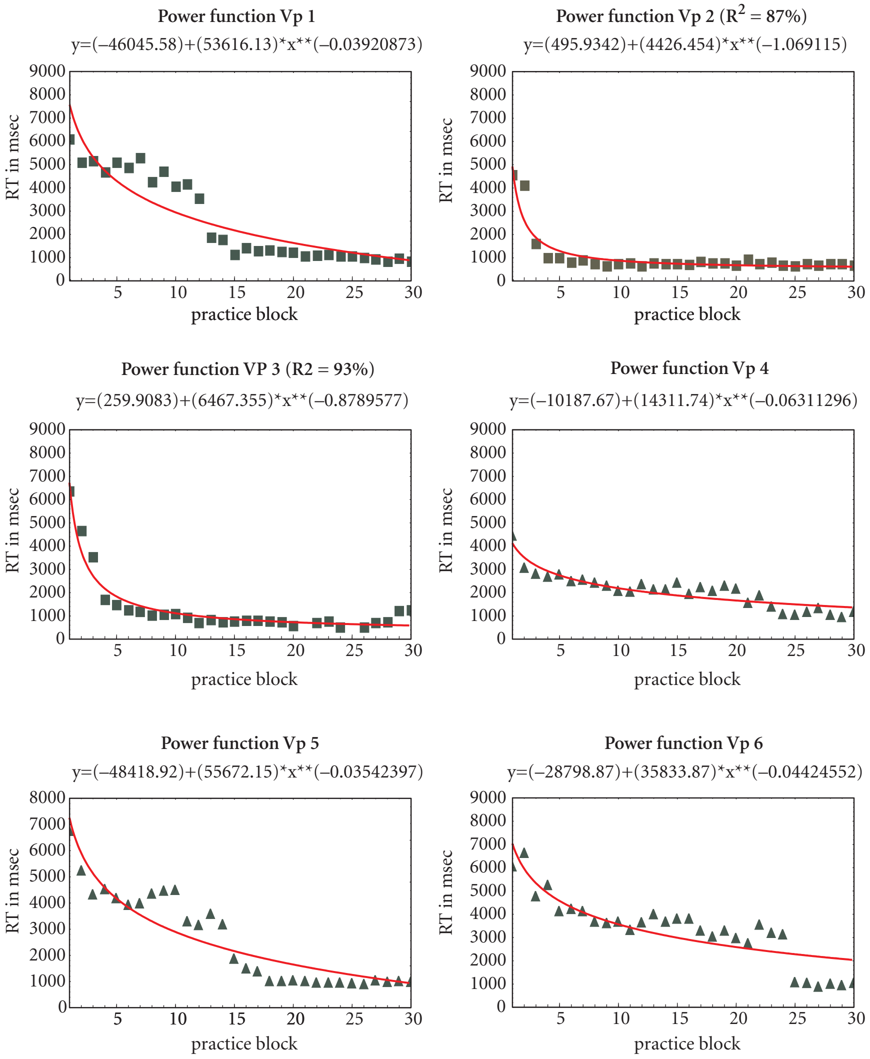


**Figure 3.** RT means for six individual participants (AVT).

Figure 3 shows the results of six single-case studies in which participants were asked to perform the AVT for a total of 30 trial blocks. In addition to plotting the individual RT curves, the Figure also shows the best-fitting power functions for the individual RT data.

Even a cursory examination of the individual RT curves shown in Figure 3 reveals that some participants show a very clear and pronounced RT-discontinuity, that is, a discrepancy between RT data and fitted power functions, whereas others do not. Using an F-test with α of 5%, we can statistically determine that four of the six participants (i.e., Vp 1, Vp 4, Vp 5, Vp 6) show a significant deviation from a power function fit, whereas the other two participants (i.e., Vp 2, Vp 3) do not.

The more interesting and important question, of course, is whether or not the observed RT-discontinuities really are a consequence, as we assume, of the ability to verbally report the regularity that was built into the task. This was indeed the case. All four of the four participants showing an RT discontinuity were also able to verbally describe the regularity when asked to do so at the end of the experiment; in contrast, the remaining two participants were not able to do so.

The obtained empirical relation between RT-discontinuity and verbal report of the task regularity was replicated with a much larger sample of participants, re-analyzing data we had collected earlier (Haider & Frensch 1999a). The results with the larger sample of 45 participants were such that of the 18 participants who showed an RT-discontinuity, 17 were able to report the regularity built into the task; 1 individual was not. By contrast, only 7 of the 27 participants showing no RT-discontinuity were able to verbally describe the regularity by the end of the experiment whereas the remaining 20 participants were not able to do so.[2]

In summary, the results described thus far are consistent with the claim that the RT-discontinuities, identified for each individual, occur indeed temporally close to the point in time at which participants acquire the ability to verbally report the built-in task regularity. However, two concerns come immediately to mind that question the validity of this conclusion. First, one might argue that perhaps verbal report of the regularity in this task does not precede the observed RT-discontinuities but rather might follow it. Second, one might argue that our findings might be unique to the task examined, the AVT.

The first argument is not only inconsistent with participants' self-reports when queried extensively at the end of the experiment, it is also made less convincing, although not entirely refuted, by the results of an additional experiment (Haider & Frensch, in preparation) in which we continuously and individually monitored, while participants were performing the AVT, whether or not an RT-discontinuity had occurred. As soon as an RT-discontinuity had been identified, we discontinued the experiment and asked participants to verbally describe any regularity they might have encountered. In this experiment, of the 14 participants who showed an RT-discontinuity, 12 were aware of the regularity built into the task; 2 participants were not. By contrast, only 4 of the 15 participants showing no RT-discontinuity were able to verbally describe the regularity by the end of the experiment whereas the remaining 11 participants were not able to do so. The available data, thus, suggests rather strongly that participants become able to verbally report the regularity built into the AVT shortly before the RT-discontinuity occurs.

The second potential counterargument, namely that the finding of an empirical relation between the occurrence of a RT-discontinuity and the ability to verbally report the task regularity is unique to the AVT is refuted by an additional experiment using a different task, the so-called Number Reduction Task, NRT (Haider & Frensch, in preparation; Thurstone & Thurstone 1941; Woltz, Gardner & Bell 2000; Woltz, Bell, Kyllonen, & Gardner 1996). In the NRT, participants are shown a string of six digits on a computer screen that they are asked to process pairwise from left to right (e.g., "9 9 9 1 4 1"). Strings contain only the digits "1," "4," and "9." Participants are asked to follow two rules when they process the digit string. Rule 1 states that the result of two identical digits is the digit itself. Rule 2 states that the result of two non-identical digits is the remaining third digit. Participants' task will become clear when we describe the mental steps that are necessary to complete an example string, "9 9 9 1 4 1."

The example string is to be processed from left to right. First, the two leftmost digits are processed, that is, the digits "9" and "9." Application of Rule 1 stating that the result of two identical digits is the digit itself, generates "9" as the result of the first comparison. All remaining comparisons are now made between the preceding result and the next digit. Thus, next, the result of the first comparison, "9," is compared with the next string digit, that is, the third digit in the sequence. The result of this comparison is, again according to Rule 1, the digit "9." Next, the last result, thus, the digit "9," is compared to the next digit in the string, a "1." Comparing the digits "9" and "1" results, according to Rule 2, in a "4." And on it goes. On any given trial, participants generate and enter a total of five responses.

The regularity in this task that participants eventually may or may not come to verbally report consists of the fact that, on any trial, Responses 4 and 5 are a mirror image of Responses 2 and 3. That is, Response 4 is equal to Response 3 and Response 5 is equal to Response 2. The regularity is not communicated to participants nor are they asked to search for any regularity that may be hidden in the task; the learning situation is thus incidental. Participants who discover the task regularity are able to speed up their responses substantially, either by entering Responses 3, 4, and 5 in very quick succession or by skipping Responses 3, 4, and 5 altogether (the latter was possible).

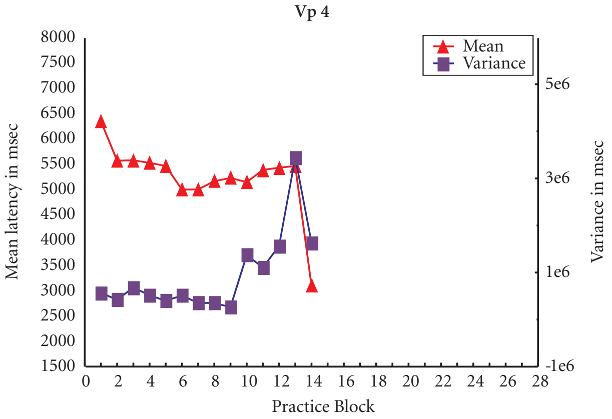
In a sample of 50 participants with the NRT, of the 13 participants who showed an RT-discontinuity, 11 were aware of the regularity built into the task; only 2 participants were not. By contrast, only 9 of the 37 participants showing no RT-discontinuity were able to verbally describe the regularity by the end of the experiment whereas the remaining 28 participants were not able to do so. These results, then, together with the findings for the AVT, strongly support the assumption that verbal report of task regularity arises during task processing and precedes an observed RT-discontinuity.

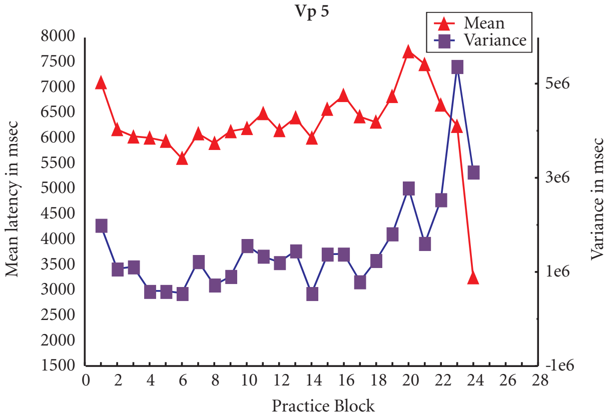
#### 4.1.2 *Explicit hypothesis testing leads to verbal reportability*

The results shown thus far do not speak to the issue of which mechanism might generate verbal report. As outlined above, at least two broad possibilities can be distinguished. First, explicit verbal report may be the result of implicit knowledge turned explicit. For example, as argued by Norman (1968) and Cowan (1995), among many others, the quality of a mental representation, that is, its strength, stability, or other quality directly determines when a representation reaches awareness. Second, as is argued by, among others, ourselves as well as Dienes and Perner (1999), explicit verbal report may be the result of explicit hypothesis testing.

How can these possibilities be empirically distinguished? As a first stab, we argue that a search for an explanation to an unexpected event, that is, intentional hypothesis testing, should lead to increased individual trial-RT variances immediately prior to the point in time at which a task regularity becomes reportable. That is, the trial block or blocks preceding individually identified RT-discontinuities should show larger RT variances than is normally to be expected for any given individual. For illustration purposes, Figure 4 shows some sample practice curves that we obtained with the Number Reduction Task NRT.

Subject 4 in Figure 4, for instance, shows a very marked RT-discontinuity between Trial Blocks 13 and 14. In the trial block preceding the RT reduction, the individual shows her largest RT variance for any trial block. Similarly, Subject 5 shows an easily detectable RT-discontinuity between Trial Blocks 23 and 24. In the trial block preceding the RT reduction, that is in Trial Block 23, the individual shows her largest variance for any trial block.





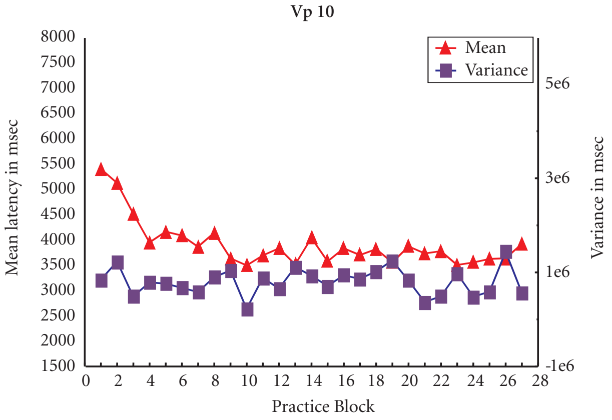


**Figure 4.** RT means and variances for three individual participants (NRT).

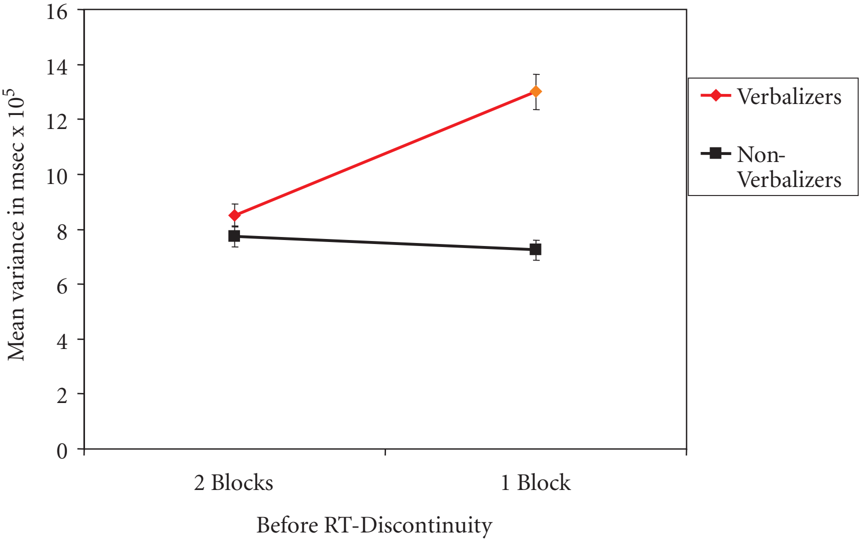


**Figure 5.** RT variances for participants able to verbally report an experienced task regularity and participants not able to verbally report the regularity prior to individually identified RT-discontinuities (NRT).

On the other hand, Subject 10 does not show an RT-discontinuity over the course of 27 trial blocks and also does not show a pronounced or out-of-the-ordinary variance in any of her trial blocks. Also, Subject 10 did not report any knowledge about the underlying task regularity, whereas Subjects 4 and 5 did.

Is, what we see for these three cases, also true in a larger sample? Figure 5 depicts the mean variance of participants showing an RT-discontinuity in the two trial blocks immediately preceding the block with the RT-discontinuity for the NRT experiment described earlier. For comparison purposes, the data of participants showing an RT-discontinuity is compared with data from participants showing no RT-discontinuity. For the latter group of participants, the trial block with the largest RT decline over the course of task practice was used as the equivalent to the RT-discontinuity. It is quite evident that the group showing an RT-discontinuity also shows a very pronounced increase in variance in the trial block immediately preceding the block in which the RT-discontinuity occurs.

In addition, of the 19 participants who were able to report the regularity in this task, almost 75% showed their largest variance in the trial block immediately preceding the block where the RT-discontinuity occurs. On the whole, thus, the systematic relation that we obtain between an increase in RT variance and a RT-discontinuity is at least strongly supportive of our claim that verbal report is the consequence of an intentional search process.

#### 4.1.3 *The experience of an unexpected event triggers hypothesis testing*

If indeed the increase in RT variance reflects intentional explicit hypothesis testing on the part of participants, then what exactly is it that triggers this hypothesis testing? According to the Unexpected-Event Hypothesis, hypothesis testing is triggered by the occurrence of an unexpected event and reflects a search for an explanation to the event.

We argue that an unexpected event can be one of many different things, some of which are more likely to lead to discovery of the regularity built into the task than others. Unexpected events can be external sounds, noises, visual irregularities on a computer screen, and so on. However, although random sounds and visual irregularities on the screen may be unexpected, they may not necessarily trigger a search for their cause. In addition, random sounds, even if they do trigger a search for their cause, will most likely not lead to detection of the regularity built into a task because the task regularity is not a likely cause of the unexpected event. Thus, in order for an unexpected event to lead to the verbal report of a task regularity, the event needs to have some plausible causal relation to the regularity.

We therefore believe that unexpected events that most likely lead to detection and verbal report of a task regularity are events that, themselves, are expressions of the regularity, such as behavioral consequences of the regularity. In other words, we argue that behavioral learning of the task regularity, as it manifests itself in, for instance motor learning, is most likely to lead to discovery and verbal report of the regularity.

An example should help to clarify this point: Assume your motor system has learned that the Responses 4 and 5 in the Number Reduction Task are the mirror image of the Responses 2 and 3. It may then happen that your motor system, i.e., your fingers, produce Response 4 too early, that is, Response 4 is emitted before the digit in the sequence that is needed to compute Response 4 is even perceived. If that happens, a research participant experiences an unexpected event, that is, an event that is not consistent with the way the task was performed in the past and with the way the task was to be performed. The unexpected event, then, triggers a search for its cause – how is it be possible that a finger produces a response already before the relevant stimulus is perceived? One plausible cause for this event and one possible result of the explicit search is the regularity built into the task.

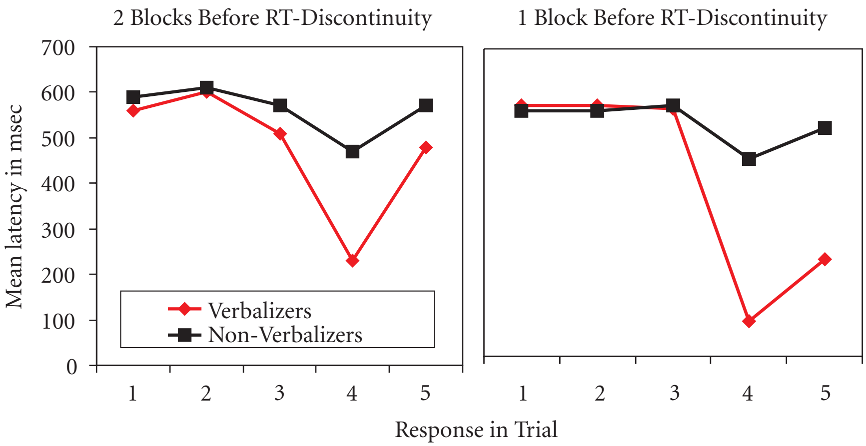


**Figure 6.** Means of the fastest observed RTs for participants later able to verbally report the regularity and participants not able to report the regularity prior to individually identified RT-discontinuities (NRT).

If this admittedly complicated explanation of ours is correct, then we should (a) be able to find "early" responses in individual participants' data just prior to their explicit hypothesis testing phases, and (b) more generally, we should find that the participants who demonstrate hypothesis testing should be those participants whose motor system has learned the regularity to a larger extent than participants who do not test hypotheses.

Figure 6 depicts the means of the fastest observed RTs for participants later able to verbally report the regularity and participants not able to report the regularity in the NRT experiment from above. The means of the fastest response times are depicted for the trial block in which, presumably, hypothesis testing occurs as well as the preceding trial block. The data are consistent with our assumption: Participants later able to report the regularity have faster minimum RTs than participants later showing no report. This is true, however, only for responses that are predictable on the basis of the regularity built into the task, that is Responses 4 and 5, and not for the remaining responses.

## 4.2 Experimental support

Next, we briefly describe some experimental evidence, collected with the NRT, that also supports the Unexpected-Event Hypothesis. With the experiments we have addressed two key assumptions of the framework, (a) the assumption that an unexpected event is the trigger for the entire process eventually leading to verbal report of the task regularity, and (b) the assumption that a search for a cause to an observed unexpected event leads to verbal report of the regularity.

#### 4.2.1 *Unexpected events trigger intentional hypothesis testing*

The first assumption we examined with a variety of experimental manipulations. In the interest of space, we focus on only one of these manipulations, an experiment in which we tried to make it more or less easy for participants to notice that an unexpected event had occurred.

In the experiment, we used the NRT in a modified version in which the six digits of a series were not displayed simultaneously, but rather were displayed successively (Haider & Frensch, in preparation). That is, a participant was first presented with the two left-most digits of a series. As soon as the first response had been entered, the next digit was presented, and so on. The regularity built into the task was the same as before. That is, Responses 4 and 5 were a mirror image of Responses 2 and 3.

There were three experimental conditions realized in this experiment that differed solely in the timing of the Response-Stimulus Interval (RSI). In the first experimental condition, RSI was fixed and set equal to 500 ms for all digits shown on a particular trial. Thus, Digits 1 and 2 were shown first. Then, 500 ms after the first response had been entered, Digit 3 was presented and so on.

In the second and third experimental conditions, the RSI varied for the presentation of the digits. In Experimental Condition 2, the RSI for the presentation of Digits 3 and 4 was set to 500 ms; the RSI for Digits 5 and 6 was set to 0 ms. In the third experimental condition, the set-up was exactly the reverse.

What were our predictions concerning the number of participants in each of the three experimental conditions that would be able to report the regularity built into the task when questioned at the end of the experiment? Earlier we mentioned that one possible "unexpected event" for an individual during the course of task practice might be an early response, that is, a response that the motor system releases before the individual has perceived the next response-relevant stimulus. The three experimental conditions were designed to make detection of an early motor response more or less likely.

Notice that an early motor response can occur only for Responses 4 and 5 because they are the only ones that are pre-determined. Giving participants an RSI of 500 ms for Responses 4 and 5 makes it quite likely that an early motor response that occurs before the next stimulus is even presented might be noticed by participants and viewed as an unexpected event. Giving participants an RSI of 0 ms makes it much more difficult to detect an early response because with an RSI of 0 ms, any response participants emit must necessarily occur after presentation of the next stimulus.

Based on this reasoning, we assumed that the unexpected event of an early motor response might be more likely noticed by participants in Experimental Conditions 1 and 3 than by participants in Condition 2. Because noticing an unexpected event triggers a search for its cause, participants in Conditions 1 and 3 should consequently be more likely to search for and become aware of the regularity built into the task than participants in Condition 2.

The main results of this experiment are qualitatively consistent with our expectations. The number of participants showing awareness of the task regularity was 19 (of 26) and 10 (of 20) in the Experimental Conditions 1 and 3, respectively, and 11 (of 26) in Condition 2.

A somewhat different way of manipulating the ease with which an unexpected event is noticed by participants is to use keyboard or mouse as external response devices. Without going into detail, it might suffice to say that there are good reasons for assuming that the unexpected event of an early response is much more likely to be noticed by participants using a keyboard as external device than by participants using a mouse. Consequently, participants using a keyboard should be more likely to become aware of the task regularity than participants using a mouse.

This is, the reader may remember, the exact result of the anagram experiment described in the Introduction to this chapter and it is a result that we have in the meantime replicated with the NRT (Haider & Frensch, in preparation).

### 4.2.2 *Verbal report is the result of a search for a cause to an observed unexpected event*

The second line of experimental research we have conducted recently is concerned with manipulating the difficulty of finding an explanation for the observed occurrence of an unexpected event. To do so, we have used the version of the NRT in which the Digits are presented successively; RSI was set equal to 500 ms for the presentation of all digits. To make finding an explanation easier or more difficult, we manipulated whether or not all of the information necessary to discover the rule that Responses 4 and 5 were the mirror image of Responses 2 and 3 was simultaneously present on the screen. In one condition, the digits and responses that appeared successively on each trial remained on the screen until the last response was entered. Thus, the information necessary to detect the regularity was all present on the screen.

In another condition, only the digit necessary to compute the next response was visible on the screen along with the last response – preceding digits and responses were erased. In the latter condition, thus, the information capturing the regularity built into the task was not all simultaneously present on screen.

Our reasoning was, of course, that participants in the condition in which all processed stimuli and responses remained visible on screen would have an easier time finding an explanation for the experienced "unexpected event" than would participants who saw only the information currently needed to compute the next response.

The results indicated that this reasoning was indeed correct. Fifteen of the 26 participants in the "all information available" condition became aware of the task regularity. By contrast, only 6 of 24 participants in the "partial information available" condition were able to verbalize the task regularity when asked to do so at the end of the experiment. It is important to add that this was true despite the fact that motor learning did not differ for the two experimental conditions.

## 5. Summary and conclusions

Starting point for the present chapter was the observation that in any experiment on implicit learning, and, indeed, in any experiment on human learning in general, some participants are able to verbally report the regularity built into the learning task whereas others are not. The main goals of the chapter were to (a) discuss the main important theoretical ideas that have been proposed in the literature to explain an individual's ability to verbally report a regularity that was experienced in the context of an incidental-learning situation, and (b) convince the reader of this chapter of the validity of our own theoretical framework. According to our own thinking, the ability to verbally report an experienced environmental regularity is a consequence of an explicit intentional search for an explanation to an unexpected event.

What are some of the implications of the empirical work and theoretical framework proposed in this chapter? First, pure "Quality-of-Representation" accounts of awareness (e.g., Norman 1968; O'Brien & Opie 1999) are unlikely to be successful candidates for explaining the data we have collected. For instance, we described an increase in RT variance immediately prior to an observed RT-discontinuity and, ultimately, prior to verbal report of the regularity. Such an increase in RT variance is difficult to reconcile with accounts that rely solely on changes in the quality of representation. This is not to say that these accounts may not explain the generation of awareness under different circumstances. It is to say, however, that for the specific situation that we are interested in, these accounts are at present not convincing theoretical contenders.

Second, in our framework it is assumed that an implicit *learning* system and an explicit *reasoning* system co-exist. The two systems do not operate in complete independence. Rather, the one system, namely the implicit learning system, results in behavioral change that, when noticed, may trigger the other system, the explicit reasoning system (e.g., Dienes & Perner 1999). We hasten to add, though, that the data presented in this chapter does little to elucidate the nature of the implicit learning processes that precede explicit hypothesis testing. Our main goal has been a different one: to identify the mechanisms that are responsible for the generation of verbal reports in incidental learning situations.

Third, according to our framework and based on the former reasoning, variables affecting the implicit learning system should also affect participants' ability to verbally describe the regularity the implicit learning system has learned. Conversely, there should exist some variables that affect participants' ability to verbally describe a regularity without affecting the implicit learning system. There is plenty of empirical evidence in the area of implicit learning supporting the first assumption and there is some, though far less, evidence supporting the second assumption.

More specifically, Table 1 contains brief summaries of published articles in the implicit learning domain (here specifically and exclusively articles using the SRTT) in which the effects of experimentally manipulated variables on the likelihood of participants' ability to verbally report an experienced task regularity have been reported. According to Table 1, the likelihood of being able to verbally express an experienced task regularity is affected by, for instance, the presence versus absence of a secondary task, the length of the used deterministic sequence in the SRTT, the complexity of the deterministic sequence, the duration of task practice, and the age of participants.

By far most of the experimental variables that have been shown to affect the likelihood of verbal report also appear to affect degree of implicit sequence learning. For example, Willingham, Nissen, and Bullemer (1989, Experiment 2) demonstrated that the amount of task practice affects both degree of implicit learning as well as verbal report. Participants in the six experimental groups in the study differed in that they practiced between one and six blocks of the SRTT. Verbal report of the experienced regularity increased systematically with the number of blocks practiced as did the amount of implicit learning.

Similarly, the presence of a secondary task reduces (relative to the absence of a secondary task) both the number of participants able to verbally report the experienced regularity and the degree of implicit knowledge acquired (e.g., Cohen, Ivry, & Keele 1990; Frensch, Buchner, & Lin 1994; Frensch, Wenke, &

**Table 1.** Recently published articles using the SRTT.

| Reference | Variable Manipulated | Effect on Verbal Report | Effect on Implicit Learning |
|---|---|---|---|
| Cherry & Stadler (1995) | Age | Yes | No |
| | Educational attainment, occupational status, verbal ability | Yes | Yes |
| Cohen, et al. (1990) | Dual/single task | Yes | Yes |
| Curran (1997) | Age | No | No |
| | First-order predictive sequence/second-order predictive sequence<br>predictive sequence | No | No for young<br>Yes for elderly |
| | Instruction | Yes | Yes |
| Curran & Keele (1993) | Intentional/incidental learning and dual/single task | Yes | Yes (single-task)<br>No (dual-task) |
| | Dual task/single task | Yes | Yes |
| Eimer, et al. (1996) | Frequency of deviant stimuli | Yes | Yes |
| Harrington & Haaland (1992) | Age | Yes | Yes |
| Heuer, et al. (1998) | Mental fatigue | Yes | No |
| | Loss of sleep | No | Yes |
| Hoffmann & Koch (1997) | S-R compatibility<br>Connectibility | No | No |
| Honda et al. (1998) | Instruction | Unclear | No |
| Howard & Howard (1997) | Age | No | Yes |
| McDowall & Allison (1995) | Anxiety | Yes | No |
| Nissen & Bullemer (1987) | Dual/single task | Yes | Yes |
| | Amnesia | Yes | No |
| Reber & Squire (1994) | Amnesia | Yes | No |
| Seger (1997) | Amount of practice | Yes | Unclear |
| | Dual/single task | Yes | Unclear |
| Willingham, et al. (1993) | Sequence/random | Yes | Yes |
| Willingham, et al. (1989) | Amount of practice | Yes | Yes |
| Ziessler, 1998 | Reaction-stimulus-regularity | No | Yes |
| | Frequency of the association between stimuli and responses | No | No |
| | Frequency of the association between responses and effect | No | Yes |

Rünger 1999). In the Cohen et al. (1990) study, for instance, only four of twelve participants were able to report part of the deterministic sequence when the experience of the sequence was accompanied by the requirement to simultaneously perform a secondary task whereas ten out of twelve participants were able to do so when there was no second task.

On the other hand, there also exists some experimental evidence supporting our contention that the ability to verbalize an experienced task regularity can be affected by experimental variables when the same variables do not affect the degree of implicit knowledge that is acquired. For example, Cherry and Stadler (1995) report that an increase in age affects the number of participants able to report an experienced task regularity negatively. However, age does not affect the amount of knowledge that is acquired implicitly. Interestingly, Harrington and Haaland (1992) found no difference between young and elderly participants when they simply asked participants whether or not they had noticed anything unusual about the experienced sequences. However, when asked to describe the experienced sequences, younger participants were able to do so far more accurately than were older participants.

Also, participants who are mentally fatigued seem to be able to verbally report more of an experienced task regularity than participants who are not fatigued, although degree of mental fatigue does not appear to affect implicit learning (Heuer, Spijkers, Kiesswetter, & Schmidtke 1998); interestingly and very surprisingly, loss of sleep appears to have exactly the opposite effect. That is, loss of sleep seems to affect implicit learning but not verbal report.

Thus, to summarize, there exists some empirical evidence that is consistent with, though certainly not conclusive of, our claim that an implicit learning system and an explicit reasoning system may co-exist, and that, more importantly, the two systems do not operate in complete independence. Rather, the one system, the implicit learning system, results in behavioral change that, when noticed, may trigger the other system, the explicit reasoning system.

Finally, we argued above that the mechanisms we have described in this chapter apply to learning situations in which (1) an individual experiences an environmental regularity in the absence of any explicit instruction to discover the regularity; and (2) the individual correctly and verbally reports the characteristics of the regularity when asked to do so The mother of all remaining questions is, thus, to what extent the mechanisms described here might be generalizable, that is, might also be operable in learning situations that differ from the situation considered in this chapter and in which awareness of what has been learned arises – we admit that, at present, we have simply no idea.

## Notes

* We thank Norman Radtke for his assistance with data collection and his comments on earlier versions of this chapter.

**1.** A somewhat similar, though not identical, categorization of psychological theories of awareness has recently been offered by Atkinson, Thomas, and Cleeremans (2000). The authors distinguish between four distinct theoretical types of explanations, specialized vehicle theories (e.g., Norman 1968), specialized process theories (e.g., Schacter 1989), non-specialized vehicle theories (e.g., O'Brien & Opie 1999), and non-specialized process theories (e.g., Crick & Koch 1995; Grossberg 1999; Tononi & Edelman 1998).

**2.** The ability to verbally report the experienced task regularity was assessed in a somewhat more complicated manner in this experiment; for details, see Haider and Frensch (1999a).

## References

Atkinson, A., Thomas, M., & Cleeremans, A. (2000). Consciousness: Mapping the theoretical landscape. *Trends in Cognitive Sciences, 4*, 372–382.

Cherry, K. E. & Stadler, M. A. (1995). Implicit learning of a nonverbal sequence in younger and older adults. *Psychology and Aging*, *10*, 379–394.

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

Cleeremans, A. & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French and A. Cleeremans (Eds.), *Implicit Learning and Consciousness* (pp. 1–40). Hove, UK: Psychology Press.

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

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford Psychology Series, No. 26. New York: Oxford University Press.

Crick, R. & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375*, 121–123.

Curran, T. (1997). Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. *Psychological Research*, *60*, 24–41.

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

Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1–37.

Dennett, D. (2001). Are we explaining consciousness yet? *Cognition*, *79*, 221–237.

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

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

Eimer, M., Goschke, T., Schlaghecken, F. & Stürmer, B. (1996). Explicit and implicit learning of event sequences: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 970–987.

Farah, M. J. (1994). Visual perception and visual awareness after brain damage: A tutorial overview. In C. Umiltá and M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and nonconscious information processing* (pp. 37–76). Cambridge, MA: MIT Press.

Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 567–584.

Frensch, P. A., Wenke, D., & Rünger, D. (1999). A secondary tone-counting task suppresses performance in the Serial Reaction Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 260–274.

Frijda, N. H. (1994). Emotions require cognitions, even if simple ones. In P. Ekman and R.J. Davidson (Eds.), *The Nature of Emotion* (pp.197–202). New York: Oxford University Press.

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8*, 1–44.

Haider, H. & Frensch, P. A. (in preparation). *Verbal report of incidentally experienced environmental regularity*.

Haider, H. & Frensch, P. A. (1999a). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 172–190.

Haider, H. & Frensch, P. A. (1999b). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied, 5*, 129–151.

Haider, H. & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*, 304–337.

Harrington, D. L. & Haaland, K. Y. (1992). Skill learning in the elderly. Diminished implicit and explicit memory for a motor sequence. *Psychology and Aging, 7*, 425–434.

Heuer, H., Spijkers, W., Kiesswetter, E. & Schmidtke, V. (1998). Effects of sleep loss, time of day, and extended mental work on implicit and explicit learning of sequences. *Journal of Experimental Psychology: Applied, 4*, 139–162.

Hoffmann, J. & Koch, I. (1997). Stimulus-response compatibility and sequential learning in the serial reaction time task. *Psychological Research*, *60*, 87–97.

Honda et al. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning. A PET study. *Brain*, *121*, 2159–2173.

Howard, J. H. & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology & Aging*, *12*, 634–656.

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*, *79*, 89–113.

Mandler, G. (1984) *Mind and Body: Psychology of Stress and Emotion*. New York: Norton.

Mandler, G. (1992) Emotions, evolution, and aggression: Myths and conjectures. In K. T. Strongman (Ed.), *International Review of Studies on Emotion* (Vol. 2, pp. 97–116). Chichester: Wiley.

McDowall, J. & Allison, C. (1995). Anxiety and learning: A dissociation between explicit and implicit processes. *New Zealand Journal of Psychology*, *24*, 44–50.

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

Norman, D. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*, 522–536.

O'Brien, G. & Opie, J. (1999). A connectionist theory or phenomenal experience. *Behavioral and Brain Sciences*, *22*, 127–148.

Perruchet, P. & Vinter, A. (in press). The self-organizing consciousness: A framework for implicit learning. In R.M. French and A. Cleeremans (Eds.), *Implicit Learning and Consciousness*. Psychology Press.

Perruchet, P., Vinter, A., & Gallego, J. (1997). Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin & Review, 4*, 43–48.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *77*, 317–327.

Reber, P. J. & Squire, L. R. (1994). Parallel brain systems for learning with or without awareness. *Learning & Memory, 1*, 217–229.

Schacter, D. L. (1989). On the relations between memory and consciousness : Dissociable interactions and conscious experience. In H. L. Roediger and F. I. M. Craik (Eds.), *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving* (pp. 355–389). Mahwah, NJ: Erlbaum Associates.

Scherer, K. (2001). Appraisal considered as a process of multilevel sequential checking. In Scherer, K., Schorr, A., and Johnstone, T. (Eds.), *Appraisal Processes in Emotion*. Oxford: Oxford University Press.

Seger, C. A. (1997). Two forms of sequential implicit learning. *Consciousness & Cognition, 6*, 108–131.

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

Stadler, M. & Frensch, P. A. (Eds.) (1998). *Handbook of Implicit Learning*. Thousand Oaks, CA: Sage Publications.

Thurstone, L. L. & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, No 2 94.

Tononi, G. & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282* (5395), 1846–1851.

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

Willingham, D. B. & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. *Psychological Science*, *10*, 531–534.

Whittlesea, B. W. A. & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 547–565.

Whittlesea, B. W. A. & Williams, L. D. (2001a). The discrepancy-attribution hypothesis I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 3–13.

Whittlesea, B. W. A. & Williams, L. D. (2001b). The discrepancy-attribution hypothesis II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 14–33.

Willingham, D. B., Greeley, T., & Bardone, A. M. (1993). Dissociation in a serial response time task using a recognition measure: Comment on Perruchet and Amorin (1992). *Journal of Experimental Psychology: Learning, Memory & Cognition*, *19*, 1424–1430.

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

Woltz, D. J., Bell, B. G., Kyllonen, P. C., & Gardner, M. K. (1996). Memory for order of operations in the acquisition and transfer of sequential cognitive skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 438–457.

Woltz, D. J., Gardner, M. K., & Bell, B. G. (2000). Negative transfer errors in sequential cognitive skills: Strong-but-wrong sequence application. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 601–625.

Wundt, W. (1896). *Grundriß der Psychologie* [Basics of psychology]. Leipzig: Engelmann.

Ziessler, M. (1998). Response-effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *24*, 962–978.

# Author index

# Subject index

In the series ADVANCES IN CONSCIOUSNESS RESEARCH (AiCR) the following titles have been published thus far or are scheduled for publication:

1. GLOBUS, Gordon G.: *The Postmodern Brain.* 1995.
2. ELLIS, Ralph D.: *Questioning Consciousness. The interplay of imagery, cognition, and emotion in the human brain.* 1995.
3. JIBU, Mari and Kunio YASUE: *Quantum Brain Dynamics and Consciousness. An introduction.* 1995.
4. HARDCASTLE, Valerie Gray: *Locating Consciousness.* 1995.
5. STUBENBERG, Leopold: *Consciousness and Qualia.* 1998.
6. GENNARO, Rocco J.: *Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness.* 1996.
7. MAC CORMAC, Earl and Maxim I. STAMENOV (eds): *Fractals of Brain, Fractals of Mind. In search of a symmetry bond.* 1996.
8. GROSSENBACHER, Peter G. (ed.): *Finding Consciousness in the Brain. A neurocognitive approach.* 2001.
9. Ó NUALLÁIN, Seán, Paul MC KEVITT and Eoghan MAC AOGÁIN (eds): *Two Sciences of Mind. Readings in cognitive science and consciousness.* 1997.
10. NEWTON, Natika: *Foundations of Understanding.* 1996.
11. PYLKKÖ, Pauli: *The Aconceptual Mind. Heideggerian themes in holistic naturalism.* 1998.
12. STAMENOV, Maxim I. (ed.): *Language Structure, Discourse and the Access to Consciousness.* 1997.
13. VELMANS, Max (ed.): *Investigating Phenomenal Consciousness. Methodologies and Maps.* 2000.
14. SHEETS-JOHNSTONE, Maxine: *The Primacy of Movement.* 1999.
15. CHALLIS, Bradford H. and Boris M. VELICHKOVSKY (eds.): *Stratification in Cognition and Consciousness.* 1999.
16. ELLIS, Ralph D. and Natika NEWTON (eds.): *The Caldron of Consciousness. Motivation, affect and self-organization – An anthology.* 2000.
17. HUTTO, Daniel D.: *The Presence of Mind.* 1999.
18. PALMER, Gary B. and Debra J. OCCHI (eds.): *Languages of Sentiment. Cultural constructions of emotional substrates.* 1999.
19. DAUTENHAHN, Kerstin (ed.): *Human Cognition and Social Agent Technology.* 2000.
20. KUNZENDORF, Robert G. and Benjamin WALLACE (eds.): *Individual Differences in Conscious Experience.* 2000.
21. HUTTO, Daniel D.: *Beyond Physicalism.* 2000.
22. ROSSETTI, Yves and Antti REVONSUO (eds.): *Beyond Dissociation. Interaction between dissociated implicit and explicit processing.* 2000.
23. ZAHAVI, Dan (ed.): *Exploring the Self. Philosophical and psychopathological perspectives on self-experience.* 2000.
24. ROVEE-COLLIER, Carolyn, Harlene HAYNE and Michael COLOMBO: *The Development of Implicit and Explicit Memory.* 2000.
25. BACHMANN, Talis: *Microgenetic Approach to the Conscious Mind.* 2000.
26. Ó NUALLÁIN, Seán (ed.): *Spatial Cognition. Selected papers from Mind III, Annual Conference of the Cognitive Science Society of Ireland, 1998.* 2000.
27. McMILLAN, John and Grant R. GILLETT: *Consciousness and Intentionality.* 2001.

28. ZACHAR, Peter: *Psychological Concepts and Biological Psychiatry. A philosophical analysis.* 2000.
29. VAN LOOCKE, Philip (ed.): *The Physical Nature of Consciousness.* 2001.
30. BROOK, Andrew and Richard C. DeVIDI (eds.): *Self-reference and Self-awareness.* 2001.
31. RAKOVER, Sam S. and Baruch CAHLON: *Face Recognition. Cognitive and computational processes.* 2001.
32. VITIELLO, Giuseppe: *My Double Unveiled. The dissipative quantum model of the brain.* 2001.
33. YASUE, Kunio, Mari JIBU and Tarcisio DELLA SENTA (eds.): *No Matter, Never Mind. Proceedings of Toward a Science of Consciousness: Fundamental Approaches, Tokyo, 1999.* 2002.
34. FETZER, James H.(ed.): *Consciousness Evolving.* 2002.
35. Mc KEVITT, Paul, Seán Ó NUALLÁIN and Conn MULVIHILL (eds.): *Language, Vision, and Music. Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, 1999.* 2002.
36. PERRY, Elaine, Heather ASHTON and Allan YOUNG (eds.): *Neurochemistry of Consciousness. Neurotransmitters in mind.* 2002.
37. PYLKKÄNEN, Paavo and Tere VADÉN (eds.): *Dimensions of Conscious Experience.* 2001.
38. SALZARULO, Piero and Gianluca FICCA (eds.): *Awakening and Sleep-Wake Cycle Across Development.* 2002.
39. BARTSCH, Renate: *Consciousness Emerging. The dynamics of perception, imagination, action, memory, thought, and language.* 2002.
40. MANDLER, George: *Consciousness Recovered. Psychological functions and origins of conscious thought.* 2002.
41. ALBERTAZZI, Liliana (ed.): *Unfolding Perceptual Continua.* 2002.
42. STAMENOV, Maxim I. and Vittorio GALLESE (eds.): *Mirror Neurons and the Evolution of Brain and Language.* 2002.
43. DEPRAZ, Natalie, Francisco VARELA and Pierre VERMERSCH.: *On Becoming Aware. A pragmatics of experiencing.* 2003.
44. MOORE, Simon and Mike OAKSFORD (eds.): *Emotional Cognition. From brain to behaviour.* 2002.
45. DOKIC, Jerome and Joelle PROUST: *Simulation and Knowledge of Action.* 2002.
46. MATHEAS, Michael and Phoebe SENGERS (ed.): *Narrative Intelligence.* 2003.
47. COOK, Norman D.: *Tone of Voice and Mind. The connections between intonation, emotion, cognition and consciousness.* 2002.
48. JIMÉNEZ, Luis: *Attention and Implicit Learning.* 2003.
49. OSAKA, Naoyuki (ed.): *Neural Basis of Consciousness.* 2003.
50. GLOBUS, Gordon G.: *Quantum Closures and Disclosures. Thinking-together post-phenomenology and quantum brain dynamics.* 2003.
51. DROEGE, Paula: *Caging the Beast. A theory of sensory consciousness.* 2003.
52. NORTHOFF, Georg: *Philosophy of the Brain. The 'Brain problem'.* n.y.p.
53. HATWELL, Yvette, Arlette STRERI and Edouard GENTAZ (eds.): *Touching for Knowing. Cognitive psychology of haptic manual perception.* 2003.
54. BEAUREGARD, Mario (ed.): *Consciousness, Emotional Self-Regulation and the Brain.* n.y.p.