Implicit Learning

50 Years On

EDITED BY Axel Cleeremans,
Viktor Allakhverdov,
and Maria Kuvaldina

A Psychology Press Book

ROUTLEDGE

IMPLICIT LEARNING

Can we learn without knowing we are learning? To what extent is our behavior influenced by things we fail to perceive? What is the relationship between conscious and unconscious cognition? *Implicit Learning: 50 Years On* tackles these key questions, fifty years after the publication of Arthur Reber's seminal text. Providing an overview of recent developments in the field, the volume considers questions about the computational foundations of learning, alongside phenomena including conditioning, memory formation and consolidation, associative learning, cognitive development, and language learning.

Featuring contributions from international researchers, the book uniquely integrates 'Western' thinking on implicit learning with insights from a rich Russian research tradition. This approach offers an excellent opportunity to contrast perspectives, to introduce new experimental paradigms, and to contribute to ongoing debates about the very nature of implicit learning.

Implicit Learning: 50 Years On is essential reading for students and researchers of consciousness, specifically those interested in implicit learning.

Axel Cleeremans is a research director with the Fund for Scientific Research (F.R.S.-FNRS), Belgium, and a professor of Cognitive Psychology at the Université libre de Bruxelles, where he directs the Center for Research in Cognition & Neuroscience and the ULB Neuroscience Institute.

Viktor Allakhverdov is a prominent Russian psychologist and chairperson of the division of General Psychology, St. Petersburg State University. He is the author of more than 5 books on consciousness and the head of a research group that investigates implicit learning and unconscious cognition.

Maria Kuvaldina is a postdoctoral researcher. She received her doctorate in cognitive psychology from St. Petersburg State University. Her research interests include mechanisms of attention and cognitive control.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

IMPLICIT LEARNING

50 Years On

Edited by Axel Cleeremans,
Viktor Allakhverdov, and Maria Kuvaldina

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2019
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record has been requested for this book

ISBN: 978-1-138-64429-8 (hbk)
ISBN: 978-1-138-64430-4 (pbk)
ISBN: 978-1-315-62890-5 (ebk)

Typeset in Bembo
by Swales & Willis Ltd, Exeter, Devon, UK

CONTENTS

ACKNOWLEDGEMENTS

This work was made possible through the involvement and support of different persons and institutions to which we are very grateful. Ceri Griffiths, our Editor at Routledge, accepted the proposal (almost) immediately during a meeting of the European Society of Cognitive Psychology in Budapest that both Axel Cleeremans and Maria Kuvaldina attended. Her initial efforts getting this project off the ground were followed up with great care by herself and Charlotte Mapp over the entire course of the process. Axel Cleeremans' involvement in the project was supported by Advanced Grant "Radical" from the European Research Council. Viktor Allakhverdov's and Maria Kuvaldina's efforts were supported by Saint Petersburg State University research grant # 8.38.287.2014 as well as by a number of research grants from the Russian Foundation for Basic Research and by the Russian Foundation for Humanities. The Belgian Fonds de la Recherche Scientifique (F.R.S.-FNRS) and St. Petersburg State University supported the organization of the third Implicit Learning Seminar in Saint Petersburg – the meeting where the very idea of composing this book was born. Special thanks must go to Agnieszka Poplawska, Elisabeth Norman, Michał Wierzchoń, Patrick Rebuschat, Dezső Németh and Adrian Opre – all organizers of the Implicit Learning Seminars. They are the ones whose dedication made this community exist. None of the work reported in this volume would have been possible without the colleagues and students who carried out the research and took part in the discussions during our meetings – thank you! Finally, the editors would also like to thank A. S. Reber for his efforts reviewing the current volume and composing a foreword.

CONTRIBUTORS

Viktor Allakhverdov, Department of Psychology, Saint Petersburg State University, Russia and Senior Researcher, National Research University Higher School of Economics (HSE), St. Petersburg, Russia.

Natalia V. Andriyanova, Department of Psychology, Saint Petersburg State University, Russia.

Laura J. Batterink, Department of Psychology, Northwestern University, Evanston, IL USA.

Andrey A. Chetverikov, Laboratory for Visual Perception and Visuomotor Control, Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland; Cognitive Research Lab, Russian Academy of National Economy and Public Administration, Moscow, Russia; Department of Psychology, Saint Petersburg State University, Russia.

Axel Cleeremans, Fund for Scientific Research (F.R.S.-FNRS, Belgium); Université libre de Bruxelles; Center for Research in Cognition and Neuroscience; the ULB Neuroscience Institute.

Monika Derda, Consciousness Lab, Institute of Psychology, Jagiellonian University, Krakow, Poland.

Arnaud Destrebecqz, Université Libre de Bruxelles, Center for Research in Cognition and Neurosciences (CRCN).

Zoltan Dienes, School of Psychology, University of Sussex, UK; Sackler Centre for Consciousness Science, University of Sussex, UK.

Margarita G. Filippova, Department of Psychology, Saint Petersburg State University, Russia.

Valeria A. Gershkovich, Department of the Problems of Convergence in Natural Sciences and Humanities, Saint Petersburg State University, Russia; N. P. Bechtereva Institute of the Human Brain, Russian Academy of Sciences.

Ivan I. Ivanchei, Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia; Department of Psychology, Saint Petersburg State University, Russia.

Emma Jones, Faculty of Psychology, University of Bergen, Norway.

Valeriia Yu. Karpinskaia, Department of Psychology, Saint Petersburg State University, Russia and Senior Researcher, National Research University Higher School of Economics (HSE), St. Petersburg, Russia.

Anatolii D. Karpov, Department of Psychology, Saint Petersburg State University, Russia.

Ferenc Kemény, Institute for Psychology, University of Graz, Graz, Austria; Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary.

Maria Kuvaldina, Department of Psychology, Saint Petersburg State University, Russia; New School for Social Research, New York, USA.

Ágnes Lukács, Department of Cognitive Science, Budapest University of Technology and Economics, and MTA-BME Lendület Language Acquisition Research Group, Budapest, Hungary.

Nadezhda V. Moroshkina, Department of Psychology, Saint Petersburg State University, Russia, Laboratory for Cognitive Studies, Saint Petersburg State University, Russia.

Elisabeth Norman, Faculty of Psychology, University of Bergen, Norway.

Alexandr S. Odainic, Department of Psychology, Saint Petersburg State University, Russia.

Irina Ovchinnikova, Department of Psychology, Saint Petersburg State University, Russia.

Agnieszka Popławska-Boruc, Sopot Faculty of Psychology, SWPS University of Social Sciences and Humanities, Poland.

Mark C. Price, Faculty of Psychology, University of Bergen, Norway.

Paul J. Reber, Department of Psychology, Northwestern University, Evanston, IL, USA.

Ben Reuveni, Department of Psychology, Northwestern University, Evanston, IL, USA.

Marta Roczniewska, Sopot Faculty of Psychology, SWPS University of Social Sciences and Humanities, Poland.

Estibaliz San Anton, Université Libre de Bruxelles, Center for Research in Cognition and Neurosciences (CRCN); Fonds de la Recherche Scientifique – FNRS.

Ryan B. Scott, School of Psychology, University of Sussex, UK; Sackler Centre for Consciousness Science, University of Sussex, UK.

Tanya V. Scott, City University of New York, Staten Island; NTENT.com.

Radosław Sterczyński, Sopot Faculty of Psychology, SWPS University of Social Sciences and Humanities, Poland.

Kelsey R. Thompson, Department of Psychology, Northwestern University, Evanston, IL, USA.

Natalia P. Vladykina, Early Intervention Center, Saint Petersburg, Russia.

Michał Wierzchoń, Consciousness Lab, Institute of Psychology, Jagiellonian University, Krakow, Poland.

FOREWORD

I'm delighted to be given the opportunity to write the foreword to this new and exciting volume of research on the seemingly ever-growing field of theory and research on the cognitive unconscious. It's a bit of cliché, when examining a field of study, to refer to the "breadth" and "depth" of the efforts – but I'm going to do it anyway, with a twist. Typically, in this way of framing the issue, "depth" refers to the extent to which the research has drilled down to more basic, more fundamental, core issues, principles, and causal links (as, for example, in Cleereman's aptly named chapter where he voyages into the deeper realms of cognitive functioning and encounters methodological difficulties at every level) whereas "breadth" takes note of how widely the research has spread, how many new domains have come under scrutiny, how new applications of findings and principles have spread, how much translation research has been carried out.

All this is easily seen in this new volume overseeing the work in implicit learning but what's noteworthy is that other meaning of "breadth": *geography*. Here we encounter researchers in countries and in laboratories that were and still are right in the thick of things though very few of us knew about them. And this is a good thing. Whenever new scientists with novel ways of viewing issues and different takes on theory, methodology, and interpretation emerge in a field it usually augurs well. What I found more than a little fascinating was how much research on implicit learning and related processes in the cognitive unconscious was carried out in Eastern Europe, in particular Russia. Much of the work, which is described in chapters that comprise roughly half of the contributions to this volume, was, of course, published in Russian and well outside the scholarly sweep of most Western cognitive psychologists. Until this volume landed in my inbox, I must admit, I was totally unaware of this work – though I am pleased to see that most of the findings reported fit within and extend the general theoretical framework we have become familiar with. I cannot help but wonder how and in which directions research on these topics might have developed had these publications been more readily available and had we known about the depth and richness of the work of, for example, Viktor Allakhverdov and his colleagues at St. Petersburg University. A round of thanks goes out to Cleeremans and his collaborators and co-editors for bringing this literature to the attention of what will certainly be a welcoming audience.

The other half of the contributions are reports, reviews, and analyses from more familiar, Western laboratories. P. J. Reber and co-authors give us an in-depth historical overview of the empirical and theoretical work on the cognitive unconscious with due attention to the underlying neuro-correlates. San Anton, Cleeremans, and Destrebecqz's chapter reveals what most of us have long suspected, but without much empirical study, that learning to function effectively in complex, multi-process domains requires a delicate mix of the implicit and the explicit, a combination of top-down, code-breaking strategies and a bottom-up unconscious extraction of associative patterns. Norman and her co-workers explore the limits to strategic control of implicit functions. Kemény and Lukács review the extensive literature on whether or not abstract representations underlie implicit knowledge. All are most welcome.

But, as is my wont when given the kind of free rein that comes with writing a foreword, I have a couple of points to make. I note that in the Introduction and several chapters (in particular Moroshkina and colleagues) there are discussions of the oft-raised question about whether humans can learn without at least some conscious awareness. I suppose this isn't unreasonable since there are still scientists interested in the cognitive unconscious and related issues that hold fast to this position. In my way of thinking, however, the studies that show some explicit "leakage" or evidence that participants in implicit learning experiments have some reportable knowledge say more about methodology than about underlying mechanisms and processes. They aren't challenges to the ontological status of implicit learning for the simplest of reasons: implicit learning is routinely observed in populations where explicit functions are either not operative or have been degraded by circumstance. Children learn language without awareness of either the process of acquisition or knowledge of the underlying syntactic, semantic, and paralinguistic aspects of what they have learned. We all become socialized in the mores, rules, traditions, and practices of the culture we are raised in. Patients with virtually complete anterograde amnesia function essentially normally in implicit learning and memory studies. There is compelling evidence for intact implicit functioning in a number of special populations including Williams Syndrome children, individuals with autism spectrum disorder, patients in psychiatric institutions, and aged individuals whose explicit cognitive functions have declined.

All these instances of learning take place largely (elderly participants) or completely (infants) outside awareness of either the processes or products of acquisition. Just a moment's contemplation, a few seconds of recognition of what happens during childhood or in these other, special populations, should make it clear that implicit learning is real and plays a compelling role in acquiring abstract and complex knowledge about abstract and complex domains. As San Anton and co-authors hint in their chapter, virtually all the interesting things that humans learn to do is a blend of implicit and explicit operations and, as P. J. Reber and collaborators note, we're beginning to get a pretty good idea about the underlying neuro-correlates of these functions and where and how they operate distinct from those modulated by top-down, explicit functions.

I suspect that much of the skepticism seen in some quarters comes from a tendency to be dependent on data collected from adult, college and university educated participants. When these cohorts are the ones in the experimental spotlight you're likely to see top-down functions being engaged. It would be, however, an existential error to conclude that the presence of some explicit and reportable information challenges the ontological status of the cognitive unconscious.

Along these same lines, I worry about a general failure to take into account a distinction made over six decades ago by Michael Polanyi between knowing "that" and knowing "what." We're often in situations where we know for certain that we know something in the absence of knowledge of what is it we know. This sentence now you've reading be nongrammatical. You spotted that instantly and, I'm certain, you did so while being utterly unaware of what syntactic rules were violated. You can repair the sentence because, again, you know that certain changes need to be made but in the absence of the specific syntactic rules you're using to do so. These kinds of things happen routinely in daily life but are rarely encountered in the typical, tightly controlled experiments carried out to examine mechanisms and processes expressed by the cognitive unconscious.

But, no matter. These are minor quibbles. This volume marks a significant contribution to an area of research that's now over a half-century old and, with the innovative and provocative contributions of the Eastern European laboratories, is primed to make even more compelling contributions in the years to come.

Arthur S. Reber
Point Roberts, WA

INTRODUCTION

Axel Cleeremans, Viktor Allakhverdov, and Maria Kuvaldina

Research about implicit learning, that is, broadly defined, learning that takes place in the absence of intention to learn and in such a way that the acquired knowledge cannot be easily verbalized, was initiated about fifty years ago with the publication, in 1967, of Arthur Reber's article about artificial grammar learning. The first aim of this book is to take stock of the state of the field today. Where do we stand insofar as the very concept of implicit learning stands? The second goal of this volume is to give broader visibility to work carried out in Eastern European countries, in particular, Russia, Poland and Hungary, each of which includes an active community of cognitive scientists interested in the complex relationships between what we can and cannot do without consciousness.

The volume itself originates in a series of informal meetings held since 2012 in different spots in Europe. The first such meeting, dubbed the "Implicit Learning Seminar", was organized at the end of August 2012 at the Warsaw School of Social Psychology in Sopot (Poland) by Agnieszka Poplawska and colleagues. About twenty people, many of whom contributed to this book, attended the meeting, which felt like a first of sorts. Indeed, while most of the people involved in the Sopot meeting identified themselves with implicit learning, the field itself never had a regular meeting of its own. This may have something to do with the fact that implicit learning, despite the substantial influence that the concept itself exerted on many other domains, was always a comparatively small field. More on this later; suffice it to say now that implicit learning research is conceptually complex, methodologically arduous and theoretically controversial.

The second Implicit Learning Seminar was organized in Bergen (Norway) by Norman and colleagues, at the end of June 2013. A slightly larger group of colleagues assembled at the University of Bergen under the long days of the Norwegian summer to exchange views about their current research.

This was followed by the third seminar, which had a decidedly different character than the first two meetings, for it was much larger. Organized in May 2014 at the Psychology Department of Saint Petersburg State University (Russia) by Viktor Allakhverdov and colleagues, in particular Maria Kuvaldina, the meeting assembled over a hundred people and was subtitled "Interactions between consciousness and the unconscious". During the meeting, three things became clear. First, there is a lot to learn about science through comparing approaches and perspectives stemming from different research traditions. Second, the Saint Petersburg meeting generated so much interest from all parties involved that it became clear that this volume was needed. Third, it was also during that meeting that the decision to continue the seminar series was taken. In 2015, the meeting returned to Poland, this time in Krakow under the supervision of Michał Wierzchoń. The fifth meeting was again much larger. Organized by Patrick Rebuschat in Lancaster, it welcomed not only the small implicit learning group but also the much larger community of researchers interested in statistical learning, featuring for instance Morten Christiansen and Linda Smith. Statistical learning, since the seminal work of Saffran and colleagues in 1996, has become a field of its own, and one that is also, in different ways, divorced from the earlier implicit learning community. For while the main issue that focused much of the earlier efforts was consciousness and its role in adult learning, the statistical learning community was from the start more interested in language acquisition and in its underlying mechanisms. Consciousness, per se, remains a side issue for the statistical learning community. The meeting in Lancaster highlighted some of these differences in approach and proved a fertile ground for interactions between the two groups. The sixth meeting, held in Budapest at ELTE University, was organized by Dezső Németh at the end of May 2017. Again a smaller meeting, it involved a slightly different mix of researchers, featuring in particular people interested in sleep and in memory consolidation. Finally, the seventh meeting was held in Cluj-Napoca (Romania) in the psychology department of Babeş-Bolyai University during May 2018. Organized by Adrian Opre and colleagues, it was again an outstanding opportunity for the group to exchange ideas and find out about ongoing research.

Out of this series of events emerges the sense of a tight-knit community of like-minded researchers – some of whom have been active in the field for decades – as well as junior scientists who see implicit learning as a way of exploring some of the most fundamental questions in cognitive science, amongst which consciousness, but also, centrally, the nature of mental representation. Implicit learning, quite uniquely, is a field that stands at the crossroads between these two fundamental questions, as the relevant paradigms typically involve incidental learning tasks that engage comparatively sophisticated processing (i.e. complex decisions, abstract knowledge).

How did the field come about over its short history? There are but a few landmarks along the way. 1993 stands out, for it is when A. S. Reber published his book titled *Implicit Learning and Tacit Knowledge: an essay on the cognitive unconscious* – probably the best rendition of his work over the previous 30 years. Cleeremans's monograph, also published in 1993, was the first to offer a computational theory of implicit learning. This was closely followed by Berry and Dienes' (1993) volume, which offered a sensible overview of the field at the time, as well as a tribute of sorts to the influence of Donald Broadbent on the development of ideas in this domain. The first journal review of implicit learning as a field was published by Carol Seger in 1994. In 1996, Saffran and Aslin published their report on "Statistical learning by 8-months infants" – a *Science* article that would veer the field in a completely different direction, focused more on cognitive development than "traditional" implicit learning research. In 1997 Diane Berry published a collection tellingly titled "How implicit is implicit learning?" The next year saw the publication of the first true compendium of implicit learning research in the form of Stadler and Frensch's *Handbook of Implicit Learning*, as well as that of another journal review of the field (Cleeremans et al., 1998). French and Cleeremans' (2002) collection, titled *Implicit learning and consciousness*, marked the progressive merging of the field of implicit learning into the then booming literature dedicated to consciousness itself. Luis Jiménez produced another collection of chapters for his 2003 volume titled *Attention and Implicit Learning*. Patrick Rebuschat edited a collection of articles dedicated to *Implicit and Explicit Learning of Languages* in 2015, with an interesting mix of authors from the implicit learning and statistical learning fields.

Alongside this substantial activity, a number of influential articles also appeared in *Behavioural and Brain Sciences*. Of note here are several articles by David Shanks and colleagues (Shanks and St. John, 1994, Shanks and Johnstone, 1998, Newell and Shanks, 2014), all of which offered critical overviews of the field, as well as in-depth treatments of theories of implicit knowledge (Dienes and Perner, Perruchet and Vinter).

This is not the place to engage in a detailed history of implicit learning research, but it is worth emphasizing that the development of ideas in this domain has followed distinct courses in the West and in the East. We briefly overview each in the following, noting along the way how some "Western" research findings had in fact already been observed in the East well before.

Implicit learning in the West

In Chapter 1, Paul Reber and colleagues offer a cogent historical overview of implicit learning research in the West that we will not reiterate here. They trace the origin of the field to Arthur's Reber's chance encounter with George Miller at a time when Chomsky was revolutionizing linguistics. This was also the time – the 1960s – when psychology was shifting away from behaviorism to embrace theoretical approaches rooted in the idea that the mind contains rich sets of abstract representations through which the more complex abilities of human beings – foremost, language and abstract thought – can be carried out. And yet, as Paul Reber discusses, the problem of how such abilities can come to be remained intact, and in fact, seemed rather intractable. Solving this problem is what motivated

Arthur Reber to begin exploring simple "language–learning–like" situations in the laboratory. Interestingly, Arthur Reber (personal communication) himself notes that his work was not so much, if at all, influenced by Freud – to the contrary, Reber himself sought to distance himself from anything having to do with psychoanalysis, hence the expression "implicit learning" rather than "unconscious learning". The original motivation to explore the mechanisms behind implicit learning was thus driven by an attempt to answer the question of how humans are able to learn language. But it is fair to say that the main reason why implicit learning proved so enduring a field is the question of awareness and its role in learning.

The idea that behavior can be driven by contents of which we are not aware is an old idea that remains controversial even today. It is of course to Freud and some of his predecessors, in particular the oft-ignored Eduard Von Hartmann, that one owes the very notion that there exists an "unconscious mind", that is, that the mind itself contains contents of which we remain unaware. This, in and of itself, however, says little about what really counts, namely the idea that the choices we make can be driven by unconscious determinants. In other words: is it the case that perception, action and even learning can take place outside of conscious awareness?

That very possibility is the one claim that elicits the most debate, and it is very difficult to state with any certainty that it has been solved. Subliminal perception remains controversial. Implicit learning, as explored through its different paradigms, remains controversial. Behavioral priming, particularly in social psychology, remains controversial. All these phenomena continue to elicit lively debate in their respective literatures. Why is that the case?

The answer is straightforward: All these phenomena are suggestive that processing can take place without awareness – a fascinating idea in its own right. But by the same token, none of these phenomena have been demonstrated with sufficient force that skeptics can be laid to rest (e.g., Shanks and St. John, 1994). The greatest empirical challenge here is probably that it is impossible to prove the null: absence of evidence is not evidence for absence, and hence, finding that some task can be carried out in the absence of awareness can often be explained by lack of sensitivity in the measures used to assess awareness, in participants' biases, and so on. Point well taken – the methodological issues raised by critics of this literature are real and worrisome. But the field has now moved on in substantive ways that make it possible to address such worries.

This being said, it is undeniable that cognitive systems can exhibit what Dennett, in his latest book (Dennett, 2017), has cogently dubbed "competence without comprehension". All sorts of living systems exhibit competence without comprehension. A bacterium argues Dennett, is competent in its own little *ümwelt*, but exhibits neither comprehension of what makes it so capable nor of much of anything else. Dennett writes "We know there are bacteria; dogs don't; dolphins don't; chimpanzees don't. Even bacteria don't know there are bacteria" (Dennett, 2017, p. 5). Artificial intelligence also perfectly illustrates this point, as some algorithms are now capable of learning to achieve superhuman performance in restricted domains such as the game of Go or Chess, while remaining not only wholly incapable of explaining their strategy but also completely unaware of their own successes and failures, or indeed of their own existence. There is thus no doubt that competence can take place without comprehension. The question is how, and more specifically, based on what kinds of processes and on what kinds of mental representations. One particular conundrum that faces all research on implicit cognition in general is the fact that awareness cannot be turned off. Hence, in the vast majority of cases, performance will always be driven by a mixture of implicit and explicit processes. Here again, implicit learning provides a rich ground to explore such fundamental issues. Today, the field is fundamentally connected to the core issues that concern the study of consciousness.

Implicit learning in the East

In the East, however, the development of ideas followed a rather different course, shaped as it was by Soviet ideology. The study of the phenomena of the *conscious and* the *unconscious* was nearly impossible during the Soviet era. There was merely a brief period, right after the revolution, when the revolutionary euphoria allowed new ideas and new people to appear and burst onto the scientific scene. In 1924, A. Luria – at the time 22 years old himself and already a secretary at the Institute of Psychology in Moscow! – listened to a talk given by the 27-year-old Lev Vygotsky "Consciousness as a Subject of the Psychological Study". Luria had invited Vygotsky, a provincial literary and theater critic, to work at his Institute, where Vygotsky, despite his short life, would become a world star. Some of Vygotsky's research ideas became a prelude to the study of the implicit learning. Using a modified version of Asch's method of artificial concept creation, Vygotsky discovered that at a certain age, children are able to categorize objects correctly without yet being able to explain how. In 1974, these findings were confirmed by Petrenko. Petrenko presented participants with geometric shapes that were either congruent or incongruent with the concept set by the experimenter. When a geometric shape was congruent with a concept, its presentation was reinforced by a slight electric shock. The findings show that even when participants had failed to categorize a shape correctly, they would exhibit a defensive reaction to it (Petrenko, 1974).

Vygotsky himself did not put much emphasis on the significance of his findings: the implicit learning responses. He was more interested in understanding the role that consciousness plays in cognitive processes. Overall, all major Soviet psychologists were trying to build an integrated, overarching theory of the human mind – a task that seemed impossible to accomplish without understanding the role of consciousness. Soviet ideology supported this approach based on the main claim that consciousness is shaped by society and culture. Therefore, it appeared plausible to promulgate a "new consciousness", based on which a new, better human could be shaped – a human that would "build" Communism. Moreover, the dominant direction of the entire Soviet scientific endeavor consisted of the building of large-scale general application theories.

During tempestuous post-revolutionary times, psychoanalysis became increasingly popular and interest in research on the unconscious grew rapidly. However, at the same time Marxism grew into an all-consuming State ideology (in its most dogmatic sense). In order to continue research, many attempts were made to combine psychoanalysis and the ideas of Marxism. These attempts proved unsuccessful. To make matters worse, Leon Trotsky – then recently proclaimed an ideological enemy of the Soviet State – had been sympathetic towards psychoanalysis; this resulted in the official declaration of psychoanalysis as a "false bourgeois pseudoscience". Russian works on psychoanalysis started to disappear off the shelves, and reading Freud's works now required special access permission. The works of Vygotsky were taken out of print for many years and largely forgotten. In this respect, it is ironic that Vygotsky himself was a convinced Marxist, but he knew very well that Karl Marx was no psychologist – and that the new Marxist psychological theory had yet to be built. Any mention of the "unconscious" in the USSR was considered to be a deviation from the "only true teaching".

There was one exception, however. Far from Moscow, in Tbilisi, Georgia, a group of psychologists led by Uznadze studied the role of unconsciousness in perception. Uznadze (1958) had conducted experiments demonstrating that when a participant is presented with a pair of objects that noticeably differ in size, and the bigger object is always presented on the same side of the visual field, then when two objects of the same size are presented, the participant perceives them to be of unequal size. Uznadze called this effect a *fixed perceptual set*. It is important to note that the fixed perceptual set is produced even when the participant is not aware of the objects' unequal sizes (for example, when a participant was under a post-hypnotic suggestion, or when the size difference was not in the focus of attention). In other words, as we might say today, the experiment's participants had learned without awareness of what they had learned. Uznadze concluded that there is something connecting the physical aspects of the environment and the contents of consciousness which is not directly recognized by consciousness, but nevertheless has a decisive influence on its macro-contents. It was only in 1978 that Uznadze's followers managed to organize a major conference on the unconscious in Tbilisi, thus finally "legalizing" the usage of the term in Soviet psychology.

It must be noted that the fires of discourse on the unconscious were slowly kept up within another – and even officially sanctioned – field, namely the (favorite Soviet) study of reinforced reflexes initiated by the legendary Pavlov. Humans are not aware of the physiological mechanisms of reflexes, but these reflexes are somehow linked to conscious processes. It had been established that it took longer to develop a conditional reflex in participants who knew that they were being conditioned. But what happens when a participant is not aware of the conditioned stimulus? A. N. Leontyev (1981), the most renowned Soviet psychologist, studied defensive reflexes developed as a response to a green light projected on the palm (the red light signal was not reinforced). It turned out that the defensive reflex would develop only when the subject knew in advance that she was being conditioned by a signal of which she had otherwise been unaware.

Anokhin (1968) studied conditional reflexes in dogs. Once, instead of the dried powder meat meal (consisting of meat, flour and bones) that was usually used as reinforcement, a much more attractive food (real meat) was supplied. Far from being excited, the hungry dog – receiving an unexpected stimulus – turned away and refused to eat altogether. Anokhin interpreted this behavior as evidence for the existence of a predictive reflex of a future situation in dogs (a reflex expected to be even larger in humans), which suggests that previously accumulated, but not necessarily conscious, experiences continue to influence behavior. It is worth noting that in 1950, during the USSR Academy of Science meeting, Anokhin was taken to task for "serious deviations from Pavlovian doctrine".

Bernstein (1991), one of the brightest Soviet thinkers, managed to approach the problem of unconscious influences in the process of motor learning. He had realized that motor skills could not be acquired through repetition of the same movements because it is impossible to repeat each and every movement precisely. Even if it were possible, none of the clumsy movements executed at the beginning of the acquisition could lead to a developed skill. Training, in his opinion, is a repetition of a motor task without repeating the actual movement. Bernstein's experiments show an unconscious regulation of the motor task during skill acquisition. In essence, his position may be interpreted as follows: all learning processes involve both conscious and unconscious components; during learning of complex movements, conscious control switches from monitoring lower levels to monitoring progressively higher levels of movement regulation. Once a motor skill is acquired and becomes automatic, conscious control is no longer needed; attempts to consciously control a motor skill that has already been developed at lower skill levels would lead to deautomatization and disruption of the skill.

Political turnaround after Stalin's death allowed for fresh new ideas to appear. In the 1950s, Ponomarev, continuing Duncker's work, set to study the value of clues in the process of creative problem solving. Ponomarev showed that under some circumstances, the completion of a task unrelated to the main problem facilitates the solution of the target problem even when participants are unaware of such an influence. Based on his research, he posited the existence of a layer of experience that is unavailable for conscious recall by the subject. It is only when the participant begins acting to solve the problem that this covert experience may be triggered as a helping tool for problem solving. To solve the problem, one needs a conscious recognition of the target solution; however, the intuitive experience is formed beside conscious realization of the target. The contemporary Russian psychologist D. V. Ushakov sees Ponomarev's findings as anticipating modern ideas of implicit learning and implicit knowledge. Ushakov goes so far as to claim (Ushakov and Valueva, 2006) that Ponomarev had discovered the relevant phenomena 15 years before Reber's introduction of the concept of "implicit learning".

Ponomarev was not the only Soviet scientist who studied implicit learning. In the 1960s, Schechter conducted a series of experiments on visual recognition. In his experiments, participants were given a set of general features in order to identify whether or not the presented visual stimulus matched the target feature.

In the learning phase, however, only one example (one special interpretation) of a general feature was presented. For example, the explicit rule would state that one line should always be at an angle to another line. In the learning phase, a participant would only see lines at an acute angle. Then, in the experimental phase, participants would see stimuli containing lines angled at an obtuse angle as representing another example of the rule, but not the one he was exposed to in the learning phase. Participants would give a fast response that the figure is not congruent with the rule, but then, immediately change their decision, stating that it is. Schechter interpreted these results as evidence that the example of the rule was learned unconsciously even when a more general explicit rule was given, and then applied instead of the explicit rule.

In the early 1970s, V. M. Allakhverdov discovered a group of effects collectively dubbed "unconscious negative learning": people tend to repeat, without noticing that anything is amiss, their own errors when performing trivial iterated sensorimotor, perceptual, mnemonic and arithmetic tasks. Moreover, under certain conditions, participants showed a tendency to repeat even their omission errors: not to notice, not to recognize, not to recall exactly what had been not noticed, recognized nor recalled in previous trials. Allakhverdov coined a term for this: the aftereffect of negative choice. In his endeavors to explain this phenomenon, Allakhverdov developed his original model of consciousness (which is, in part, presented in the current volume).

Implicit learning itself became relevant during the 1970s, when studies of unconscious recognition of probabilistic sequences became popular in the USSR. I. Feigenberg measured the reaction time when observers saw a sequence of four digits (1, 2, 3, 4) presented one by one and were asked to press a corresponding key. Digits were presented in a uniform random order: a digit in the first place could be any out of four, the digit in the second place could be any of the three others left, the digits in the third and fourth places could be only the digits that were left. Participants were not aware of this sequence but their reaction time reflected the probabilities of the presented digits. The first digit in a sequence elicited the longest reaction time and the last digit in a sequence elicited the shortest reaction time. This result did not depend on whether participants were aware of the presented sequence or not. Later, Lee (1997) conducted the same experiment using a sequence of digits from 1 to 6, and found the same effect (A. Cleeremans even refers to it in his articles as the *Lee effect*).

At the beginning of 1980s cognitive psychology in the USSR was still regarded as a "bourgeois science". Studies of cognition would not resonate with the Marxist dogma, and could therefore not develop. By the end of the 1980s, the dogmatic labels were done away with; and the time finally came for Russian psychology to blossom. In the 1990s, the Russian psychologists finally obtained access to Western professional journals to review the studies firsthand and not from reinterpreted and inaccurate recounts.

At this time, studies of implicit learning in Russia become goal oriented and systematic. In Moscow, Ushakov's laboratory at the Psychology Research Institute of the Academy of Sciences, studies are conducted of the unconscious transfer of the skill acquired in one task onto another. This research attempts to combine Ponomarev's findings with Reber's ideas and findings. In Samara, Agafonov (and others, 2010) conducted a series of experiments on the solution priming effect in problem solving. Participants performed a lexical decision task with some of the trials being primed by items congruent with the solution of the task and some trials being primed by items incongruent with the solution of the task. When presented with a subliminal congruent priming for several dozens of trials, participants showed a gradual increase of the positive priming effect, i.e. they reduced the reaction time in a lexical decision task. When presented with subliminal incongruent primes for several dozens of trials, participants showed a gradual decrease of the priming effect. The positive priming effect became less pronounced when a series of subliminal congruent priming trials was preceded by a series of subliminal incongruent priming trials. Thus, when a participant received a series of valid pre-threshold primed cues, she relied on them more than when she received a series of the invalid ones. These results demonstrate that despite priming being subliminal, participants implicitly estimate the validity of the primed cue.

In St. Petersburg, a series of implicit learning studies have been conducted by Allakhverdov's research group. The results of these studies are represented in the current volume by two articles: latest experiments on the negative choice aftereffect–implicit negative learning are discussed by Kuvaldina and others; while Moroshkina et al. are concerned with the role of the participant's chosen strategy (analytic vs. holistic) in the acquisition of implicit knowledge, as well as with the factors triggering switches between the relevant strategies.

Outline of the chapters

This volume contains two broad sections. The first is devoted to overviewing the theoretical and historical aspects of implicit learning research; the second consists in a compendium of experimental studies on implicit learning conducted in Western and in Eastern Europe respectively. Together, the two sections offer a snapshot of the field as it stands today.

The theoretical section opens with a historical overview by Reber, Batterink, Thompson and Reuveni (Chapter 1). The context of the discovery of the phenomenon of implicit learning and some of the theoretical complications that ensued are described. Reber and colleagues warn against overlooking the importance of implicit learning. They contrast laboratory-based approaches to applied research, and give a broad overview of the usage of implicit learning principles in various research domains. The authors show how the fields of statistical learning, decision-making and skill acquisition use theoretical frameworks that presuppose the existence of multiple memory systems, and describe implicit learning as a central aspect of complex cognition.

In the next two chapters Cleeremans (Chapter 2) and Allakhverdov and colleagues (Chapter 3) offer overviews of their respective theoretical frameworks.

In his chapter, Cleeremans revisits some of the ideas associated with the radical plasticity thesis – the proposal that consciousness is something that the brain learns to do. The chapter begins with an analysis of the differences between conscious and unconscious information processing and continues with a brief overview of the core ideas of the radical plasticity thesis, which offers a way of rooting such differences in principles derived from Higher-Order Thought Theory. The chapter concludes with an overview of the different manners in which information can be made subjectively unconscious.

Allakhverdov, Filippova, Gershkovich, Karpinskaia, Scott and Vladykina provide an overview of the Theory of Consciousness created by Allakhverdov in 1970s. They put it together as a number of principles aimed at resolving a central paradox of learning: why do people, when they perform the same trivial action several times in a row, start to perform it better and faster? In a consistent series of theoretical claims, the authors defend the view that even before the process of learning, a person is implicitly able to do what s/he is learning to do, but is unable to explicitly realize this skill. The chapter describes the principles of independent verification that occur on several levels of cognition including implicit processing. The indirect mediator for verification is the signal of the accuracy of the unconscious decisions. However, this subjective signal is not translated to awareness directly. Conscious experience is characterized by avoiding contradictions with previous experience, hence, some of the previously rejected signals continue to be rejected despite their correctness. Allakhverdov introduces the idea of recurrent mistakes, which become possible only with an underlying process of implicit learning. Some broader implications of this idea are discussed.

The Experimental section of the current volume contains a number of studies conducted by the members of the Implicit Learning Seminars group over different years. Chapters are written by Russian, Polish, Norwegian and English collaborators and participants of the Implicit Learning Seminar.

The authors of Chapter 4 – Kuvaldina, Chetverikov, Odainic, Filippova and Andriyanova – continue the theoretical message of Allakhverdov et al. by presenting their research on recurring mistakes. The authors describe the framework of "negative choice", which assumes that we may implicitly learn our own mistakes in simple cognitive tasks. In a series of experiments, they demonstrate that recurring mistakes may happen in response to the same stimulus when this stimulus is presented several times. Recurring mistakes occur in different cognitive tasks involving either retrieval or recognition processes. This type of error leaves traces in memory that include information not only about the identity of the item that was associated with the erroneous response, but also the information about the location of this item and the previous response to this item. Recurring mistakes are different from the set of single mistakes: they elicit shorter reaction times than single errors. In addition, recurring errors correlate with higher confidence ratings in comparison with single errors. The authors compare their findings to a number of similar studies and conclude that the learning mechanism that allows for "erroneous" (from the experimenter's point of view) association of response and stimulus that is stored in memory for quite some time could be somewhat similar to the Hebbian learning rule and the bidirectional associative memory model.

In Chapter 5, San Anton, Cleeremans and Destrebecqz explore the tension between associative and inferential theories of learning. According to inferential theories, learning is the result of inferences and reasoning on propositional representations. As a consequence, awareness of a rule always precedes the corresponding behavioral changes. According to associative theories, however, learning involves the gradual updating of the associative strength between mental representations of the stimuli. On this view, consciousness is not mandatory for learning to take place. The authors test contrasted predictions from both theories in a video game-based paradigm in a series of two experiments. They conclude that their results support evidence of unconscious associative learning and embrace dual-process theories of learning.

Strategic control in implicit learning tasks is usually used as an indicator of conscious knowledge. Chapter 6, authored by Norman, Scott, Price, Jones and Dienes carefully disambiguates this statement. First, they show that strategic control does not necessarily entail awareness. An approach offered by Dienes and Scott (2005) and Scott and Dienes (2008) distinguishes between structural and judgement knowledge, i.e. knowledge of the rule and knowledge that this rule exists. Conscious structural knowledge implies conscious judgement knowledge. However, unconscious structural knowledge could be accompanied by either conscious or unconscious judgement knowledge, thus allowing for strategic control to be present in the task without participants being fully aware of it. With the support of this theoretical approach, the authors describe an experiment where they show that participants who did not express conscious structural knowledge nevertheless showed some ability to strategically control the application of two learned grammars on a trial-by-trial basis. Strategic control was measured through a "strategic score", a measure offered by Dienes, Altmann, Kwan and Goode (1995) that reflects ability to decide which of the two grammars to apply. The authors conclude that strategic control may occur even when participants express global unawareness of the nature of the rule that governs letter strings and respond on the basis of intuitive feelings, i.e., "implicit" decision strategies.

The question of whether implicitly acquired knowledge is abstract or not remains one of the most hotly debated issues in implicit learning studies. Kemény and Lukács (Chapter 7) look into the role of abstraction in sequence learning. The authors discuss different situations in which abstraction may influence the results of implicit learning: transfer, learning of category sequences, task complexity-related effects and stimulus-dependence. Skill transfer studies are described as an evidence in favor of abstraction. Nevertheless, the authors notice that some of the tasks used for this aim cannot surmount the shortcomings of a mapping problem, i.e. the degree to which pre- and post-transfer stimuli can be linked to each other. This problem certainly leaves the question of whether or not abstraction happens in implicit learning unsolved. The other variable that might be influenced by the level of abstraction is task complexity. The authors argue that due to the various degrees of complexity in the tasks used for implicit learning studies its inevitable that effects of abstraction will be observed in some of them. Finally, the authors point out that in most of the studies of abstraction it is assumed that the learning mechanism is modality- and domain-independent. To contrast this widely shared opinion Kemény and Lukács describe the studies of Conway and Christiansen and their own that presume stimulus-dependent and modality-constrained learning mechanisms of abstraction. In conclusion of this review the authors describe task requirements as playing a crucial role in determining abstraction of the acquired representation. The question about abstraction in implicit learning turns into a question about abstraction in specific tasks like Task Sequence Learning and some forms of Artificial Grammar Learning (AGL).

Another theoretical question that is imminent with respect to implicit learning is the measurement of whether the rule was acquired consciously or unconsciously. Moroshkina, Ivanchei, Karpov and Ovchinnikova (Chapter 8) wrote an extensive review of studies addressing this issue. Firstly, they provide a description of a variety of rule awareness measurements that include information and sensitivity criteria (Shanks and St. John, 1994), reliability and immediacy criteria (Newell and Shanks, 2014), exclusiveness and exhaustiveness criteria (Timmermans and Cleeremans, 2015), decision strategy attribution test (Dienes and Scott, 2005) and others. Secondly, authors argue that most of these approaches and criteria provoke verbalization and thus affect learning and application of implicit knowledge. They observe evidence for the effect of strategy, the effect of concurrent verbalization and the effect of retrospective verbalization that might block the application of implicit knowledge in subjects. Moroshkina, Ivanchei and Ovchinnikova claim that fade-out of the implicit learning effect that is observed in many studies happens as soon as subjects transit to conscious rule or regularity search, thus changing sensitivity to hidden rules and decision-making criterion accuracy. The authors conclude with several statements that appear to address the influence of goal-directed behavior, verbalization and attention selection on the implicit learning more carefully. Recognition of these issues may resolve a lot of contradictions in the implicit learning field.

Echoing the message discussed in the previous chapter, Popławska-Boruc, Sterczyński and Roczniewska (Chapter 9) address a role of a goal-directed learning in the AGL task. The authors point out that it is hard to believe that implicit learning is totally independent of the cognitive resources and mechanisms of attentional selection which are engaged in it. The testing phase of the AGL task and transferral of the tacit knowledge certainly requires cognitive resources as shown by a number of experiments described in the chapter. Another important factor that modulates results of implicit learning tasks is instruction or relevance of the feature to be memorized for the participant. The authors discuss experiments by Eitam and colleagues that demonstrated that implicit learning is related only to those features of the stimulus which are relevant to the goal of the task. However, in studies by Eitam and colleagues, the goals were externally set for participants – they had been informed beforehand what to focus on. In three experiments conducted by

Popławska-Boruc, Sterczyński and Roczniewska they demonstrated that participants did not need to be told directly to focus their attention on a particular feature to implicitly learn it. Moreover, participants were selective towards the features that were guiding their attention, i.e. they were more successful in learning the grammar of letters than the grammar of colors. Their results thus point to the resource-consuming nature of implicit learning, which calls for attention to be allocated to certain aspects of objects in the perception field.

Contrary to the conclusion of Popławska-Boruc, Sterczyński and Roczniewska, a chapter by Wierzchoń and Derda (Chapter 10) instantiates the opposite claim, that is, that attentional resources (attentional load) do not affect implicit learning. The authors first review the literature on "dual task" in implicit learning to clarify the circumstances under which discrepancies in implicit learning studies are observed. They conclude that discrepancies in the experimental results could arise from a weak manipulation of attention load in a secondary task and insufficient statistical analysis that doesn't, in most of the cases, include a test of null hypothesis by means of a Bayesian approach. To prove their point, authors present the data from four experiments of their own. In these experiments Wierzchoń and Derda vary the secondary task – Divided Attention, Random Interval Generation, Random Number Generation, Mental Arithmetic tasks – and calculate the probability of null hypothesis being true using a Bayesian approach. Two of their experiments used AGL as a primary implicit learning task, another two, a serial reaction time task. AGL experiments showed that a secondary task applied over an acquisition phase did not affect classification accuracy. Bayesian analysis conducted over these results presented evidence in favor of null hypothesis. The results observed in context of the serial reaction time task were less clear. Despite this, the authors conclude that their data adds to evidence supporting the thesis that attentional load does not affect implicit learning.

In closing, we very much hope this volume will stand as a snapshot of contemporary implicit learning research as it takes place fifty years after its inception, and as a unique opportunity for the many actors in this domain to become acquainted with work that has so far not been very visible in the West. It is a truism to state that science progresses by confronting different perspectives – but sometimes, even truisms ring truer than usual. . .

References

Agafonov, A., Kudel'kina, N. and Vorozheikin, I. (2010). Fenomen neosoznavaemoy semanticheskoy chuvstvitelnosti: noviye eksperimentalnie facti [The phenomenon of the unconscious semantic sensitivity: new experimental facts (article 1).] *In* Lisicki, K.S and Shpuntova, V. V. (eds.) *Psychological Studies: Collection of Articles 8*, pp. 6–25. Samara: "Univers-Group".

Anokhin, (1968). Biologia i neurofisiologia uslovnogo reflexa [*Biology and neurophysiology of conditional reflex*]. *Medicine*. Moscow.

Bernstein, N. (1991). O lovkosti i iyie razvitii [On agility and its development]. *Physical Education and Sports*. Moscow.

Berry, D. C. (ed.) (1997). *Debates in Psychology: How Implicit Is Implicit Learning?* New York, NY, US: Oxford University Press.

Berry, D. C. and Dienes, Z. P. (1993). *Implicit Learning: Theoretical and Empirical Issues*. London: Psychology Press.

Cleeremans, A. (1993). *Mechanisms of Implicit Learning: Connectionist Models of Sequence Processing*. Cambridge, MA: MIT Press.

Cleeremans, A., Destrebecqz, A. and Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416.

Dennett, D. C. (2017). *From Bacteria to Bach and Back: the Evolution of Minds*. New York: Norton & Company.

Dienes, Z., Altmann, G., Kwan, L. and Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1322.

Dienes, Z. and Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*(5), 735–808.

Dienes, Z. and Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*(5–6), 338–351. doi:10.1007/s00426-004-0208-3.

French, R. M. and Cleeremans, A. (2002). *Implicit Learning and Consciousness: an Empirical, Philosophical, and Computational Consensus in the Making*. London: Psychology Press.

Jiménez, L. (ed.) (2003). *Attention and Implicit Learning* (Vol. 48). Amsterdam, Philadelphia; John Benjamins Publishing.

Lee, Y. S. (1997). Learning and awareness in the serial reaction time task. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society: August 7–10, 1997, Stanford University* (Vol. 19, p. 424). Mahwah, NJ, London: Lawrence Erlbaum Associates.

Leontyev, A. N. Problemi vozniknovenia ochucheniy [A problem of sensation formation] // *Problemi Rzvitiya psihiki* (pp. 15–218). Moscow, 1981.

Newell, B. R. and Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *Behavioral and Brain Sciences*, *37*(1), 1–19.

Perruchet, P., Vinter, A. and Gallego, J. (1997). Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin & Review*, *4*(1), 43–48.

Petrenko, V. F. (1974). Dinamika semanticheskogo poiska: Isledovanie rechevih actov i mishleniaya [Semantic search dynamic: the study of the speech acts and thinking.] Alma-Ata.

Ponomarev, Yu. (1976). *Psihologia tvorchestva* [*Psychology of Creativity*]. Moscow.

Reber, A. S. (1993). Implicit Learning and Tacit Knowledge: an Essay on the Cognitive Unconscious (Oxford Psychology Series, No 19).

Rebuschat, P. (ed.) (2015). *Implicit and Explicit Learning of Languages* (Vol. 48). Amsterdam: John Benjamins Publishing Company.

Saffran, J. R., Aslin, R. N. and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Scott, R. B. and Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1264–1288. doi: 10.1037/a0012943.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*(2), 163.

Shanks, D. R. and St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367–395.

Shanks, D. R. and 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, US: Sage Publications, Inc.

Stadler, M. A. and Frensch, P. A. (1998). *Handbook of Implicit Learning*. Thousand Oaks, CA, US: Sage Publications, Inc.

Timmermans, B. and Cleeremans, A. (2015). How can we measure awareness ? An overview of current methods. In M. Overgaard (Ed.), *Behavioral Methods in Consciousness Research* (pp. 21–46). Oxford: Oxford University Press.

Ushakov, D. V. and Valueva, E. A. (2006). Parallelnye otkrytija v otechestvennoj i zarubezhnoj psihologii: primer intuicii i implicitnogo nauchenija [Parallel findings in Russian and international psychology on the example of implicit learning]. In *Nauchnye materialy mezhdunarodnogo foruma i shkoly molodyh uchenyh IP RAN/Razdel 1. Obraz rossijskoj psihologii za rubezhom* [*Proceedings of the international forum of the young scientists in IP RAS*].

Uznadze, D.N. (1958) *Experimentalnieye osnovi psihologii ustanovki* [*Experimental foundations of psychology of set*]. Moscow.

1

IMPLICIT LEARNING

History and applications

Paul J. Reber, Laura J. Batterink, Kelsey R. Thompson, and Ben Reuveni

Introduction and history

The term "implicit learning" was first published 50 years ago in a report titled, "Implicit learning of artificial grammars" (A. S. Reber, 1967). This paper described studies with a novel paradigm aimed to create a laboratory analogue of language learning. The new approach was based on using mathematical formalisms for stimulus creation that were similar to ones being developed to help understand human language function. Surprisingly, participants exhibited an unusual behavioral pattern in their learning process. They appeared to be learning to be sensitive to the statistical structure of the underlying formalism, but seemingly without any awareness that there were any underlying rules. This report established the possibility of a dissociation between learning that could only be exhibited through performance and more traditional learning and memory that was available to conscious awareness. Over the next several decades a wide variety of additional studies and many more novel paradigms were constructed to drive research into understanding the phenomenon of implicit learning (A. S. Reber, 1989; P. J. Reber, 2013).

The historical context in which the original report was published provides some insight into why this finding had such a widespread, enduring impact and how the idea of implicit learning came to be foundational to the modern characterization of memory systems theory (Squire, 1992) and the cognitive neuroscience of memory. The story of the basis of the original studies (A. S. Reber, personal communication) starts with a chance meeting between the author, Arthur S. Reber, and George A. Miller in the early 1960s. Miller had fairly recently published the seminal paper on "the magic number 7" and working memory (Miller, 1956) which is often cited as one of the core reports demarking the shift in the field of psychology away from behaviorism and to cognitive psychology, known as the Cognitive Revolution. Other notable publications also considered in the same vein include Broadbent

(1958), Newell, Shaw, and Simon (1958) and a review written by Chomsky (1959) highly critical of a book by B. F. Skinner (1957) titled *Verbal Behavior*.

The Cognitive Revolution was effectively a movement against and away from the Behaviorist school that had attempted to put psychology on a robust scientific footing through the use of simple, well-characterized tasks with quantifiable measures that allowed for robust, reliable experimental paradigms. In practice, this meant using tasks from the tradition of physiologists (e.g., Pavlov's conditioning research) that could also be studied in animal models. However, the extrapolation from animal cognition to human cognition has always posed some difficult questions, in particular when considering complex human cognition and especially the process of language, which is effectively unique to humans. Skinner's suggestion (1957) that language could be explained from reinforcement and conditioning studies was forcefully rejected by Chomsky (1959), implying that the study of human cognition needed a different approach.

The new approach favored by Chomsky led to his seminal work developing the field of computational linguistics. Early explorations of this work appeared in the *Handbook of Mathematical Psychology* (Luce, 1963) which includes three chapters authored or co-authored by Chomsky outlining how language production and comprehension might be modeled with formal grammars. Two of these chapters were co-authored with Miller, which provides some context for how Reber, as a graduate student at nearby Brown University, came into contact with these formalisms through Miller (at Harvard) via their occasional interactions.

While Chomsky's research program can be seen as characterizing mathematical formalisms that would account for human language production and comprehension, Miller and Reber were considering a separate but related problem. If these grammars were how humans accomplished language, how does a human acquire them? The formalisms seemingly required to account for language use appeared to be exceedingly complex and possibly entirely unlearnable, especially considering the cognitive abilities of newborns. One approach was to assume they were not learned, necessitating the existence of a pre-wired "universal grammar" embedded in the human brain (e.g., genetically endowed). Another approach was to try to capture this learning process in the laboratory using simplified "artificial grammars," which then led to the seminal finding (A. S. Reber, 1967) and observation of a novel type of human learning that might solve this "unlearnability" problem for language.

Researchers familiar with this history are aware that the idea of implicit learning did not immediately revolutionize the study of memory or language. In fact, for much of the next several decades, there followed a great deal of debate centered on the difficult problem of establishing the "implicit" part of this kind of learning. With a definition of implicit learning founded on "not available to consciousness," establishing even the existence of this phenomenon depends critically on proving a universal null, no awareness, which is an essentially intractable problem (Merikle, 1994). While experimental techniques and measurement approaches eventually began to provide guidelines for tackling this issue (Dienes and Berry, 1997), important support for the concept also emerged from ideas being developed separately and in parallel from research in neuropsychology and neuroscience.

Cognitive neuropsychology and systems neuroscience

At around the same time as the several famous publications in cognitive psychology that launched the Cognitive Revolution were published, a landmark paper in human cognitive neuropsychology was also reported. Scoville and Milner (1957) described the famous case of the patient H. M., who exhibited severe and selective impairment to his ability to acquire new conscious memories after bilateral medial temporal lobe (MTL) removal to treat otherwise intractable epilepsy. While there had been a few prior reports of selective cognitive loss following localized brain regions (e.g., the patients described by Paul Broca and Karl Wernicke in the 19th century) the theoretical model of the time was dominated by Lashley's (1929) theory of equipotentiality that hypothesized that any region of the brain could support high-level cognitive function. The case of patient H. M. established that memory was dependent on a specific neural region and did not arise from mass action of neural changes across the entire brain.

Research over the next 35 years characterized the structure and function of the memory circuitry within the MTL (hippocampus and adjacent cortical areas) and established that this system was critical for the acquisition and consolidation of memories for facts and events (Squire, 1992). Patients with damage similar to H. M. are unable to acquire new explicit memories, but are able to retrieve remote episodic memories of events that occurred prior to the damage to the MTL. More recent memories are partially affected by a temporal gradient of retrograde amnesia (see Lechner et al., 1999, for a review and history), leading to the development of a theory of memory consolidation dependent on a gradual process of memory strengthening and reorganization that depends on the MTL after initial learning.

However, detailed neuropsychological assessment of H. M.'s memory capabilities subsequently indicated that not all learning processes in his brain were entirely disrupted. Corkin (1968) and Milner, Corkin, and Teuber (1968) documented improvements in performance in procedural tasks (mirror tracing), maze learning, and picture identification from fragments. Shortly after, Weiskrantz and Warrington (1970) described a broader phenomenon of intact memory from fragmentary information in amnesic patients (priming) that would come to be known as "implicit memory" and very widely studied (Schacter, 1987). Together these findings indicated that another type of memory existed that did not operate in the same manner as memory for new facts and events that depended on the MTL memory system.

These findings were foundational to the development of a "memory systems framework" that aimed to connect these observations about human memory to research going on in parallel on the neuroscience of memory. The field of neuroscience also progressed remarkably over the course of the 20th century (c.f. Gross, 1999 for a highly readable overview) with a notable moment in this progression being the founding of the Society for Neuroscience in 1969. With respect to specifically the neurobiology of learning and memory, an important early paper was the work of Kandel and Spencer (1968), who began to characterize the underlying biology of synaptic change in the nervous system. It is of note that all three of these then independent lines of research on learning and memory saw significant results in a similar time frame in the second half of the 1960s. However, integration of the related ideas across these research areas did not emerge until somewhat later during the development of the interdisciplinary field of cognitive neuroscience.

A great deal of neuroscientific memory research through the subsequent years was focused on establishing and characterizing the role of the MTL in explicit, declarative memory (facts and events). While observations from patients such as H. M. were fascinating, it was understood that it would require the establishment of a model system to be able to characterize how MTL damage affected memory with experimental control. The roles of the hippocampal formation, the adjacent cortical areas (entorhinal, perirhinal, parahippocampal), and the amygdala were all studied in detail (Squire, 1992). Systems-level analysis eventually converged on the key importance of the hippocampus and the adjacent cortical areas with the amygdala playing largely a modulatory role related to emotional memory. In addition, examination of the phenomenon of retrograde amnesia following MTL damage led to the characterization of memory consolidation processes as a key feature for how the MTL operates to store information.

Evidence for consolidation theory was also accumulating in parallel in research on the neurobiology of synaptic change (McGaugh, 2000). Synergy across these areas demonstrated how cellular and systems neuroscience could inform each other in building a theory of memory (Milner, Squire, and Kandel, 1998). Connections to research on psychological phenomena directed at studies of complex cognition were not immediately evident. Animal models do not allow for research on processes related to language or subjective measures of consciousness. Instead, many of the paradigms used to characterize and quantify learning and memory processes in these animal model systems were closely related to the tasks developed by the Behaviorist researchers (e.g., conditioning models of learning) which were very well suited to neuroscientific study of learning and memory.

Implicit learning and the problem of assessing awareness

Studies of implicit learning through two decades following the original description of the AGL task aimed to better characterize this kind of learning (A. S. Reber, 1989) but struggled with the question of how to firmly establish when learning was outside awareness. Assessing a lack of awareness depends on an accurate model of the information learned by participants to guide assessments of conscious knowledge. Dulany, Carlson, and Dewey (1984) and Perruchet and Pacteau (1990) found that asking participants about the letter strings used in the AGL paradigm specifically elicited some additional knowledge related to determining whether the strings followed the grammar rules or did not. This raised the possibility that participants were inferring another type of representation that allowed them to make "grammaticality" judgments without being aware of the specifics of the formal grammar. However, it was also possible that these assessments were not of the awareness of the knowledge that drove the grammaticality judgment, but reflected concomitant explicit memory for the study stimuli (which would naturally be acquired by cognitively healthy participants but might not contribute to AGL performance).

Similar questions were being raised about studies of implicit memory (e.g., Roediger, 1990). To show that this type of memory did not depend on explicit memory for previously seen stimuli, it would be necessary to show robust priming in the absence of conscious memory. In cognitively healthy participants, this proved to be extremely difficult as a participant with an intact MTL memory system will always have some explicit memory of the study items. The inability to show a strong dissociation made it impossible to rule out the hypothesis that implicit memory phenomena simply reflected a weaker form of explicit memory (similar to familiarity) rather than a separate form of memory entirely.

A new paradigm for studying implicit learning was described by Nissen and Bullemer (1987), the Serial Reaction Time (SRT) task that became quite widely popular. This task embedded a covert repeating sequence into a simple choice reaction time task. Participants were found to increase their speed of responding to a practiced sequence compared with unpracticed sequences without seemingly being aware of the repetitions. In addition to the dissociation with awareness, this paradigm was also shown to exhibit intact learning in memory-impaired patients (Korsakoff's) in the original report. Like with the AGL paradigm, concerns emerged over the content of the representation (Reed and Johnson, 1994) which led to protocol improvements without changing the basic character of the finding. However, the development of increasingly sensitive measures of explicit sequence knowledge (Perruchet and Amorim, 1992; Willingham, Greeley, and Bardone, 1993) started to show the same pattern observed in other tasks used in implicit memory research. Participants with intact explicit memory tended to have at least some memory for the covertly embedded (implicit) information, even if it was not clear that it contributed to task performance.

Memory systems theory

The emergence of an integrated memory systems theory that used an interdisciplinary cognitive neuroscience approach eventually showed how the neural basis of memory function in the brain could be used to help understand the type of learning observed in implicit learning paradigms. Squire (1992) described a taxonomy of memory types within a single major subdivision based on the importance of the MTL memory system. Declarative memory referred to information that required the MTL memory system to store (and consolidate) and produced representations that were generally available to awareness and verbal report. Nondeclarative memory described a collection of other phenomena that did not depend on the MTL memory system but were instead supported by synaptic change in other circuits.

Applying this framework to phenomena of implicit learning, Knowlton, Ramus, and Squire (1992), and Knowlton and Squire (1996) showed that as predicted, AGL was intact in patients with severely impaired memory due to MTL damage. P. J. Reber and Squire (1994; 1998) established the same parallel finding for the SRT task with techniques in protocol design and awareness assessment that had been advanced since Nissen and Bullemer (1987). Research on implicit memory with particularly severely memory-impaired patients indicated that it was possible to observe intact priming in the complete absence of explicit (declarative) memory for stimuli (Hamann and Squire, 1997; Stark and Squire, 2000). In each case, the tasks studied with cognitively healthy participants characterized as implicit learning, were the same as those that neuropsychological studies showed an important role for nondeclarative memory. P. J. Reber (2013) reviewed these areas and described a general framework for memory based on the MTL memory system together with general, pervasive neuroplasticity mechanisms that shape processing everywhere else in the brain to adaptively improve functioning via practice (repetition).

This framework provides a neurocognitive foundation for studies of memory that depend on implicit or explicit learning, or a complex interaction between the two types of memory. It also allows for a theoretical approach to the small handful of exceptions in which memory phenomena that appear implicit with cognitively healthy participants appear to depend on the MTL memory system. The contextual cuing paradigm (Chun and Jiang, 1998) has been used to study implicit learning in attentional search such that improved search performance occurs with repeated stimuli, even when the participants are unaware of the repetition. However, this type of learning is disrupted with hippocampal damage (Chun and Phelps, 1999). The pattern is similar to observations from a paradigm of "priming of new associations" (Graf and Schacter, 1985; Shimamura and Squire, 1989) that described a type of priming that was not preserved in amnesic patients. However, if the mechanism of implicit learning is pervasive throughout the brain, we can expect that it would apply even to shaping representations that were initially acquired from MTL-based (explicit) memory processes. This type of process would also support the statistical effects on explicit memory retrieval processes hypothesized by Anderson (Anderson and Milson, 1989) to account for how human memory adaptively responds to the observed demands of the environment.

Allowing for this interplay between types of memory allows for a very flexible theoretical account of a wide variety of observed human memory phenomena. However, it is based on a different approach than the original findings of robust dissociations between types of memory and might be criticized as exceedingly difficult to falsity. Even though the description is consistent with a very wide range of findings across memory systems research, it does not directly rule out alternate hypotheses. The primary alternate view of memory has historically been that human memory is largely based by a single system with the idea that this more parsimonious approach needs to be ruled out before accepting the more complex memory systems framework (Shanks and St. John, 1994; Nosofsky and Zaki, 1998). A single system or type of memory is largely inconsistent with neuroscientific observations of memory and the many systems demonstrating synaptic plasticity. However, a skeptic might suggest that although there is clearly neuroplasticity in the brain that operates outside awareness, the cognitively important aspects of human cognition depend exclusively on operations of explicit memory. Because human implicit learning phenomena have traditionally been studied with artificial paradigms aimed to dissociate implicit and explicit memory, it could be suggested that implicit learning is merely a vestigial reflex or a trick that can be elicited in the psychology or neuroscience laboratory.

To address this concern, it is necessary to examine how implicit learning affects cognitive behavior in designs that better capture the demands of memory imposed by activities in the world outside the laboratory. The utility of the memory systems framework needs to be shown as leading to a better understanding of complex learning processes and should be driven by a program of research in Applied Implicit Learning. This will entail eventually moving past reliance on the creative and unusual learning and memory paradigms (e.g., AGL, SRT) that were highly effective for isolating types of memory and developing the scientific framework. Among the immediate challenges for this new approach is that a theory of memory systems interactions is needed (e.g., Nomura and Reber, 2012) that the focus on dissociation has often overlooked (with notable exceptions, such as Poldrack et al., 2001).

In the remainder of this review, three research areas will be presented in which there is already evidence of influence of the core ideas behind implicit learning and memory systems and in which it appears further integration of the neurocognitive framework will be valuable. The first of these, "statistical learning" (Saffran, 2003), reflects a research area very much in the same tradition as the original AGL paradigm aimed at understanding the automatic extraction of statistical regularities to support language learning. Second, the process of "skill learning" and performance also naturally incorporates ideas about separate forms of learning from explicit instruction and repetitive practice. The memory systems framework captures these descriptions well and can guide theoretical accounts of the development of skilled expertise. Third, research on decision making (Tversky and Kahneman, 1975) developed in parallel a structurally similar multi-system approach to differentiate processes for rapid, intuitive decision making and slower, deliberate reasoning. This approach maps on fairly well to the memory systems framework and highlights interesting questions about the interaction of systems. This framework has been highly valuable in helping to understand certain classes of errors where implicit learning can lead to implicit bias affecting judgments. Across these three areas, consideration of the roles and interplay of multiple types of memory allows for better characterization and understanding of complex, real-world, human learning processes than can be supported by a simple, single system theory.

Statistical learning and language

The original AGL paradigm used to introduce the idea of implicit learning was developed in response to the introduction of computational linguistics. If human language can be represented in formal structures (finite state machines) that account for important aspects of syntax, how are these structures learned? While the AGL paradigm explicitly represented the underlying formal grammar structure, a different approach to the same idea was taken by Saffran, Aslin, and Newport (1996) with a paradigm described as "statistical learning." This approach used much simpler stimuli but was designed to be used to assess how pre-verbal infants could extract statistical structure from auditory speech-like input. The findings that emerged from this field of research were strongly influenced by considerations of the formal linguistics model of Chomsky (Saffran, 2003), just as the original A. S. Reber (1967) paper was. The paradigm developed by Saffran and colleagues focused on the statistics embedded in speech that could be used to determine word boundaries, rather than the syntactic structure implied by an AGL, and were designed to be amenable for developmental studies with pre-verbal infants.

In the statistical learning paradigm, infants (or other participants) listened to 2–3 minutes of artificial speech (synthesized) that contained an essentially undifferentiated stream of syllables. Statistical structure was covertly embedded by constraining the transitional probabilities between syllables in a manner similar to natural speech. In natural speech, phonemes within words are highly constrained but phonemes at the end of a word (on the boundary) can be followed by the initial phoneme of a much wider range of possibilities. After familiarization with artificial phoneme streams following this structure, infants exhibit differential preferential looking to stimuli that follow or violate this statistical structure. By careful control of the underlying frequency and conditional probabilities (Aslin et al., 1998) in a manner reminiscent of the controls discovered to be necessary with the SRT task (Reed and Johnson, 1994), it was established that these very young infants were essentially computing the transitional conditional probabilities among phonemes.

The statistical learning paradigm established a key idea behind the original AGL paradigm in that it showed that pre-verbal infants, in the process of natural language acquisition, exhibited a sophisticated learning ability that could support key aspects of language learning. In addition to the findings showing that word boundaries could be statistically extracted from continuous auditory input, additional findings extended this type of learning to more abstract relational rules (Marcus et al., 1999) and to some kinds of non-adjacent dependencies (Newport and Aslin, 2004). These paradigms do not attempt the complexity of formal linguistic structures necessary to acquire and produce well-formed, syntactic language. However, the statistical learning findings do show a core learning ability that emerges from experience and shapes processing of auditory input to support language processing. Of particular note, this implicit learning mechanism is available and relatively computationally complex even in young infants who are acquiring language.

Extensions of this line of research further suggested that statistical learning ability is not restricted to linguistic stimuli, with statistical learning being exhibited by infants in the visual domain as well (Fiser and Aslin, 2002; Kirkham et al., 2002). Using paradigms that parallel the auditory presentation of covertly embedded statistical information, infants and adults exhibit sensitivity to this structure in sequences of visual objects (Fiser and Aslin, 2002; Turk-Browne et al., 2005). These findings suggest that the ability to extract statistical structure are present across sensory modalities, generally supporting the idea of widespread neural plasticity supporting implicit learning to reshape processing throughout the brain (P. J. Reber, 2013).

Although the statistical learning paradigm was also extended to adults, attempts to assess the conscious accessibility of the statistical structure did not immediately follow. Since this research area emerged from developmental studies, the tools developed to assess awareness of learning were not applied to the adult learning paradigms. Even so, the commonalities between implicit and statistical learning were noted as likely emerging from the same underlying mechanism (Perruchet and Pacton, 2006). Batterink et al. (2015) systematically evaluated contributions from both implicit and explicit memory to statistical learning to support the idea that even in adults, this form of statistical learning depends on mechanisms that support implicit learning.

While statistical learning is able to play an important role in language learning, it is clear that not all language processing depends on being or can be learned entirely implicitly. Some crucial elements such as reference and word meaning seem to depend on the MTL memory system that is better suited to supporting memorization of the connection between a vocabulary word and its referent. This observation has led to the description of language processing as depending on both kinds of memory (Ullman, 2004; Paradis, 2004) contributing materially to different aspects of this complex process. Morgan-Short et al. (2010) applied this theory to questions of second language acquisition suggesting that a multiple systems model of language acquisition can provide valuable insight into how a second language is learned.

As seen across the three areas of "applied implicit learning" considered here, connections of implicit learning and memory systems theory to non-laboratory applications generally require considering both types of memory and also potential interactions between memory systems. Considering learning a second language as an example, we would hypothesize that the new syntactic structures to be learned might be best acquired by high levels of exposure to speech to allow for statistical learning to proceed. Memorization of new vocabulary would be facilitated by strategies that facilitate explicit learning (e.g., deep semantic encoding). However, an unanswered question in this area is how these two types of memory interact during the learning process. Do statistical learning and word memorization support each other, proceed independently, or even interfere with each other? To date, within language studies questions about system interactions have not been thoroughly explored. As in many areas within implicit learning, the drive to isolate this type of learning has led to the development of tasks aimed at separating memory types rather than examining interactions.

Skill learning

A research area in which the potential importance of interactions among memory types has begun to be considered is the acquisition of expert skill. While skills are often initially learned with some explicit instruction, the importance of practice in acquiring expert levels of skilled performance has long been understood. What is learned during the process of repetition is not easily available to conscious awareness but accrues through experience. Early research in psychology aimed to characterize this process of skill learning and improvements in performance due to repeated practice (e.g., Fitts, 1964). The course of learning measured as performance improvements from practice has been extensively studied and is often described as following a power-law (Newell and Rosenbloom, 1981; or some similarly negatively-accelerated curve) that continues over remarkably extended periods of time, even up to millions of repetitions (Crossman, 1959). Within this field, there are active debates over the role of rote practice, structured deliberate practice (Ericsson et al., 1993) and other factors (such as talent) that predict expertise (Campitelli and Gobet, 2011). However, this process is fundamentally a memory phenomenon that must be supported by the learning and memory mechanisms of the brain.

There is a basic assumption embedded in any approach based on practice that the information acquired during practice could not have been acquired by explicit, verbal instruction, which would otherwise be much more efficient. The information learned during practice is generally not available to later verbal report, suggesting that implicit learning mechanisms are playing an important role. The nonverbal nature of this knowledge might alternately be ascribed to the type of representation, i.e., "motor learning" might not support verbally accessible representations. However, the memory systems framework incorporates this idea by including learning within specific neural systems such as motor execution (or perceptual learning) as varieties of implicit learning in that they do not depend on the MTL memory system and produce knowledge representations that cannot be described.

Many of the tasks examined in the general domain of skill learning are not simple motor or perceptual learning tasks. Cognitively complex skill such as playing chess are initially learned through explicit instruction but expertise only emerges after extensive practice (Ericsson et al., 1993). Within music cognition, the different roles of explicit memorization and learning from practice are well-understood. Chaffin, Logan and Begosh (2009) describe in detail two parallel processes of preparing for expert music performance, one based on building associative chains while repeatedly practicing and a separate process of explicitly memorizing the written score (as a backup in case of error). The memory systems framework provides a useful way of characterizing these learning processes. Memorization of the music piece depends on explicit memory and the MTL memory system. Practicing the piece allows for the pervasive neuroplasticity mechanisms supporting implicit learning to hone neural processing to make execution of performance smooth, precise, and accurate.

However, the fact that this framework is consistent with descriptions of skill learning does not establish that the account is accurate. One of the challenges in studies of complex skill learning is the necessity of both explicit and implicit instruction during the learning process. Because these always co-occur, alternate hypotheses about skilled knowledge representations need to be considered. One possibility is that skill learning produces a functionally integrated representation across memory types such that an independent systems model cannot aid our understanding of this process. Another possibility is that repeated practice changes the character of an initially explicit memory representation such that retrieval becomes so rapid, effortless, and automatic that there is no role (or need) for implicit learning processes. Laboratory research aimed to capture the skill-learning process in order to address these alternatives has largely focused on tasks of perceptual-motor skill learning such as the SRT task.

The SRT task (Nissen and Bullemer, 1987) is a highly studied task that appears to be largely supported by implicit learning. Participants perform a serial 4-alternate forced choice response task in which the sequence of cues covertly follows an embedded sequence. Faster reaction times when the cues follow a practiced sequence compared with conditions where the cues follow an unfamiliar sequence are evidence that the practiced sequence has been learned. Establishing that this learning is solely implicit would provide robust evidence for the memory systems framework in skill learning. This kind of direct implicit learning without initial explicit memorization makes it clear that skilled performance does not necessarily depend on either an integrated implicit/explicit knowledge representation or automation of initially explicit knowledge.

However, while the first demonstration of learning with the SRT task suggested both knowledge outside of awareness and intact learning by memory-impaired patients (indicating lack of dependence on the MTL memory system), debates about the character of knowledge acquired during the SRT task have persisted. Thorough investigations of the conscious access of sequence knowledge (Perruchet and Amorim, 1992; Willingham, Greely, and Bardone, 1993) suggested that sensitive tests of sequence recognition almost always indicate some explicit knowledge in healthy participants. Having some conscious memory of the repeating sequence might reflect the concomitant operation of the MTL memory system (in cognitively healthy undergraduate participants) or might reflect evidence for the alternate hypothesis that skill learning depends on integrated representations. Studies of amnesic patients (Reber and Squire, 1994; 1998) showed that reliable learning could be observed in the absence of explicit memory, but concerns remained about the ability to prove intact learning in patients (which necessarily depends on a finding of a null difference between patients and controls). Destrebecqz and Cleeremans (2001) reported a strong dissociation between implicit and explicit sequence knowledge for a specific variant on the SRT design (zero delay in the interval between response and next cue). Overall, while the evidence supported the idea that learning was implicit, the difficulty of regularly finding evidence for process-pure implicit learning with the SRT task meant questions about representation persisted.

A new variation of the sequence learning paradigm was described by Sanchez, Gobel, and Reber (2010) as a Serial Interception Sequence Learning (SISL) task. This paradigm changed the basic task performed by the participant. Rather than a simple speeded response to the onset of a cue (in one of four locations), cues appear, then move vertically down the screen towards a target area, and the participant has to time an "interception" response of pressing the correct response key precisely as the cue reaches the target area. Just like in the SRT task, the cues follow a repeating covertly embedded sequence but the additional cognitive demands of the response task appears to reduce the degree to which explicit knowledge is acquired. Sequence knowledge is measured by accuracy (a properly timed response is correct, a mis-timed or incorrect keypress is incorrect) during the repeating sequence compared with accuracy during an unfamiliar sequence. Sanchez, Gobel, and Reber (2010) showed that learning on the SISL task is solely implicit for a substantial subset of cognitively healthy participants in a typical experiment. Showing robust learning with zero apparent explicit knowledge in approximately a third of participants provided strong evidence that the SISL task could be learned in the absence of explicit knowledge, arguing against integrated representations. In a follow-up study, Sanchez and Reber (2013) found that giving participants full explicit knowledge of the embedded sequence did not affect performance on the core task, providing strong evidence that implicit learning drives performance and that any explicit knowledge obtained by noticing the repeating cues does not materially contribute to accurate responding in the SISL task (in contrast to the SRT task where explicit knowledge can lead to negative reaction times where participants respond before cue onset). Neuroimaging during the SISL task found that learning was associated with greater efficiency in neural processing for the practiced, repeated sequence as would be predicted by adaptive neuroplasticity (Gobel, Parrish and Reber, 2011). Neural changes were largely in cortical regions, although increased activity suggested a role for the ventral striatum. A neuropsychological study of memory-impaired (amnestic MCI) patients and patients with Parkinson's disease (PD) found impaired learning in the PD patients but intact learning in the MCI patients, reinforcing the importance of the basal ganglia rather than the MTL for sequence learning (Gobel et al., 2013).

Across each of these studies, knowledge of the embedded repeating sequence was found to be extracted implicitly from practice and used to enhance task performance (when the cues follow that sequence). Because this happens without initial explicit cue knowledge and independently of the MTL, the memory systems framework provides the best account of the learning process as based on separate neural systems for implicit and explicit sequence knowledge. Theoretical accounts based on skilled performance emerging from integrated representations or explicit knowledge automated through practice cannot account for these findings. Beyond a consistent description, the memory systems framework also guided a series of additional studies seeking to better characterize the implicit learning component of skill learning with a goal of understanding skill training and education outside the laboratory.

A key challenge in skill learning is the degree to which learning is inflexible, leading to poor performance in novel but related transfer tasks (Adams, 1987) which may be due to the role of relatively inflexible implicit learning (Cleeremans, Destrebecqz, and Boyer, 1998). Using the SISL task, sequence learning was found to be highly specific and inflexible such that small changes in inter-cue timing (Gobel, Sanchez, and Reber, 2011) or perceptual characteristics (Sanchez, Yarnik, and Reber, 2015) led to nearly complete elimination of the accuracy advantage for practiced sequences. This inflexibility may have the practical consequence of making implicit knowledge occasionally inaccessible, perhaps explaining the need for expert musicians to separately memorize the written score prior to performance (so that explicit memory could be used to rescue performance if implicit knowledge was unexpectedly unavailable).

In contrast to this constraining aspect of implicit learning, Sanchez and Reber (2012) found robust implicit learning for surprisingly long repeating sequences in the SISL task (up to 90 items) and that learning appeared to be log-linear with practice regardless of sequence length, meaning long sequences were learned as rapidly as short ones (except that they took longer to complete). The ability to learn very long sequences indicates that implicit learning can support the kind of memory described by musicians during the "learning" phase of repetitive practice in which performance is honed for a piece that will contain large numbers of sequential actions. An extension of this kind of long sequence learning into an applied context was described by Bojinov et al. (2012) in which participants implicitly learned a long sequence that was then used as part of security authentication as an implicit password. A password learned this way has useful security implications as it cannot be shared (or coerced) and reflects the use of implicit learning and memory systems theory in an attempt to guide non-laboratory applications.

This theoretical framework has also been applied to research examining the effect of stress (pressure) on the performance of trained skills (DeCaro et al., 2011). Beilock and colleagues described a theory of "choking" under pressure in which explicit monitoring of a skill learned implicitly led to decrements in expert performance (Beilock and Carr, 2001). Flegal and Anderson (2008) reported a similar phenomenon in skilled performance in experts as verbal overshadowing reflecting competition between memory systems. These types of findings are difficult to understand without utilizing the memory systems framework that incorporates different types of memory with different operating characteristics and separate neural mechanisms. Being able to study each system independently has also recently revealed that implicit learning and/or performance can be influenced by factors such as mental fatigue (ego depletion; Thompson et al., 2014), motivation (Chon et al., 2018) or even hypnosis (Németh et al., 2013). As skill learning is foundational to education (cognitive skills), training and the development of expertise, implicit learning and the memory systems framework will provide critical guidance to basic science research applied to improving learning in skill learning contexts.

Decision making

A research area in which the roles and interactions among multiple systems has a fairly substantial history is the process of decision making. In his remarks upon accepting the Nobel Prize in economics, Kahneman (2003) described the framework developed by his work with Amos Tversky as emerging from two cognitive systems. Intuitive reasoning depends on System 1, a processing system characterized as: rapid, automatic, effortless, associative, slow-learning, and emotional. In contrast, System 2 reasoning is deliberate, slow, controlled, effortful, rule-governed, and flexible. These system definitions mirror the memory systems model of implicit and explicit learning with many of the same descriptive terms applied to features of each type of processing. However, this line of research was largely developed independently and was primarily applied to research on behavioral economics and decision making without direct connection to the role of memory.

Within this line of research, a notable difference is the focus on the speed of processing rather than the availability of knowledge to conscious awareness or underlying neural systems. Descriptions of decision making within this framework typically describe a fast System 1 response that can be then reviewed and potentially overridden by a slower System 2 response. This type of interaction across systems is different than those considered within skilled expertise (or language processing) but might be hypothesized to play a role in those domains as well.

This approach lends itself to research examining the phenomenon of intuition (e.g., Klein, 2004), defined as a System 1 process that rapidly identifies an action to take that is often subjectively described as based on a "gut hunch" or instinct. This type of intuitive decision making (IDM) has been studied for its potential to support rapid, expert, accurate decisions that are of great value in complex and/or stressful environments. That approach is somewhat different than the early focus of Tversky and Kahneman (1974) that focused on erroneous (non-rational) decisions driven by biases that could emerge from System 1 processes. Kahneman and Klein (2009) contrasted and compared a System 1 and 2 account of this process with research obtained through naturalistic decision-making research based on analysis of experts making complex, high leverage decisions in the field. They determined that their approaches were largely in sync and suggested that prior experience in the decision-making context was an effective predictor of the accuracy of intuition. This idea was explored and directly supported empirically by Dane, Rockmann, and Pratt (2012) by comparing the accuracy of intuition across different levels of domain expertise. High domain expertise led to much more accurate intuitive judgments, as would be expected if intuition was supported by implicit extraction of the statistics of the environment during the acquisition of domain knowledge.

This conclusion fits well with the memory systems model derived from laboratory studies of implicit and explicit memory. Implicit knowledge of a specific domain is accumulated as part of the development of expertise based on refining and honing processing (as in skill learning) in addition to statistical learning of environmental features (as in language learning). The resulting implicit knowledge structures reside outside awareness due to their dependence on plasticity separate from the MTL memory system (which will provide episodic memory of specific examples and salient events from experience). We can also connect this idea to laboratory studies of implicit learning where participants are asked to make a response, e.g., about grammaticality of an unfamiliar letter string, but report they feel as though they are just guessing even when their performance is significantly above chance (Reber, Beeman, and Paller, 2013).

However, an unanswered question in memory research is the route by which this information proceeds through the brain in order to actually guide action selection. Using laboratory studies of visual category learning in which participants are required to learn categories through a process of trial and error as a model, Nomura and Reber (2012) described a multi-system model of category learning and performance, PINNACLE, that incorporated two separate processing streams for information extracted implicitly during learning and memorized knowledge of the category stimuli. This model has the structure of a "mixture of experts" model at the decision-making level with the response decision (the participant's response about which category the stimulus was thought to belong to) being influenced by either implicit knowledge via intuition or explicit knowledge by deliberate application of conscious task knowledge.

The PINNACLE model was developed in reference to a well-established laboratory paradigm for studying category learning (Ashby and Alfonso-Reese, 1998; Ashby and Maddox 2005) in which known manipulations to the underlying category structure could lead participants to rely on an explicit, rule-based (RB) strategy or an implicit strategy based on integration information across dimensions (II). In this task, participants are shown artificial stimuli that vary in two dimensions, such as sine-wave gratings that vary in spatial frequency (line thickness) and tilt. They attempt to learn how the stimuli are organized into two underlying categories by trial and error with feedback after each response. When the structure is determined by a simple rule, participants generally discover the rule, use it to make their category membership decisions and verbally report the rule after learning. Complex, multi-dimensional rules often drive behavior differently with participants exhibiting gradually increasing accuracy at the task but without being able to report the basis of their judgments.

Using neuroimaging, Nomura et al. (2007) showed that neural activity associated with RB learning occurred within the MTL memory system. In contrast, II learning was associated with increased activity in posterior regions of the caudate, brain areas often associated with implicit learning plasticity (Seger and Miller, 2010). Using the PINNACLE model to probe the neuroimaging data in more detail, regions in the prefrontal cortex were identified that were associated with the cognitive process of selecting which strategy to apply on a single decision. The resulting model lines up well structurally with the Kahneman (2003) framework with separate neural systems contributing to rapid, intuitive decisions and slower, deliberate, and explicit decisions. Interactions between the two modes of decision making would occur within the dorsolateral prefrontal cortex which would reflect a meta-level decision such as knowing when to "trust one's instinct" to guide behavior.

The PINNACLE model provides a method for translating the laboratory studies of multiple brain systems into non-laboratory applications. P. Squire et al. (2014) described how research in this direction could be used to study the processes of IDM and generate hypotheses about how decision-making expertise could be trained more rapidly. A similar approach was used by Dane and Pratt (2007) in their analysis of how treating intuitive and non-IDM in managerial contexts could be informed by the multiple memory systems model. In a number of these cases, attention is also paid to erroneous decision making that can emerge from reliance on intuition (Kahneman and Klein, 2009). A balanced model of the value of IDM emerges naturally from an implicit learning approach. Implicit learning can only reflect experience and the statistical structure of the environment in which it was acquired. Thus intuitions may be quite inaccurate in novel contexts where the environmental statistics are different than prior experience. In addition, implicit learning through practice can just as easily reinforce consistently erroneous decisions, i.e., bad habits can be learned as easily as expert performance.

A research area focused directly on the potential negative consequences of our automatic implicit learning is studies of stereotype prejudice that are based on "implicit attitudes" (Greenwald et al., 2002). The core idea in these studies is that experience in an environment shaped by the existence of stereotypes will tend to shape individuals' cognitive processes to reflect these prejudices. The result of this process is that stereotypical information is represented outside awareness, leading individuals to not even realize that their responses and decisions are being influenced by this implicitly acquired bias. This implies a very different model of prejudice in which stereotype-driven decisions and responses are not knowingly based on dislike of an outgroup but are based on something closer to a negative form of intuition. This model fits very well with the initial descriptions of decision-making biases originally characterized by Tversky and Kahneman (1975) in accounts of apparently non-rational decision-making behavior. Thus, statistically-induced biases in cognition that are acquired via implicit learning can enhance decision-making performance, but there is also a potential negative side where environmental bias will become reinforced through the same mechanism.

Conclusions

While neuroscientific studies of memory leave little doubt that there are multiple mechanisms of synaptic change in different systems across the brain, this observation does not indicate what roles different types of memory play in complex human behavior. Phenomena characterized as implicit learning in laboratory studies have shown how prior experience can influence current behavior without awareness of the information previously acquired. However, the cognitive consequences of this type of memory are most clearly seen when examining applications of memory systems theory outside the laboratory. Research examining human decision making, skill acquisition and language learning have all converged on theoretical frameworks that are highly consistent with the basic multiple memory systems model derived from cognitive neuroscience research. Whether these separate processes are called System 1 and 2, instruction and practice, or syntax and semantics, independent roles are seen for learning from both the statistics of experience and also conscious memorization of prior episodes. Thus applied learning and memory research is well captured by the memory systems model of P. J. Reber (2013) which posits widespread non-MTL neuroplasticity as the basis for implicit learning as a separate type of memory than that supported by the MTL.

Applying the memory systems framework to questions of learning in non-laboratory contexts highlights some gaps in many current programs of memory research. In language use, skill learning and decision making, identifying important roles for both types of memory immediately indicates a need for hypotheses about how these systems interact. Since the main focus in memory systems research to date has been focused on isolating memory types, most laboratory paradigms have not confronted questions about the interplay among systems. In contrast, in decision-making research the potential for slower, deliberate processing to override a fast, intuitive response is a core hypothesis. In addition, basic questions about how we learn to trust and use our intuition are not addressed within the memory systems model. Trusting one's gut instinct appears to imply a meta-cognitive process for evaluating the quality of our implicitly held knowledge, which is a counter-intuitive construct since the implicit knowledge is theoretically outside of awareness. Within skilled performance, the example of expert musicians both learning and memorizing a piece to be performed indicates a different kind of interaction. Here, the implicit, practice-based knowledge is not seen as inaccurate but occasionally and unpredictably unavailable, requiring redundant memory representations to support performance. Skilled performance can also reveal negative interactions between memory systems in a model of "choking" where explicit processing interferes with expert implicit processing (Beilock and Carr, 2001). Within the domain of language, the role of extracted statistics from prior experience seems as if it must function in a much more closely synergistic manner with conscious aspects of linguistic processes in order to communicate, a fundamentally conscious process.

Laboratory studies of phenomena related to implicit learning over the past 50 years have established tools and paradigms for characterizing and studying these phenomena. Applied, non-laboratory research instead leapt ahead, assuming a basic implicit/explicit multi-system model and found it provided explanatory power in a range of domains. A common framework unifies these approaches and builds on the understanding of the cognitive neuroscience of memory mechanisms in the brain. Widespread neuroplasticity leads to adaptive rewiring of neural circuitry to improve performance and increase neural efficiency. This mechanism leads to knowledge embedded in performance structures that is implicit and unavailable to conscious report. The MTL memory system supports acquisition and consolidation of episodic memory, prior experiences of facts and events, that are retrieved consciously and used flexibly and creatively. Both research approaches then point to the need for theoretical development at the interaction between systems to understand how information represented in such different ways can support complex human cognition.

Acknowledgements

Preparation of this chapter was supported by grants from the Office of Naval Research (N00014-16-1-2251; PJR) and National Science Foundation (CNS-1228357; PJR) as well as NINDS and NICHD (T32 NS047987 and F32 HD078223; LJB). The authors would also like to thank Chaya Tabas for her great help in manuscript preparation.

References

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin*, *101*(1), 41.

Anderson, J. R. and Milson, R. (1989). Human memory: an adaptive perspective. *Psychological Review*, *96*(4), 703.

Ashby, F. G. and Alfonso-Reese, L. A. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442.

Ashby, F. G. and Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.*, *56*, 149–178.

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Batterink, L. J., Reber, P. J., Neville, H. J., and Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78.

Beilock, S. L. and Carr, T. H. (2001). On the fragility of skilled performance: what governs choking under pressure? *Journal of Experimental Psychology: General*, *130*(4), 701–725.

Bojinov, H., Sanchez, D.J., Boneh, D., Reber, P.J., and Lincoln, P. (2012). Neuroscience meets cryptography: designing crypto primitives secure against rubber hose attacks. Usenix Annual Technical Conference, Boston, MA.

Broadbent, D. (1958). *Perception and Communication*. London: Pergamon Press.

Campitelli, G. and Gobet, F. (2011). Deliberate practice: necessary but not sufficient. *Current Directions in Psychological Science*, *20*(5), 280–285.

Chaffin, R., Logan, T. R., and Begosh, K. T. (2009). Performing from memory. In: S. Hallam, I. Cross and M. Thaut (eds.), *The Oxford Handbook of Music Psychology* (pp. 352–363). Oxford: Oxford University Press.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior: Language*, *35*(1), 26–58.

Chon, D., Thompson, K. R., and Reber, P. J. (2018). Motivation to avoid loss improves implicit skill performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 327–333.

Chun, M. M. and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71.

Chun, M. M. and Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, *2*(9).

Cleeremans, A., Destrebecqz, A., and Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416.

Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, *6*(3), 255–265.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill*. *Ergonomics*, *2*(2), 153–166.

Dane, E. and Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. *Academy of Management Review*, *32*(1), 33–54.

Dane, E., Rockmann, K. W., and Pratt, M. G. (2012). When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes*, *119*(2), 187–194.

DeCaro, M. S., Thomas, R. D., Albert, N. B., and Beilock, S. L. (2011). Choking under pressure: multiple routes to skill failure. *Journal of Experimental Psychology: General*, *140*(3), 390–406.

Destrebecqz, A. and Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*(2), 343–350.

Dienes, Z. and Berry, D. (1997). Implicit learning: below the subjective threshold. *Psychonomic Bulletin & Review*, *4*(1), 3–23.

Dulany, D. E., Carlson, R. A., and Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, *113*(4), 541.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363.

Fiser, J. and Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Fitts, P. M. (1964). Perceptual-motor skill learning. *Categories of Human Learning*, *47*, 381–391.

Flegal, K. E. and Anderson, M. C. (2008). Overthinking skilled motor performance: or why those who teach can't do. *Psychonomic Bulletin & Review*, *15*, 927–932.

Gobel, E. W., Blomeke, K., Zadikoff, C., Simuni, T., Weintraub, S., and Reber, P. J. (2013). Implicit perceptual-motor skill learning in mild cognitive impairment and Parkinson's disease. *Neuropsychology*, *27*(3), 314.

Gobel, E. W., Parrish, T. B., and Reber, P. J. (2011). Neural correlates of skill acquisition: decreased cortical activity during a serial interception sequence learning task. *NeuroImage*, *58*(4), 1150–1157.

Gobel, E. W., Sanchez, D. J., and Reber, P. J. (2011). Integration of temporal and ordinal information during serial interception sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 994.

Graf, P. and Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 501.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., and Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3.

Gross, C. G. (1999). *Brain, Vision, Memory: Tales in the History of Neuroscience*. Cambridge, MA: MIT Press.

Hamann, S. B. and Squire, L. R. (1997). Intact perceptual memory in the absence of conscious memory. *Behavioral Neuroscience*, *111*(4), 850.

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *The American Economic Review*, *93*(5), 1449–1475.

Kahneman, D. and Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American Psychologist*, *64*(6), 515.

Kandel, E. R. and Spencer, W. A. (1968). Cellular neurophysiological approaches in the study of learning. *Physiological Reviews*, *48*(1), 65–134

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Klein, G. A. (2004). *The Power of Intuition: How to Use Your Gut Feelings to Make Better Decisions at Work*. New York: Crown Business.

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

Knowlton, B. J. and 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*(1), 169.

Lashley, K. S. (1929). *Brain Mechanisms and Intelligence: a Quantitative Study of Injuries to the Brain*, Chicago: University of Chicago Press.

Lechner, H. A., Squire, L. R., and Byrne, J. H. (1999). 100 years of consolidation—remembering Müller and Pilzecker. *Learning & Memory*, *6*(2), 77–87.

Luce, R. D. (Ed.) (1963). *Handbook of Mathematical Psychology*, New York; London: John Wiley & Sons Inc.

Marcus, G. F., Vijayan, S., Rao, S. B., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80.

McGaugh, J. L. (2000). Memory—a century of consolidation. *Science*, *287*(5451), 248–251.

Merikle, P. M. (1994). On the futility of attempting to demonstrate null awareness. *Behavioral and Brain Sciences*, *17*(3), 412.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Milner, B., Corkin, S., and Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM. *Neuropsychologia*, *6*(3), 215–234.

Milner, B., Squire, L. R., and Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron*, *20*(3), 445–468.

Morgan-Short, K., Sanz, C., Steinhauer, K., and Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154–193.

Németh, D., Janacsek, K., Polner, B., and Kovacs, Z. A. (2013). Boosting human learning by hypnosis. *Cerebral Cortex*, *23*(4), 801–805.

Newell, A., Shaw, J. C., and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*(3), 151.

Newell, A. and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition*, *1*, 1–55.

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance I: statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.

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

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., . . . and Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*(1), 37–43.

Nomura, E. M. and Reber, P. J. (2012). Combining computational modeling and neuro-imaging to examine multiple category learning systems in the brain. *Brain Sciences*, *2*(2), 176–202.

Nosofsky, R. M. and Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. *Psychological Science*, *9*(4), 247–255.

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism* (Vol. 18). Amsterdam; Philadelphia: John Benjamins Publishing.

Perruchet, P. and 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*(4), 785–800.

Perruchet, P. and Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*(3), 264.

Perruchet, P. and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., and Shohamy, D. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546.

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

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219.

Reber, P. J. (2013). The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*, 2026–2042.

Reber, P. J., Beeman, M., and Paller, K. A. (2013, July). Human memory systems: a framework for understanding the neurocognitive foundations of intuition. In *International Conference on Augmented Cognition* (pp. 474–483). Berlin, Heidelberg: Springer.

Reber, P. J. and Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & Memory*, *1*(4), 217–229.

Reber, P. J. and Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, *10*(2), 248–263.

Reed, J. and 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*(3), 585.

Roediger, H. L. (1990). Implicit memory: retention without remembering. *American Psychologist*, *45*(9), 1043.

Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sanchez, D. J., Gobel, E. W., and Reber, P. J. (2010). Performing the unexplainable: implicit task performance reveals individually reliable sequence learning without explicit knowledge. *Psychonomic Bulletin & Review*, *17*(6), 790–796.

Sanchez, D. J. and Reber, P. J. (2012). Operating characteristics of the implicit learning system supporting serial interception sequence learning. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 439.

Sanchez, D. J. and Reber, P. J. (2013). Explicit pre-training instruction does not improve implicit perceptual-motor sequence learning. *Cognition*, *126*(3), 341–351.

Sanchez, D. J., Yarnik, E. N., and Reber, P. J. (2015). Quantifying transfer after perceptual-motor sequence learning: how inflexible is implicit learning? *Psychological Research*, *79*(2), 327–343.

Schacter, D. L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 501–518.

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11.

Seger, C. A. and Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203–219.

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

Shimamura, A. P. and Squire, L. R. (1989). Impaired priming of new associations in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 721.

Skinner, B. F. (1957). *Verbal Behavior*. Englewood Cliffs, NJ: Prentice Hall.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195.

Squire, P., Cohn, J., Nicholson, D., Nolan M., Reber, P. J., Oudiette, D, Niehaus, J., Geyer, A., and O'Neill, L. (2014). Towards enhancing intuitive decision making through implicit learning. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2014.

Stark, C. E. and Squire, L. R. (2000). Recognition memory and familiarity judgments in severe amnesia: no evidence for a contribution of repetition priming. *Behavioral Neuroscience*, *114*(3), 459.

Thompson, K. R., Sanchez, D. J., Wesley, A. H., and Reber, P. J. (2014). Ego depletion impairs implicit learning. *PLoS One*, *9*(10), e109370.

Turk-Browne, N. B., Jungé, J. A., and Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, *92*(1), 231–270.

Weiskrantz, L. and Warrington, E. K. (1970). Verbal learning and retention by amnesic patients using partial information. *Psychonomic Science*, *20*(4), 210–211.

Willingham, D. B., Greeley, T., and 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.

2

THE MIND IS DEEP

Axel Cleeremans

The central question that drives research on unconscious cognition is rather straightforward: can what we do be influenced by unconscious knowledge? As simple as it sounds, this question remains largely unanswered today, mostly because of the considerable methodological challenges involved in collecting accurate data about what people actually know at some point in time and of the epistemological challenge of demonstrating absence of knowledge. And yet, we all share the compelling intuition that it *must be the case* that what we do is influenced by factors we are not aware of, simply because we seem to experience this all the time. For instance, we are not continuously paying attention to whatever it is that we do, but attention is necessary for consciousness. Breathing is perhaps the best example, and so are countless other bodily actions such as posture, reflexes, aim, eye movements, and so on. While such examples are not controversial in and of themselves, critics will be quick to point out that are barely *cognitive*. The kind of knowledge involved in driving such habitual behavior is perhaps best described as neural knowledge—it does not involve the sorts of representations that we think of as *mental representations*, that is, words and images, which have an existence in our minds because we can readily experience them as mental objects. It is these sorts of representations that the study of unconscious cognition is concerned with, and in such cases, it is much less clear whether it can be the case that there exist unconscious mental representations capable of causally influencing behavior in the absence of awareness. I will not overview the many continuing controversies that pervade research on unconscious cognition here—from unconscious decision-making to subliminal perception—but simply note, despite general sentiment to the contrary, that the question is overall far from settled (e.g., Balsdon and Clifford, 2018).

Beyond the genuinely hard methodological and epistemological challenges associated with demonstrating unconscious cognition, the most significant challenge is a conceptual one. It is still not clear, in my mind, and, I would surmise, in anyone else's mind, how exactly we should think of the difference between conscious and unconscious cognition. Some authors have argued, and one finds echoes of this in Allakhverdov's chapter (this volume), that there are reasons to consider that unconscious and conscious cognition involve radically different systems. Freud (1949), but also Kahneman (2011), have in that sense adopted an almost architectural approach to the mind, with neatly distinct, yet interconnected spaces for information and affective processing. Others, most notably Searle (1992), have argued for a strong connection between conscious and unconscious processing. There is no sense, says Searle, in which we can envision unconscious knowledge if it isn't by assuming that it *could* be conscious. Some further authors have argued just as forcefully for the perspective that unconscious cognition simply does not exist—one thinks here of authors such as David Shanks (Shanks and St. John, 1994) or Pierre Perruchet (Perruchet and Vinter, 2002a), both staunch critics of the field of implicit learning and of unconscious cognition in general.

Joining the ranks of the skeptics, Nick Chater also argues forcefully against the existence of the unconscious in his book titled *The Mind is Flat* (Chater, 2018). There are no "hidden depths" in our minds, says Chater. The mind is not structured like the proverbial iceberg, with its bulk sitting unseen under the surface of the ocean. Rather, Chater argues, there simply is no iceberg; instead, there is only what one can see above the water—a chunk of ice floating on the surface. Hence, "The mind is flat—the surface is all there is" (p. 220). Everything mental takes place in the moment; each of our thoughts is an improvised creation; we are continuously reinventing ourselves as we go.

In defending this argument, Chater appeals to the observation that much of what we experience is in fact illusory. We see things that are not there—motion where there is none, as in contrast-induced motion illusions; color in the nevertheless cone-sparse periphery of our visual field; animal shapes in clouds. We see differences where there are none, as in the stupefying Adelson illusion, which makes two identically colored gray squares appear to be of strikingly different shades. But we also fail to see things that are there, as in change blindness or inattentional blindness demonstrations. Likewise, memory is also subject to considerable malleability. Not only do we often fail to retrieve information from memory, we can also be convinced that something that never actually happened did in fact take place (Loftus and Hoffman, 1989). We confabulate reasons for choices we did not make (Johansson, Hall, Sikström, and Olsson, 2005). In short, we are subject, says Chater, to a neural hoax of sorts. As Seth (2016) wrote, perhaps "We hallucinate all the time. It's just that when we agree about our hallucinations, that's what we call reality."

This all acts as a strong and justified reminder that perception, rather than being a faithful rendering of what is out there in the world, should rather be thought of as an active construction, whereby sensory information is continuously integrated with our own beliefs and expectations about what will come next—this is, in a nutshell, the core of the predictive processing framework (Clark, 2013; Friston, 2006; Hohwy, 2013). But Chater goes one step further. What if our views on the unconscious were likewise illusory? What if it turned out, in fact, that there was no submerged chunk of ice under the surface, that the depth we sense is nothing but another illusion? Consider insight, says Chater. You're thinking hard, over days, about a problem. One day the solution presents itself to you. Now, you can have two stories about what actually happened over the course of those few days. One story assumes that your unconscious mind worked continuously on the problem, precisely as you were sitting there hopelessly trying to solve it consciously. The other story is simply that there was no unconscious work going on at all—it's just that by rearranging the pieces of the puzzle consciously over and over again, at some point, you found an arrangement *that you then suddenly recognized as the solution*. Hence the feeling of "insight". I vividly experienced this myself. For years I drove my daughter Emilie to school, taking the same route everyday at the same time. On one such trip, Emilie and I were searching for the name of an actor in a movie we had both seen. Neither she nor I could even remember the name of the movie, and so the entire drive was spent in silence fruitlessly rummaging through our respective minds for any scrap of information that would help us towards the solution. I dropped her off, drove back thinking about this some more and then soon forgot about it. A few days later, on the way back from school again, the name of the actor (which I have now ironically forgotten again) popped to my mind for no apparent reason, and I instantly recognized that name as that of the actor we were looking for. "How wondrous!" I thought. It truly felt as if some inscrutable search process had finally come up with an answer to the query launched days before! But then, Chater's alternative account immediately came to my mind as well. Perhaps, I thought, there had not been any unconscious search going on all that time. Perhaps I simply thought of that name at that time for reasons that have nothing to with the fact that I had been looking for it so intensely a few days ago. Perhaps something on the radio made me think that name at that moment. Who knows? What made the experience interesting was the instant, conscious, puzzled recognition that that was the name I had been looking for.

Fair enough—I find the opposite view that the unconscious is more powerful than consciousness (e.g., Dijksterhuis and Nordgren, 2006) just as untenable. But is this view that the "mind is flat" defensible? Chater himself expresses doubt when he writes:

> Over a lifetime, the flow of thought shapes, and is shaped into complex patterns: our habits of mind, our mental repertoire. These past patterns of thought, and their traces in memory, underpin our remarkable mental abilities, shape how we behave and make each of us unique. So, in a sense, we do, after all, possess some inner mental landscape (p. 203).

Developing this line of thought, Chater adopts a hydraulic metaphor which I am fond of: each act of thought is like a river that progressively carves its way into the landscape, producing "mental channels". And in turn, that very landscape gives the flow of the river its own unique character. But that view is one that is in fact entirely consistent with the idea that what we do is influenced by contents that we remain unaware of, just as the flow of the river is constrained by the very channels it created through erosion.

What, then, is it that Chater is arguing against here? There is a distinct rendition of the unconscious that Chater aims to deconstruct, namely the Freudian proposal that there exist unconscious mental representations that are structured like conscious mental representations and that can be just as causally efficacious. Searle (1992) argued in roughly the same direction as Chater, that is, specifically, against Freudian ideas that the unconscious has structure and that it works in ways that are akin to the mechanisms involved in conscious processing, namely symbol processing. Not so, says Searle; the only way in which we can make sense of unconscious representations is by connecting them with conscious representations (this is what Searle calls the "connection principle"): "The notion of an unconscious mental state implies accessibility to consciousness. We have no notion of the unconscious except as that which is potentially conscious" (p. 152). Later on, Searle elaborates as follows:

> Our naïve, pretheoretical notion of an unconscious mental state is the idea of a conscious mental state minus the consciousness. But what exactly does that mean? How could one subtract the consciousness from a mental state and still have a *mental* state left over? (p. 152).

Likewise, but in a different direction that is perhaps closest to Chater's own perspective, Perruchet and Vinter (Perruchet and Vinter, 2002a, 2002b) have defended the idea that consciousness is "self-organizing"—that is, the idea that the conscious representations are, in general, sufficient to account for all of cognition: "Accordingly, the most salient feature of the mentalistic framework is the denial of the very notion of unconscious representations. The only representations that exist, in this view, are those that are embedded in the momentary phenomenal experience" (p. 299). The mental world, according to Perruchet and Vinter, thus divides squarely into two realms: neural information processing on the one hand, and conscious processing on the other.

In his article "Styles of mental representation", Dennett (1982) also attempted to clarify the subtle distinctions between explicit, implicit, and tacit representations. Explicit representation, says Dennett, is:

> a physically structured object, a formula or string or tokening of some members of a system (or "language") of elements for which there is a semantics or interpretation, and a provision (a mechanism of some sort) for reading or parsing the formula" (p. 216).

This definition is a bit long-winded, but two key elements stand out: (1) an explicit representation is a symbolic proposition that (2) requires interpretation, which in turn requires a mechanism to do so. Implicit representations, for Dennett, are those that are logically implied by something that is stored explicitly. Tacit representations, finally, are representations that capture the "knowhow" of a cognitive system:

> The knowhow has to be built into the system in some fashion that does not require it to be represented (explicitly) in the system. People often use the word "implicit" to describe such information-holding; what they mean is what I mean by "tacit" (p. 218).

This usage of "tacit" is wholly congruent with Reber's own notion of "tacit knowledge" (Reber, 1993).

Finally, in an influential article, Dienes and Perner (1999) also offered a careful analysis of the implicit–explicit distinction. According to Dienes and Perner:

> a fact is explicitly represented if there is an expression (mental or otherwise) whose meaning is just that fact; in other words, if there is an internal state whose function it is to indicate that fact. Supporting facts that are not explicitly represented but must hold for the explicitly known fact to be known are implicitly represented (p. 736).

Dienes and Perner distinguish two "sources of implicitness". First are linguistic presuppositions—when I say "The mind is flat", I necessarily presuppose that there exists something called a "mind", yet this is not explicitly conveyed. The second source of implicitness comes from semantics. Each fact I state explicitly necessarily implies other facts that remain implicit. Thus when I state that someone is a bachelor, to use Dienes and Perner's example, I also state, implicitly, that that person is a man and that he is unmarried. In Artificial Intelligence, the problem of handling such cases, which are ubiquitous, is known as the frame problem—the problem an artificial agent is faced with when it comes to updating its beliefs about the world when it acts. The frame problem is in turn connected to the common-sense problem, which also appeals to vast amounts of implicitly represented facts that support explicit reasoning, and which appears simply intractable when all the possible consequences and non-consequences of any explicit fact have themselves to be represented explicitly.

These different perspectives highlight how complex a problem it is to think of such distinctions in a conceptually coherent manner. In Cleeremans (2014), I offered my own analysis of this conceptual riddle and suggested that connectionism, *contra classical approaches*, offered a whole new way of rethinking the distinction between implicit and explicit knowledge. I remain convinced by these arguments, which I reproduce *in extenso* below for the lack of a better way of conveying it otherwise.

The trouble with classical approaches

In Cleeremans (1997) and also in Cleeremans and Jiménez (2002), I suggest that the central reason why we have such a hard time thinking of the unconscious is that the phenomena of implicit cognition cannot be reconciled with classical perspectives on information processing. While nobody truly defends such classical perspectives today, they nevertheless continue to shape our thinking about the unconscious and make it hard for us to come to terms with alternative frameworks.

Empirically, the central characteristic of unconscious processing is the observation that an agent's behaviour is influenced by knowledge of which it remains unaware. In Cleeremans (1997), I define implicit knowledge as follows:

> At a given time, knowledge is implicit when it can influence processing without possessing in and of itself the properties that would enable it to be an object of representation.

Thus, unconscious knowledge is knowledge that is causally efficacious yet unavailable to form the contents of conscious experience. Now, consider the manner in which knowledge is represented in classical models of cognition. Such models—roughly speaking, the "Computational Theory of Mind" (Fodor, 1975)—take it as a starting point that cognition consists of symbol manipulation. The flow of information processing in classical models goes roughly like this: there is a central processor that fetches or stores information in knowledge bases, and processes it. The processor interacts with the world through input/output systems. Knowledge (either "programs" or "data") is represented symbolically. Bates and Elman (1993) dubbed this perspective on cognition "The First Computer Metaphor of Cognition", and characterized it as follows:

> At its core, the serial digital computer is a machine that manipulates symbols. It takes individual symbols (or strings of symbols) as its input, applies a set of stored algorithms (a program) to that input, and produces more symbols (or strings of symbols) as its output. These steps are performed one at a time (albeit very quickly) by a central processor. Because of this serial constraint, problems to be solved by the First Computer Metaphor must be broken down into a hierarchical structure that permits the machine to reach solutions with maximum efficiency (e.g., moving down a decision tree until a particular subproblem is solved, and then back up again to the next step in the program) (p. 630).

In such systems, thus, knowledge always takes the form of symbolic propositions stored in a mental database (i.e., production rules or declarative statements). There are two important and problematic features about such representations. First, they remain causally inert until activated or otherwise accessed by the processor. Second, their shape (symbolic propositions) makes their contents immediately accessible by the processor. The conjunction of these two features renders the entire approach incapable of accounting for unconscious cognition, for it entails that representations cannot influence processing independently of being accessed (activated, manipulated) by the processor. However, this property—causal influence without access—is precisely what one means when one says that knowledge is unconscious. A production rule, for instance, cannot influence ongoing processing unless the algorithm that drives the entire system has established that the rule's preconditions match current input and that the rule could now be applied. Intuitively, this is akin to a human participant who figures out, when confronted with a mathematical problem for instance, that an arithmetical expression can be simplified in certain ways described by a rule that he has learned. But this process is clearly a conscious process. Now consider what happens when a chess expert intuitively decides to move a particular chess piece. One could claim that the same process described above in the case of the arithmetic problem now takes place: a heuristic rule is identified as being relevant to the particular situation at hand and applied. However, the chess expert claims that he is unable to justify his choice: the move he made is just what came to mind. Perhaps he could explain the specific reasons why he chose that particular move given sufficient time and effort, but the move itself simply appeared to pop in his mind. The difference between the arithmetical problem and the chess move is one of consciousness: one seems to have access to the relevant knowledge in the first instance, but not in the second.

Now here is the key argument: if one assumes, as do thoroughly classical approaches to cognition, that the mechanisms involved in each case always entail accessing and activating the relevant rule, then one is left with no principled difference between cognition with and without awareness, for in both cases, the very same mechanisms (specifically: access to the relevant knowledge) are involved.

More formally, the argument could be spelled out in this way:

(1) awareness of some knowledge entails access to the relevant representations;
(2) in classical models, representations take the form of symbolic propositions;
(3) symbolic propositions cannot be causally efficacious unless they are accessed.

Therefore, in classical models, causally efficacious representations are necessarily conscious.

Briefly put thus, the argument I introduced in Cleeremans (1997) is this: if you believe that cognition consists exclusively of manipulating structured, symbolic, propositional representations, then you only have two possibilities of accounting for the phenomena of implicit cognition. You can either (1) ascribe them to a separate "psychological unconscious" (Kihlstrom, 1987) that is capable of performing exactly the same sorts of computations as your conscious system is (specifically: access to the relevant knowledge), only minus consciousness (Searle, 1992), or (2) explain them away by rejecting existing evidence for implicit cognition altogether and claim that all of cognition involves conscious knowledge (Chater, 2018; Perruchet and Vinter, 2002a; Shanks and St. John, 1994).

There is also a third possibility, which consists of rejecting the idea that unconscious cognition always involves symbol manipulation. This is the "third way" that connectionism (McClelland and Rumelhart, 1986; Rumelhart and McClelland, 1986b) and other subsymbolic approaches to cognition have made so salient over the past thirty years. While the classical perspective takes it as a starting point that information processing involves operations modeled after conscious cognition, connectionism turns this perspective on its head and proposes that information processing begins with unconscious cognition. It is worth pointing out here that many contemporary approaches rooted in symbolic processing (e.g., ACT-R, CLARION) have evolved to the point that they share many features more typically associated with connectionist models, such as associative processing.

Once we eliminate the idea that all of cognition, be it with or without consciousness, involves symbol manipulation, we can then focus on exploring what we can do without symbols. We are then facing the great challenge of figuring out how we can get symbols in the game after all, but at least we begin with more plausible assumptions. In the next section, I briefly overview how such approaches, and connectionism in particular, have changed our understanding of unconscious cognition.

What have we learned from connectionism?

Connectionist models have provided genuine insights into how knowledge can influence processing without access—a hallmark of unconscious processing—and of how change can accrue as a result of mere information processing—a hallmark of the phenomena of implicit learning. Numerous models of implicit learning based on connectionist models have now been proposed (Cleeremans and Dienes, 2008), and it is fair to say that such models have been very successful in accounting for the mechanisms that subtend performance in a wide range of relevant empirical paradigms, from artificial grammar learning (Dienes, 1992) and sequence learning (Cleeremans and McClelland, 1991) to process control (Gibson, Fichman, and Plaut, 1997) or priming (Mathis and Mozer, 1996).

The first fully implemented connectionist models of implicit learning are found in the early efforts of Dienes (1992) and of Cleeremans and McClelland (1991). While authors such as Brooks (1978) and Berry and Broadbent (1984) had already suggested that performance in implicit learning tasks such as Artificial Grammar Learning or Process Control may be based on retrieving exemplar information stored in memory arrays, such models have in general been more concerned with accounting for performance at retrieval rather than on accounting for learning itself. The connectionist approach, by contrast, has been centrally concerned with the mechanisms involved during learning since its inception, and therefore constitutes an excellent candidate framework with which to think about the processes involved in implicit learning.

My purpose here is not to review these developments in detail, but rather to focus on how four fundamental principles that characterize the connectionist approach (but also have much wider implications) are relevant to our understanding of the differences between conscious and unconscious processing. In the following, I discuss each in turn.

Active representation

As discussed above, this first principle highlights a fundamental difference between classical and connectionist representations, namely that the former are inherently passive whereas the latter are continuously active. Indeed, the symbolic, propositional representations characteristic of classical models of cognition (i.e., production rules and declarative knowledge) are intrinsically passive: they are objects (data structures) stored in mental databases and can only influence ongoing processing when an algorithm (i.e., an inference engine) has determined that certain trigger conditions are met. Thus, for a classical representation to be causally efficacious, it first needs to be accessed or otherwise made active in some way. But, as discussed above, this necessary link between causal efficacy and access is immediately problematic for our conceptualization of the differences between information processing with and without awareness. The difficulty stems from the (tacitly) assumed equivalence between causal efficacy, access, and consciousness. This equivalence in turn stems from the fact that in classical perspectives on cognition, there is a complete separation between representation and processing. Connectionism solves this quandary very elegantly by proposing that access is not necessary to drive information processing. Nothing "accesses" anything in a connectionist network. Instead, connectionist models assume that all the long-term knowledge accrued over experience is embedded in the very same structures that support information processing, that is, the connection weights between processing units. Such knowledge therefore does not need to be accessed in any way to be causally efficacious; it simply exerts its influence automatically whenever the units whose activation propagates through the relevant connections are active. Thus, knowledge in connectionist networks is active in and of itself, and fundamental phenomena such as priming are accounted for naturally without the need to postulate additional mechanisms.

An important consequence of the fact that long-term knowledge in connectionist networks accrues in connection weights as a mandatory consequence of information processing is that connectionist models capture, without any further assumptions, two of the most important characteristics of implicit learning, namely (1) the fact that learning is incidental and mandatory, and (2) the fact that the resulting knowledge is difficult to express. A typical connectionist network, indeed, does not have direct access to the knowledge stored in connection weights. Instead, this knowledge can only be expressed through the influence that it exerts on the model's representations, and such representations may or may not contain readily accessible information, that is, information that can be retrieved with no or low computational cost (Kirsh, 1991). Arguably, symbolic approaches may capture the same distinction through the difference between compiled and interpreted code. It would take too long to discuss the finer issues raised by this possibility here, but two points are worth mentioning. First, all compiled code necessarily existed as interpreted code before compilation took place. This makes the strong prediction, under the assumption that compiled code corresponds to unconscious knowledge and that interpreted code corresponds to conscious knowledge, that all unconscious knowledge we possess at some point previously existed as conscious knowledge. Whether this holds true or not is a matter for empirical investigation, but there is evidence that we are sensitive to regularities that were never made explicit (Pacton, Perruchet, Fayol, and Cleeremans, 2001). Second, whether compiled or interpreted, symbolic computer code always needs a processor to execute it (and hence access it) for it to be causally efficacious. This stands in sharp contrast to the patterns of connection weights that drive processing in connectionist networks, which exert their effects directly, merely as a result of transmitting activation.

Emergent representation

A second principle simply states the following: sensitivity to some regularity does not necessarily imply that the regularity is itself represented as an object of representation. What I mean by this is the following: it is not because you observe that the actions of an agent indicate that it is sensitive to certain regularities (such as in implicit learning situations) that you can conclude that these regularities are represented in its cognitive system as objects of representation that the agent can manipulate intentionally. There are so many examples of the importance of this principle that entire books have been written about it—see for instance the nice popularized treatment of this issue by Steven Johnson (Johnson, 2002), simply titled *Emergence*. Thus, bees construct complex nests and perfectly regular hexagonal cells without any evidence that they even have simple representations of the overall structure of the nest. It's hard not be reminded of behaviorism in this context, but this is certainly one thing behaviorism got right: you don't always need internal representations to account for complex behavior. Of course, one must always be careful not to throw away the baby with the bathwater, to revisit an old cliché: we undeniably entertain systems of complex representations that we can access, manipulate, ponder about and so on—just not always, and just not for anything.

In cognitive psychology, this principle of "emergent representation" has been expressed most clearly through dynamical approaches to Cognitive Science (van Gelder, 1998) and through the connectionist approach (McClelland, 2010). To illustrate, consider the fact that connectionist networks can be often be described as obeying rules without possessing anything like rule-like representations. A very well-known example is Rumelhart and McClelland's (1986a) model of the acquisition of the past tense morphology. In the model, not only are regular verbs processed in just the same way as exceptions, but neither are learned through anything like processes of rule acquisition.

Another example that attracted considerable attention when it was first reported is Hinton's (1986) "family trees" demonstration that a back-propagation network can, through training, become sensitive to the structure of its stimulus environment in such a way that this sensitivity is clearly removed from the surface features of the stimulus material. In Hinton's words, "The structure that must be discovered in order to generalize correctly is not present in the pairwise correlations between input units and output units" (p. 9). The model thus exhibits sensitivity to functional similarity based on the distributional information present in the input, and, as a result, develops abstract knowledge of the relevant dimensions of the domain.

Hinton's network was a relatively simple back-propagation network trained to process linguistic expressions consisting of an agent, a relationship, and a patient, such as for instance "Maria is the wife of Roberto". The stimulus material consisted of a series of such expressions, which together described some of the relationships that exist in the family trees of an Italian family and of an English family. The network was required to produce the patient of each agent-relationship pair it was given as input. For instance, the network should produce "Roberto" when presented with "Maria" and "wife". Crucially, each person and each relationship was presented to the network by activating a single input unit. Hence there was no overlap whatsoever between the input representations of, say, Maria and Victoria. Yet, despite this complete absence of surface similarity between training exemplars, Hinton showed that after training, the network could, under certain conditions, develop internal representations that capture relevant abstract dimensions of the domain, such as nationality, sex, or age. Hinton's point was to demonstrate that such networks were capable of learning richly structured internal representations as a result of merely being required to process exemplars of the domain. Crucially, the structure of the internal representations learned by the network is determined by the manner in which different exemplars interact with each other rather than by their mere similarity expressed, for instance, in terms of how many features (input units) they share—a property that characterizes sensitivity to functional rather than physical similarity. Hinton thus provided a striking demonstration of this important and often misunderstood aspect of associative learning procedures by showing that under some circumstances, specific hidden units of the network had come to act as detectors for dimensions of the material that had never been presented explicitly to the network. These results truly flesh out the notion that rich knowledge can simply emerge as a by-product of processing in structured domains. This introduces a crucial distinction, one that I will return to later, between sensitivity and awareness.

As a final example, consider also that a Simple Recurrent Network (Elman, 1990) trained on only some of the strings that may possibly be generated from a finite-state grammar will generalize to the infinite set of all possible grammatical instances (Cleeremans, Servan-Schreiber, and McClelland, 1989), thus demonstrating perfect, rule-like generalization based only on the processing of a necessarily finite set of exemplars. Interestingly, the representations developed by the network when trained on such material exhibited, under certain conditions, the remarkable property of corresponding almost perfectly with the nodes of the grammar: cluster analyses indeed showed that the similarity structure of the learned internal representations that the network had developed about the relationships between each sequence element and its possible successors reflects the structure of the very grammar the network had been trained on. Again, and crucially, such structure simply emerges out of exposure to relevant stimuli.

Graded processing

The third principle states that information processing as carried out by the brain (i.e., neural computation) is inherently graded (Munakata, 2001). Note that this is not incompatible with the observation of all-or-none outputs. In fact, the logistic function that is so central to many neural network models demonstrates how the relationship between two quantities can be simultaneously graded and dichotomous, just as continuous variations in the temperature of a body of water can make it change state (i.e., freeze) at a critical point. Again, the connectionist literature is replete with striking demonstrations of this principle (Elman et al., 1996). One of the clearest is perhaps McClelland's model of the balance scale problem (Schapiro and McClelland, 2009), in which continuous, incremental learning nevertheless produces both the plateaus and the abrupt, stage-like changes in performance characteristics of many aspects of cognitive development. Another potent illustration of how graded representations can nevertheless produce complex patterns of associations and dissociations between several aspects of behavior is provided by the work of Munakata et al. (1997) on object permanence, in which a Simple Recurrent Network was used to model children's ability to keep active representations of hidden objects. In both cases, the graded nature of the underlying representations is crucial in producing the observed effects; that is, it is precisely by virtue of the fact that representations are graded that such models are successful in accounting both for the steady-changes characteristic of plateaus and for the abrupt-changes characteristic of stage-like transitions. Again, while the implications of graded processing are perhaps clearest in the case of development, they are just as relevant for our understanding of the differences between conscious and unconscious processing for they highlight the fact that qualitative differences can accrue from purely quantitative changes. Whether consciousness is graded or all-or-none is both an important empirical debate as well as a challenging conceptual issue, for it is the case that graded output can be obtained based on the operation of all-or-none computing elements, and that all-or-none output can be obtained based on the operation of graded computing elements. Connectionism, in many cases, has given us new conceptual tools with which to think about the distinction between graded and all-or-none processing.

Mandatory plasticity

This final principle states that learning is a mandatory consequence of information processing. Thus, the brain is inherently plastic. Every experience leaves a trace in many neural pathways. William James stated that "Every impression which impinges on the incoming nerves produces some discharge down the outgoing ones, whether we be aware of it or not" (James, 1890). Donald Hebb (Hebb, 1949) later operationalized this idea in the form of what is now known as the Hebb rule, which simply states that activity between two neurons will tend to increase whenever they are simultaneously active. The Hebb rule, unlike other learning procedures, actually forms the basis for elementary mechanisms of plasticity in the brain, namely long-term potentiation (LTP) and depression (LTD).

O'Reilly and Munakata (2000) proposed an interesting distinction between what they called "model learning" (Hebbian learning) and "task learning" (error-driven learning). Their argument is framed in terms of the different computational objectives each of these types of learning process fulfills: capturing the statistical structure of the environment so as to develop appropriate models of it on the one hand, and learning specific input–output mappings so as to solve specific problems (tasks) in accordance with one's goals on the other hand. There is a very nice mapping between this distinction—expressed in terms of the underlying biology and a consideration of computational principles—and the distinction between incidental learning and intentional learning on the other hand. Thus, as made clear by the manner in which information processing is construed in the connectionist framework, (1) representations are dynamical, constantly causally efficacious objects, and (2) change occurs as soon as information processing takes place. The fact that learning is almost viewed as a by-product of information processing networks accounts very naturally (that is, without requiring further assumptions) for a host of phenomena associated with unconscious cognition, and in particular with implicit learning.

To summarize, these four connectionist principles—active representation, emergent representation, graded processing, and mandatory plasticity—help us recast the differences between conscious and unconscious cognition in a manner that is strikingly different from thoroughly classical approaches. Instead of assuming that representations take the form of inert symbolic propositions that cannot be active unless they are somehow accessed, we now have a constantly causally efficacious network of subsymbolic computational elements (units, neurons). These features make it easy to understand how knowledge can influence behavior in a way that does not entail that the relevant representations be accessed as objects of representation, which is precisely what happens in the many phenomena characteristic of implicit cognition, such as priming, implicit learning and implicit memory.

However, we now face the even greater challenge of understanding how such systems can also account for consciousness. What are the computational principles through which one can characterize the differences between conscious and unconscious representations? This is the question that I attempt to sketch an answer to in the next section.

Consciousness

Numerous theories of consciousness have been proposed over the last twenty years (Atkinson, Thomas, and Cleeremans, 2000). While it would take an entire book to attempt to summarize the state-of-the-art in this respect, it is probably sufficient for the purposes of this text to point out that most theories that have achieved some form of consensus fall into two camps: Global Workspace theories and Higher-Order Thought theories.

Global Workspace Theory (GWT) (Baars, 1988; Dehaene, Kerszberg, and Changeux, 1998) is currently the most consensual account of the functional characteristics of consciousness. According to GWT, conscious representations are globally accessible in a manner that unconscious representations are not. Global accessibility, that is, the capacity for a given representation to influence processing on a global scale (supporting, in particular, a verbal report), is achieved by means of "the neural workspace", a large network of high-level neural "processors" or "modules" linked to each other by long-distance cortico-cortical connections emanating from layer 5 of the cortex. Thus, while information processing can take place without awareness in any given specialized module, once the contents processed by that module enter in contact with the neural workspace, "ignition" occurs and the contents are "broadcast" to the entire brain, so achieving what Dennett (2001) has dubbed "fame in the brain". In this respect, it is interesting to note that in some ways, early connectionist models such as the Interactive Activation Model (McClelland, 1981) already contain the lineaments of GWT.

GWT thus solves the quandary spelled out in the introduction (i.e., which computational principles differentiate between conscious and unconscious cognition) by distinguishing between causal efficacy and conscious access through architecture: on the one hand, knowledge embedded in peripheral modules can bias and influence processing without entering the global workspace, and so remain unconscious. On the other hand, knowledge that is sufficiently supported, both by bottom-up factors such as stimulus strength and by top-down factors such as attention, can "mobilize" the neural workspace, resulting in "ignition" and so become conscious and available for the global control of action.

Higher-Order Thought (HOT) theories of consciousness (Rosenthal, 1997) have a very different flavor. According to HOT, a mental state is conscious when the agent entertains, in a non-inferential manner, thoughts to the effect that it currently is in that mental state. Importantly, for Rosenthal, it is by virtue of occurring HOTs that the target first-order representations become conscious. In other words, a particular representation, say, a representation of the printed letter "J", will only be a conscious representation to the extent that there exists another (unconscious) representation (in the same brain) that indicates the fact that a (first-order) representation of the letter "J" exists at time *t*. Dienes and Perner (1999) have elaborated this idea by analyzing the implicit–explicit distinction as reflecting a hierarchy of different manners in which a given representation can be explicit. Thus, a representation can explicitly indicate a property (e.g., "yellow"), predication to an individual (the flower is yellow), factivity (it is a fact and not a belief that the flower is yellow), and attitude ("I know that the flower is yellow"). Fully conscious knowledge is thus knowledge that is "attitude-explicit". A conscious state is thus necessarily one that the subject is conscious of. While this sounds highly counter-intuitive to some authors (Block, 2011), it captures the central intuition that it is precisely the fact that I know (that I experience the fact, that I feel) that I possess some knowledge that makes this knowledge conscious.

HOT thus solves the problem of distinguishing between conscious and unconscious cognition in a completely different manner, specifically by assuming the involvement of specific kinds of representations the function of which it is to denote the existence of and to qualify target first-order representations. Such HOTs, or metarepresentations, need not be localized in any particular brain region, but of course the densely interconnected prefrontal cortex is a good candidate for such metarepresentations to play out their functions.

Regardless of whether one takes GWT or HOT to best characterize the differences between conscious and unconscious cognition, one question that connectionist thinking about this issue prompts us to ask is: *how do we get there*? How do we *build* the global workspace? Where do metarepresentations come from?

Considering existing theories of consciousness through a connectionist lens offers the tantalizing possibility not only of unifying the two accounts, but also of rooting them both in mechanisms of learning. On this view, unconscious representations constantly compete with each other to capture the best interpretation of the input (Maia and Cleeremans, 2005). This competition is biased by further representations that capture the system's high-level, learned knowledge (its expectations and its goals). The "winning coalitions" come to dominate processing as the result of prior learning, and hence afford the global availability claimed to be constitutive of consciousness by GWT. Global availability is not sufficient, however, for one can perfectly imagine all of the aforementioned taking place without consciousness (as any interactive neural network readily demonstrates). What I surmise to be also necessary, congruently with the assumptions of HOT, is that the winning representations *be known as objects of representation* by the system that possesses them. In other words, that first-order representations be redescribed by other representations in such a way as to make the former be identified or recognized by the system as familiar states of knowledge, that is, "attitude-explicit" in the terminology of Dienes and Perner.

In the following, I first attempt to flesh out the main computational principles that differentiate GW-like theories from HOT theories of consciousness. Next, I describe recent simulation work in which we specifically explore how it may be possible to build connectionist models that capture the central intuition of HOT, namely that knowledge is conscious when it is appropriately redescribed by means of metarepresentations.

The radical plasticity thesis

In other works (Cleeremans, 2008, 2011), I have defended the idea that consciousness is itself the result of learning. From this perspective, agents become conscious by virtue of learning to redescribe their own activity to themselves; a perspective that finds strong echoes in Allkahverdov's (this volume) own theory. Taking the proposal that consciousness is inherently dynamical seriously opens up the mesmerizing possibility that conscious awareness is itself a product of plasticity-driven dynamics. In other words, from this perspective, we learn to be conscious. To dispel possible misunderstandings of this proposal right away, I am not suggesting that consciousness is something that one learns like one would learn about the Hundred Years War, that is, as an academic endeavor, but rather that consciousness is the result (vs. the starting point) of continuous and extended interaction with the world, with ourselves, and with others. The brain, from this perspective, continuously (and unconsciously) learns to anticipate the consequences of its own activity on itself, on the environment, and on other brains, and it is from the practical knowledge that accrues in such interactions that conscious experience is rooted. This perspective, in short, endorses the enactive approach introduced by O'Regan and Noë (2001), but extends it both inwards (the brain learning about itself) and further outwards (the brain learning about other brains), so connecting with the central ideas put forward by the predictive coding approach to cognition. In this light, the conscious mind is the brain's (implicit, enacted) theory about itself, expressed in a language that other minds can understand.

The theory rests on several assumptions and is articulated over three core ideas. A first assumption is that information processing as carried out by neurons is intrinsically unconscious. There is nothing in the activity of individual neurons that makes it so that their activity should produce conscious experience. Important consequences of this assumption are (1) that conscious and unconscious processing must be rooted in the same set of representational systems and neural processes, and (2) that tasks in general will always involve both conscious and unconscious influences, for awareness cannot be "turned off" in normal participants.

A second assumption is that information processing as carried out by the brain is graded and cascades (McClelland, 1979) in a continuous flow (Eriksen and Schultz, 1979) over the multiple levels of a heterarchy (Fuster, 2008) extending from posterior to anterior cortex as evidence accumulates during an information processing episode. An implication of this assumption is that consciousness takes time.

The third assumption is that plasticity is mandatory: the brain learns all the time, whether we intend to or not. Each experience leaves a trace in the brain (Kreiman, Fried, and Koch, 2002). With these assumptions in place, the theory is articulated around three core ideas that I will now briefly expose.

Quality of representation

The first core idea is that consciousness depends on quality of representation (see Figure 2.1). "Quality of representation" (QoR), here, designates graded properties of neural representations, specifically their Strength, their Stability in time, and their Distinctiveness. QoR depends both on bottom-up factors such as stimulus properties and on top-down factors such as attention. QoR determines the extent to which a representation is available to (1) influence behaviour, (2) form the contents of awareness, and (3) be the object of cognitive control and other high-level processes. Crucially, QoR changes as a function of learning and plasticity over different time scales (processing within a single trial, learning, and development), as depicted in Figure 2.1. The first region of the figure, labeled "Weak (implicit)", corresponds to the point at which processing starts in the context of a single trial, or to some early stage of development or skill acquisition. This stage is characterized by weak, poor-quality representations. Implicit representations are capable of influencing behavior, but only weakly so (e.g., through priming).

The second region corresponds to the emergence of higher-quality explicit representations, here defined as representations over which one can exert control. Such representations are good candidates for redescription and can thus be recoded in different ways, e.g., as linguistic propositions (supporting verbal report).

The third region involves what I call automatic representations, that is, representations that have become so strong that their influence on behavior can no longer be inhibited (e.g., as in the Stroop situation). Such representations exert a mandatory influence on processing.

Importantly, however, and unlike the weak representations characteristic of implicit cognition, one is (at least potentially) aware of possessing such strong representations and of their influence on processing. Thus, both the weak representations characteristic of implicit cognition and the very strong representations characteristic of automaticity cannot be controlled, but for very different reasons. This leaves intermediate-quality (explicit) representations, that is, representations that are strong enough that their influence on behavior needs to be monitored yet not sufficiently adapted that they can be "trusted", as those representations

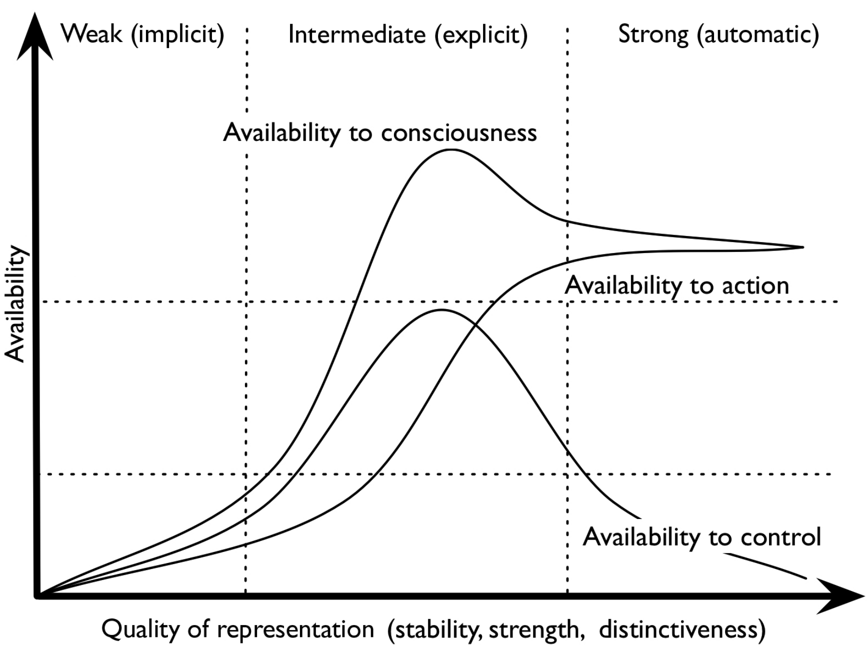

FIGURE 2.1 The QoR framework.

that require the most cognitive control. Crucially, this also predicts that intermediate-quality representations are the most susceptible to being influenced by other sources of knowledge, as they are the most flexible. One would thus expect non-monolithic effects as expertise develops, in different paradigms ranging from perception to motor skill learning.

Metarepresentation

The second core idea is that consciousness depends on the involvement of metarepresentations. Indeed, QoR cannot be the only factor that shapes availability to different aspects of consciousness. Even strong stimuli can fail to enter conscious awareness—this is what happens in change blindness (Simons and Levin, 1997), in the attentional blink (Shapiro, Arnell, and Raymond, 1997), or in inattentional blindness (Mack and Rock, 1998). States of altered consciousness like hypnosis, and pathological states such as blindsight (Weiskrantz, 1986) or hemineglect likewise suggest that high-quality percepts can fail to be represented in awareness while remaining causally efficacious. This suggests that QoR, while necessary for conscious awareness, is not sufficient.

One way of understanding what is missing is to appeal to the central hypothesis of the HOT theory of consciousness (Rosenthal, 1997; see also Lau and Rosenthal, 2011), namely that a representation is a conscious representation when one knows that one is conscious of the representation. This roots conscious awareness in a system's capacity to redescribe its own states to itself, a process ("representational redescription") also viewed as central during cognitive development (Karmiloff-Smith, 1992) and metacognition in general (Nelson and Narens, 1990). A system's ability to redescribe its own knowledge to itself depends (1) on the existence of recurrent structures that enable the system to access its own states, and (2) on the existence of predictive models (metarepresentations) that make it possible for the system to characterize and anticipate the occurrence of first-order states (Bar, 2009; Friston, 2006; Wolpert, Doya, and Kawato, 2004). Such redescription is also uniquely facilitated, in humans, by language, viewed here as the metarepresentational tool par excellence. A natural spot for such metarepresentations to perform their functions is the prefrontal cortex (i.e., Crick and Koch's "the front is looking at the back" principle (Crick and Koch, 2003)). Importantly however, here, such metarepresentational models (1) may be local and hence occur anywhere in the brain, (2) can be subpersonal, and (3) are subject, just like first-order representations, to plasticity and hence can themselves become automatic. Metacognition, just like cognition, can thus involve implicit, explicit, or automatic metarepresentations.

The theory thus proposes a novel conception of skill acquisition that links automaticity with the observation that conscious awareness seems to proceed from the top down (i.e., Crick and Koch's "the high levels first" principle, see Crick and Koch, 2003): we become aware of the higher-level aspects (the gist) of a scene before becoming aware of its lower-level features. I suggest that this stems from the fact that, from a computational point of view, metarepresentations implement what one could call cortical reflexes or shortcuts: a system that has learned to redescribe the activity of an entire feedforward pathway can now also anticipate the consequences of early activity in such a chain on its output faster than the pathway itself can compute the output. As a result, adapted metarepresentations (and only adapted metarepresentations) make it possible to bypass the first-order pathway altogether. I surmise that this accounts not only for the fact that the time course of (expert) perception seems to follow a reverse hierarchy (Ahissar and Hochstein, 2004), but also for the fact that automaticity entails loss of access to the contents computed along the first-order pathway. By the same token, this also opens up the possibility for postdictive effects in conscious experience, as metarepresentations are shaped by first-order processing. This top-down view of automaticity contrasts with extant theories (Chein and Schneider, 2012).

With these ideas in place, we can now ask "When is knowledge unconscious?" Figure 2.2 shows a simple network organized in two different pathways: a (horizontal) first-order pathway comprising five layers of units, and a (vertical) second-order pathway simplified here to show only a single layer of "metarepresentational units" (see Pasquali et al., 2010, for implemented instantiations of such networks). The figure is aimed to distinguish four stages that such networks traverse as they learn to carry out a particular task. These four stages correspond to four different ways in which knowledge may remain unconscious.

First (Figure 2.2a), knowledge embedded in synapses is assumed not be accessible at all, for such knowledge fails to be instantiated in the form of active patterns of neural activity (Koch, 2004), a necessary condition for their contents to be available to awareness. The provocative idea here is that the brain does not know, e.g., that SMA activity consistently precedes M1 activity. To represent this causal link to itself, it therefore has to learn to redescribe its own activity so that the causal link is now represented explicitly as a metarepresentation. Second, weak representations (Figure 2.2b), while they can influence behavior, remain unconscious for they fail to be sufficiently strong to be the target of metarepresentations. Third, when sufficiently strong, first-order representations can begin to be redescribed into metarepresentations (Figure 2.2c), yet, other conditions (e.g., lack of attention induced by distraction, failure to properly redescribe first-order contents) may make such redescription impossible or difficult. Fourth, the very strong representations characteristic of automaticity (Figure 2.2d) are not necessary anymore to drive behavior since the learned metarepresentations now implement a faster "shortcut" pathway from input to output. This also accounts for the fact that metacognitive accuracy often lags behind first-order performance initially, but precedes first-order performance with expertise (i.e., I know that I know the answer to a query before I can actually answer the query).

The distinctions introduced here overlap partially with the distinctions introduced by existing theories of consciousness: Dehaene's conscious–preconscious–unconscious taxonomy (Dehaene, Changeux, Naccache, Sackur, and Sergent, 2006), Lamme's Stages 1/2/3/4 framework (Lamme, 2006), and Kouider's partial awareness

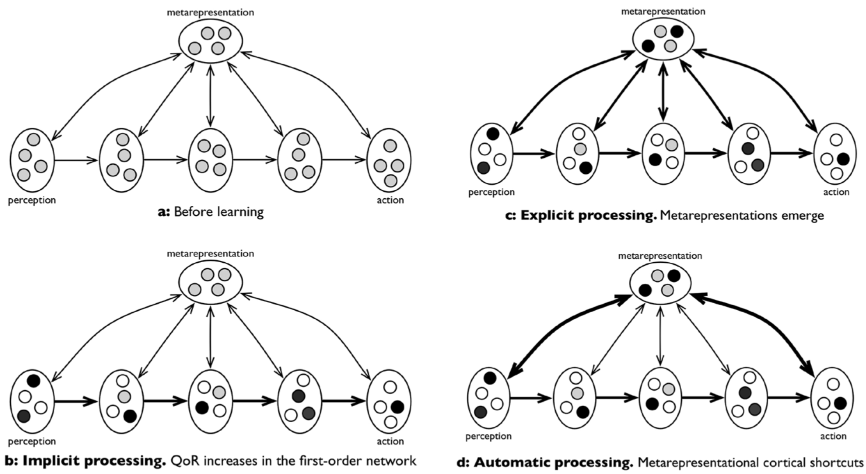

FIGURE 2.2 Implicit, explicit and automatic processing.

hypothesis (Kouider, de Gardelle, Sackur, and Dupoux, 2010), but uniquely frame the transitions dynamically as resulting from learning.

Theory of mind and self-awareness

The third core idea is that consciousness depends on theory of mind (Schilbach et al., 2013; Timmermans, Schilbach, Pasquali, and Cleeremans, 2012). The emergence of an agent's ability to redescribe its own representations to itself in the way sketched above, I argue, critically depends on the agent being embedded in interaction with other agents. From this perspective, as Frege pointed out, conscious experience cannot be understood independently from the agent who has these experiences. Yet, as obvious as this may seem, neuroscientists have approached the question as though the differences between conscious and unconscious representations could be understood independently of the subject, from a purely "objective", third-person point of view. The entire "search for the Neural Correlates of Consciousness" is, in this sense at least, misguided. As Donald (2001) put it, "the human mind is unlike any other on this planet, not because of its biology, which is not qualitatively unique, but because of its ability to generate and assimilate culture" (p. xiii). Thus, I build a model of myself not only by developing a non-conceptual understanding of how my goals are eventually expressed in action, but also by understanding how agents similar to me react to actions directed towards them. It is thus essential that we strive to understand how interactions with other agents shape our own conscious experiences.

Putting the three core ideas together, we end up with the radical plasticity thesis (Cleeremans, 2008, 2011), that is, with the idea that consciousness emerges in cognitive systems that are capable of learning to redescribe their own activity to themselves. In other words, one "learns to be conscious".

Experimental strategies

The framework outlined above suggests different experimental strategies through which to manipulate the extent to which processing is conscious vs. unconscious. To develop this, consider first that HOT theories, much like Nelson and Narens' (1990) metamemory framework, conjure an image of consciousness as a monitoring system, a radar of sorts, that continuously sweeps and redescribes our own representations, so carrying out what Lau (2008) dubbed "signal detection on our mind".

But the radar analogy only gets us this far, as there is more to consciousness than mere monitoring. First, consciousness creates and shapes its own phenomenal field. By this I mean that what one experiences changes the very structure of the phenomenal space in which those experiences take place. Think of the impact that expertise exerts on the contents of consciousness. Expert bird watchers, for instance, are capable of recognizing hundreds of species of bird on sight. Contrast that with non-expert knowledge of birds, which contains perhaps a dozen broad categories of birds at most. Expertise thus enriches conscious experience; the very perceptual contents to which we have access are now far more distinctive and detailed than they used to be before the relevant skill was acquired. But, strikingly, expertise also eliminates contents from conscious experience. Consider reading, for instance. Literacy makes us read in a very efficient manner, with fine-tuned eye movements that make us skip over function words in such a way that they literally fade away from awareness. Likewise, driving a car has a decidedly different character when one has become an expert at it than when one is learning to drive—the many control movements we perform while driving in the city, for instance, have vanished from our phenomenal field to the point that they can simply unfold without the need for conscious monitoring of any kind.

Second, while actual radars merely continuously sweep the space they are monitoring, consciousness can be intentionally directed to scrutinize particular regions of it. This, of course, is a metaphorical way of describing attention, the association of which to consciousness itself is complex (Lamme, 2006) but empirically trackable.

Third, unlike radars, human agents have preferences driven by affect. There are some parts of their phenomenal space they would rather pay no attention to, and there are some parts of their phenomenal space they would be inclined to spend a long time looking at. This, of course, is where learning enters the picture, but also *caring*—agents care about some experiences more than about others; this is what makes them do certain things rather than others. In this sense, subjective experience works almost as a currency of sorts. What would be the point of doing anything if the doing wasn't doing something to you? Beyond mere survival, all organisms seek pleasurable states and attempt to avoid unpleasant or dangerous ones. But no machine ever does. This is the singular challenge artificial consciousness is facing: how to get artificial agents to have reasons for doing things rather than merely doing things for reasons (paraphrasing Dennett). The claim here is thus

(1) that subjective experience is what makes it possible for an agent to find its own reasons for doing things, and (2) that this is only possible if the agent is able both to know about its own internal states and to learn about them in ways that make it possible for it to come to care about them.

Fourth, human consciousness, unlike a radar, is able to actively shape its own phenomenal space, that is, to control the action systems that make it possible for the agent to come to experience certain things rather than others. Controlling action to shape perception is of course possible without awareness—a good example is perhaps saccadic eye movements—but the key point here is that consciousness enables far more sophisticated control driven by the agent's mental states—its intentions, desires, hopes, and so on.

If one accepts this admittedly very rough metaphor and its many qualifications, it is interesting to consider what it takes to fool the radar, that is, to escape detection. There are essentially three strategies, each corresponding to one of three regions (implicit, explicit, automatic) depicted in Figure 2.1. One is to be small enough to simply fail to register. While a plane will register on an airspace radar, for instance, this would not be the case for a bird or a small drone. In cognitive psychology, this strategy amounts to *weakening the stimulus*, that is, reducing its energy or its temporal duration, or making it less distinguishable from noise, in such a way as to render it phenomenologically unconscious. This is the strategy that most experimental designs based on masking or crowding leverage, as in subliminal priming for instance.

The second strategy one could use to fool a radar is to *divert its attention*, or rather, the attention of the radar operator. Here, there is something that has sufficient energy to be detected, yet it fails to be noticed because the radar operator is not looking at it or because his attentional resources are otherwise engaged. This is precisely what happens in paradigms such as inattentional blindness, change blindness, or the attentional blink, but also in any paradigm that is designed to overload working memory capacity. While *weakening the stimulus* is a bottom-up strategy, *diverting attention* is clearly a top-down strategy. In the first case, one manipulates the quality of incoming stimuli in such a way that they are not available to form the contents of consciousness even when attention is directed towards them, whereas in the other case, a stimulus that has the necessary quality to be available to form the contents of awareness fails to become conscious because insufficient attention is dedicated to it.

There is a second top-down strategy that is somewhat more unusual: it consists of *changing the narrative*. To return to our radar analogy, this would consist of making the radar operator believe not that there is *nothing* out there, but rather that what is there is *not what he thinks it is*. For instance, one plane may pass itself off as another; or a flock of birds might organize itself so as to look like a plane (though why birds would do that is open to question). This is connected to the well-known issue of misrepresentation in the philosophy literature (Dretske, 1986). In cognitive psychology, perhaps the best-known contemporary example is the phenomenon of choice blindness (Johansson et al., 2005), whereby participants are made to believe that a photograph of a person they have just *not chosen* is in fact the one they chose. This trick makes a number of them confabulate reasons for their choice, just as in Nisbett and Wilson's (1977) famous "nylon stockings" experiment. But this kind of paradigm is just one instance of a wide variety of such manipulations of people's conscious beliefs. One also immediately thinks, for instance, of hypnosis, and of the placebo effect. The latter is particularly striking, since it demonstrates that the mere belief that a pill contains an active ingredient is sufficient to profoundly modify people's subjective appraisal of their symptoms. Likewise, in hypnosis, people can be convinced that their actions are not their own; that their arm is lifting up "on its own". This reduced sense of agency under hypnosis is, according to some (i.e., Dienes and Perner, 2007; Lush, Naish, and Dienes, 2016), the defining feature of the hypnotic phenomenon. Such high-level manipulations of what people believe to be the case can penetrate deeply in the cognitive hierarchy. Raz and collaborators (Raz, Fan, and Posner, 2005), for instance, have shown that a hypnotic suggestion that "words would appear as mere gibberish" is sufficient to reduce Stroop interference in a color-naming task. We have replicated this finding ourselves using a placebo-aided non-hypnotic suggestion that color perception would either be enhanced or deteriorated by a fake apparatus to which people were connected (Magalhães de Saladanha da Gama et al. 2013).

Each of these three strategies to fool consciousness comes with its own challenges, which are often thorny and extremely difficult to properly address. Subliminal perception, and specifically the claim that a stimulus can be weak enough to be available to consciousness yet strong enough that it remains causally efficacious, such as in its ability to prime subsequent decisions, continues to elicit vivid debate today. This should come as no surprise, because it turns out it is exceedingly challenging to come up with experimental designs in which a stimulus can, precisely, be weak enough to "fly under the radar" yet strong enough to be causally efficacious. This is the quandary that faces all researchers interested in demonstrating unconscious cognition. I call it the *strength–efficacy dilemma*, and it is a dilemma that comes up in any paradigm leveraging the strategy of weakening the stimulus. This gets complicated by the further challenges involved in properly assessing awareness itself. Should we ask participants to rate visibility on each trial, or ask them to take part in a visibility test after the main (e.g., priming) task is completed? In either case, should we use binary (seen/unseen) measures or graded reports such as the Perceptual Awareness Scale (PAS, see Ramsoy and Overgaard, 2004) or a fully continuous scale (Sergent and Dehaene, 2004)? Crucially, these methodological decisions, which at first sight appear to be secondary, are demonstrably crucial in that they change our conclusions about the extent to which processing was indeed unconscious.

To wit, consider the fact that a systematic comparison between different ways of collecting subjective judgements about visibility can sometimes yield strikingly different results for each method. Thus, Sandberg et al. (2010) presented participants with briefly displayed masked shapes (a square, a circle, a triangle, and a lozenge) in a psychophysical paradigm on each trial of which they had to (1) identify the shape by pressing on one of four buttons and (2) express a subjective judgement of visibility or confidence, using one of three different scales all involving four points: PAS, a confidence judgement, or post-decision wagering (Persaud, McLeod, and Cowey, 2007). Briefly put, the study revealed that PAS was the most exhaustive scale: when using PAS, people reported experiencing something ("a brief glimpse") at shorter stimulus durations than they did with the other two scales. This finding has obvious implications for our interpretation of dissociation results (e.g., between priming and visibility), for they suggest that many such dissociation findings may be a mere consequence of lack of sensitivity.

Further, the study also demonstrated that while people tend to use post-decision wagering in a binary manner, betting either high or low and mostly refraining from using the intermediate scale points, this was not the case with PAS, which showed a much more distributed use of the different scale points. This suggests that intermediate or graded states of awareness are possible and that people are quite willing to report them when offered the possibility.

In recent research, Windey, Gevers, and Cleeremans (2013) focused specifically on this question, that is, whether consciousness should be taken to be an all-or-none or a graded phenomenon. Using a simple task consisting of judging either the magnitude or the color of briefly presented, masked Arabic numerals, we found that the shape of the psychophysical functions relating stimulus duration with subjective visibility assessed by PAS were influenced by level of processing: such functions were more linear when judging color than when judging magnitude, which is suggestive that both stimulus features and level of processing modulate the extent to which perceptual awareness appears graded or dichotomous. This turns out to be important when comparing different studies, for different authors typically use different stimuli. Lamme and colleagues (see e.g., Fahrenfort, Scholte, and Lamme, 2008), for instance, who defend the idea that phenomenal consciousness has a graded character, have often used low-level stimuli such as gratings and Gabor patches, whereas researchers who defend the idea that consciousness involves a sharply non-linear transition (Del Cul, Baillet, and Dehaene, 2007) tend to use high-level stimuli such as numbers or words. Further studies in this vein confirmed and extended our results (see Anzulewicz et al., 2015; Windey, Vermeiren, Atas, and Cleeremans, 2014).

One may think that such measurement issues are only problematic when subjective measures such as visibility or confidence are collected. However, even objective measures such as *d'* can exhibit substantial variability. Thus, Vermeiren and Cleeremans (2012) showed that *d'*, as measured in a prime visibility test administered after a priming task involving metaconstrast masked arrows, was influenced by different factors such as (1) the delay that elapsed between stimulus presentation and response, (2) whether the target (visible) stimulus was neutral or oriented, and (3) whether attention was divided between the prime and the target. Observed differences in *d'* magnitude could be as large as 1.0 in some conditions, thus casting doubt on the use of *d'* as a neutral and objective measure of visibility, and again bearing strong implications for the interpretation of dissociation findings.

Thus, the general take-home message is that the measure we use to assess awareness matters very much, for different measures reveal different dynamics and can lead to very different conclusions about the extent to which processing was unconscious or not.

The second strategy, namely *diverting attention*, has challenges of its own. The most important challenge presents itself in the form of *observer paradox*, that is, the fact that in cognitive systems, observing a process actually changes that very process. Thus, if my goal is to draw people's attention away from a stimulus, I simply cannot ask participants "whether they noticed something", for doing so attracts their attention precisely to what I wish them to remain unaware of. This limits possible designs to single-trial experiments, as in Mack and Rock's (1998) well-known inattentional blindness experiments, or forces us to collect subjective reports after the entire experiment is over, a method that is problematic in and of itself, as is also the case for subliminal perception experiments.

This particular challenge is one of several listed by Newell and Shanks (2014) in a recent review of unconscious decision-making. According to the authors, any measure of awareness should fulfil the following four criteria. First, the measure should be *reliable*, that is, independent of experimental demands or social desirability. Second, it should be *relevant*, that is, probe participants about the very same knowledge that is involved in subtending performance. Third, it should *immediate*, so as to avoid forgetting and interference. Ideally thus, any measure of awareness (e.g., a visibility or a confidence judgement) should be administered on a trial-by-trial basis rather than after the entire main task is completed. Fourth, the measure should *sensitive*, which is to say based on the same material used to elicit behaviour and fine-grained enough that it can be as exclusive and exhaustive as possible (Reingold and Merikle, 1988). On the face of it, very few paradigms can claim to fulfill all four criteria.

The third strategy—*changing the narrative*—presents the challenge of being able to develop a convincing cover story to describe the situation to participants. It is also prone to the "retrospective assessment" problem (Shanks and St. John, 1994), as it is impossible to satisfy the immediacy criterion: one cannot test people's beliefs online without undermining such beliefs, the solidity of which is crucial for the effects to obtain. Nevertheless, in recent work, we used this strategy in the hope of demonstrating implicit learning. Implicit learning (see Cleeremans, Destrebecqz, and Boyer, 1998, for a review) is a notoriously challenging domain insofar as demonstrating the involvement of unconscious knowledge is concerned, for it presents most of the challenges we have examined so far.

Thus, in a recent series of experiments, Alamia et al. (2016) asked participants to detect the direction most of the dots contained in a random-dot kinematogram were moving. On some trial blocks, the entire patch of dots was colored either green, red, or blue. Two of the colors were predictive of the response and the third was neutral. There was thus a very simple way for participants to improve their performance in this otherwise very difficult task (difficulty was staircased individually). Crucially, however, participants were not told about the predictive value of the colors. Instead, they were told that the colors were introduced to make the task more difficult, and that they would have to report the color on some of the trials.

We found that people quickly learned to use the colors to improve their performance, albeit most were (1) unable to report noticing the association between color and motion direction, (2) unable to carry out generation tasks asking them to report the color associated with a particular motion direction, and (3) unable to recognize whether a color/motion association was familiar or not. Thus, albeit all of these tests of awareness had to be administered after the main experiment was over so as to avoid the observer paradox, we are confident that these results demonstrate that people were able to unconsciously use available predictive information to improve their performance, thus demonstrating implicit learning. The central feature that makes this design successful is probably the use of a convincing story to characterize the function of the colors. By *changing the narrative* in this way, we made sure people actually pay attention to the relevant stimuli (which is necessary for learning to take place), while at the same time ensuring that they form inaccurate metarepresentations about the role that such stimuli play in the task.

Conclusion

From the perspective presented here, the brain is continuously and unconsciously learning to anticipate the consequences of action or activity on itself, on the world, and on other people. Thus, we have three closely interwoven loops (Figure 2.3) all driven by the very same prediction-based mechanisms. A first internal or "inner loop" involves the brain redescribing its own representations to itself as a result of its continuous unconscious attempts at predicting how activity in one region influences activity in other regions. In this light, consciousness amounts to the brain performing signal detection on its own representations (Lau, 2008), so continuously striving to achieve a coherent (prediction-based) understanding of itself. It is important to keep in mind that this inner loop in fact involves multiple layers of recurrent connectivity, at different scales throughout the brain. A second "perception-action loop" results from the agent as a whole predicting the consequences of its actions on the world. The third loop is the "self-other loop", and links the agent with other agents, again using the exact same set of mechanisms as involved in the other two loops. The existence of this third loop is constitutive of conscious experience, I argue, for it is by virtue of the fact that as an agent I am constantly attempting to model other minds that I am able to develop an understanding of myself.

In the absence of such a "mind loop", the system can never bootstrap itself into developing the implicit, embodied, transparent (Metzinger, 2003) model of itself that forms the basis, through HOT theory, of conscious experience. The processing carried out by the inner loop is thus causally dependent on the existence of both the perception–action loop and the self–other loop, with the entire system thus forming a "tangled hierarchy" (e.g., Hofstadter's concept of a "strange loop", see Hofstadter, 2007) of predictive internal models (Pacherie, 2008; Wolpert, Doya, and Kawato, 2004).

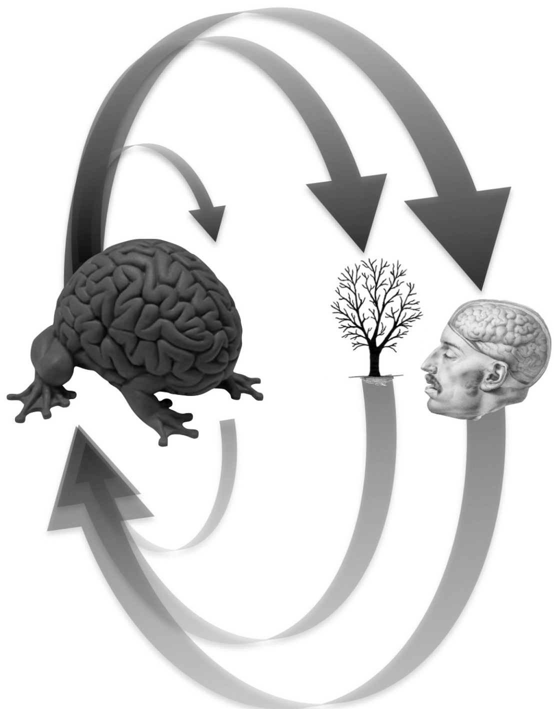

FIGURE 2.3 Tangled loops.

To conclude, I would thus like to defend the following claim (see also Cleeremans, 2014): conscious experience occurs if and only if an information processing system has learned about its own representations of the world. Consciousness, in this light, is thus the brain's implicit, embodied theory about itself, gained through experience interacting with itself, with the world, and with other people. It is subtended by continuously operating prediction-driven learning mechanisms applied to all levels of a representational hierarchy that make it possible for cognitive agents to know themselves—something that first-order systems are simply incapable of achieving. Unconscious processing and representations thus leverage the same mechanisms as those involved in conscious processing, with a crucial difference brought about by the involvement of specific kinds of learned metarepresentations geared towards redescribing first-order knowledge in increasingly informative ways.

So, the mind is deep after all—its depth, just like the depth of canyons, is the result of entire lifetimes of experience progressively carving out a mental landscape that exerts greater and greater influence on the way our conscious mental states express themselves (something Chater calls "traditions"). Exploring this complex, dynamical relationship between conscious and unconscious processing mandates using paradigms that make it possible to document the ebbs and flows of each—a challenge that implicit learning paradigms are uniquely positioned to address. As Pierre Perruchet wrote, "Learning shapes conscious experience; conscious experience shapes learning". So, it begins and ends with consciousness rather than with the unconscious indeed—but it is the unconscious that ultimately gives conscious experience its subjective character, just as its own shape is carved out by the conscious acts we engage in. That loop is deep indeed. . .

Acknowledgements

Axel Cleeremans is a Research Director with the Fund for Scientific Research (F.R.S.-FNRS, Belgium). This work was supported by ERC advanced grant RADICAL to Axel Cleeremans. Substantial portions of this article appeared in previous publications, in particular Cleeremans, A. (2014). Connecting conscious and unconscious cognition. *Cognitive science, 38*(6), 1286–1315; and Cleeremans, A. (2017). As above, so below: tangled loops between consciousness and the unconscious. In Z. Radman (Ed.), *Before Consciousness: in Search of the Fundamentals of Mind*. Exeter: Imprint Academic, pp. 83–105.

References

Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Alamia, A., de Xivry, J.-J., San Anton, E., Olivier, E., Cleeremans, A., and Zenon, A. (2016). Unconscious associative learning with conscious cues. *Neuroscience of consciousness*, *2016*(1), 1–10.

Anzulewicz, A., Asanowicz, D., Windey, B., Paulewicz, B., Wierzchon, M., and Cleeremans, A. (2015). Does level of processing affect the transition from unconscious to conscious perception? *Consciousness and Cognition*, *36*(1–11).

Atkinson, A. P., Thomas, M. S. C., and Cleeremans, A. (2000). Consciousness: mapping the theoretical landscape. *Trends in Cognitive Sciences*, *4*(10), 372–382.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Balsdon, T. and Clifford, C. W. G. (2018). Visual processing: conscious until proven otherwise. *Royal Society Open Science*, *5*, 171783.

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society B*, *364*, 1181–1182.

Bates, E. A. and Elman, J. L. (1993). Connectionism and the study of change. In M. Johnson (Ed.), *Brain Development and Cognition: a Reader* (pp. 623–642). Oxford, UK: Blackwell.

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

Block, N. (2011). The higher-order approach to consciousness is defunct. *Analysis*, *71*(3).

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch and B. Lloyd (Eds.), *Cognition and Concepts* (pp. 16–211). Mahwah, N.J.: Lawrence Erlbaum Associates.

Chater, N. (2018). *The Mind is Flat*. London: Allen Lane.

Chein, J. M. and Schneider, W. (2012). The brain's learning and control architecture. *Current Directions in Psychological Science*, *21*(2), 78–84.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

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

Cleeremans, A. (2008). Consciousness: the radical plasticity thesis. *Progress in Brain Research*, *168*, 19–33.

Cleeremans, A. (2011). The radical plasticity thesis: how the brain learns to be conscious. *Frontiers in Psychology*, *2*, 1–12.

Cleeremans, A. (2014). Connecting conscious and unconscious cognition. *Cognitive Science*, *38*(6), 1286–1315.

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

Cleeremans, A. and Dienes, Z. (2008). Computational models of implicit learning. In R. Sun (Ed.), *The Cambridge Handbook of Computational Modeling* (pp. 396–421). Cambridge: Cambridge University Press.

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

Cleeremans, A., Servan-Schreiber, D., and McClelland, J. L. (1989). Finite state automata and simple recurrent networks. *Neural Computation*, *1*, 372–381.

Crick, F. H. C. and Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119–126.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211.

Dehaene, S., Kerszberg, M., and Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the U.S.A.*, *95*(24), 14529–14534.

Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PloS Biology*, *5*(10), e260.

Dennett, D. C. (1982). Styles of mental representation. *Proceedings of the Aristotelian Society, New Series, LXXXIII*, 213–226.

Dennett, D. C. (2001). Are we explaining consciousness yet? *Cognition*, *79*, 221–237.

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

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

Dienes, Z. and Perner, J. (2007). The cold control theory of hypnosis. In G. Jamieson (Ed.), *Hypnosis and Conscious States: the Cognitive Neuroscience Perspective* (pp. 293–314). Oxford: Oxford University Press.

Dijksterhuis, A. and Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives in Psychological Science*, *1*(2), 95–109.

Donald, M. (2001). *A Mind So Rare*. New York: W.W. Horton.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, Content and Function* (pp. 17–36). Oxford, U.K.: Oxford University Press.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.
Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: a Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
Eriksen, C. W. and Schultz, D. W. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Attention, Perception and Psychophysics*, *25*(4), 249–263.
Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision*, *8*(1), 1–12.
Fodor, J. A. (1975). *The Language of Thought*. New York, NY: Harper & Row.
Freud, S. (1949). *An Outline of Psychoanalysis* (J. Strachey, Trans.). London: Hogarth Press.
Friston, K. (2006). A free energy principle for the brain. *Journal of Physiology (Paris)*, *100*, 70–87.
Fuster, J. M. (2008). *The Prefrontal Cortex* (fourth ed.). London: Academic Press.
Gibson, F., Fichman, M., and Plaut, D. C. (1997). Learning in dynamic decision tasks: computational models and empirical evidence. *Organizational Behavior and Human Decision Processes*, *71*, 1–35.
Hebb, D. O. (1949). *The Organization of Behavior*. New York, NY: Wiley.
Hinton. (1986). *Learning Distributed Representations of Concepts*. Paper presented at the 8th Annual Conference of the Cognitive Science Society.
Hofstadter, D. R. (2007). *I Am a Strange Loop*. New York: Basic Books.
Hohwy, J. (2013). *The Predictive Mind*. Oxford, U.K.: Oxford University Press.
James, W. (1890). *The Principles of Psychology*. New York: H. Holt and company.
Johansson, P., Hall, L., Sikström, S., and Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*, 116–119.
Johnson, S. (2002). *Emergence: the Connected Lives of Ants, Brains, Cities, and Software*. London: Penguin Books.
Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar Straus Giroux.
Karmiloff-Smith, A. (1992). *Beyond Modularity: a Developmental Perspective on Cognitive Science*. Cambridge: MIT Press.
Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237*(1445–52).
Kirsh, D. (1991). When is information explicitly represented? In P. P. Hanson (Ed.), *Information, Language, and Cognition*. New York, NY: Oxford University Press.
Koch, C. (2004). *The Quest for Consciousness: a Neurobiological Approach*. Englewood, CO: Roberts and Company Publishers.
Kouider, S., de Gardelle, V., Sackur, J., and Dupoux, E. (2010). How rich is consciousness: the partial awareness hypothesis. *Trends in Cognitive Sciences*, *14*(7), 301–307.
Kreiman, G., Fried, I., and Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences of the U.S.A.*, *99* (8378–8383).
Lamme, V. A. F. (2006). Toward a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501.
Lau, H. (2008). A higher-order Bayesian Decision Theory of consciousness. In R. Banerjee and B. K. Chakrabarti (Eds.), *Models of brain and mind: Physical, computational and psychological approaches: Progress in Brain Research* (Vol. 168, pp. 35–48). Amsterdam: Elsevier.
Lau, H. and Rosenthal, D. (2011). Empirical support for higher-order theories of consciousness. *Trends in Cognitive Sciences*, *15*(8), 365–373.
Loftus, E. F. and Hoffman, H. (1989). Misinformation and memory: the creation of new memories. *Journal of Experimental Psychology: General*, *118*(1).

Lush, P., Naish, P., and Dienes, Z. (2016). Metacognition of intentions in mindfulness and hypnosis. *Neuroscience of Consciousness*, *1*,1–10.

Mack, A. and Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: MIT Press.

Magalhães de Saladanha da Gama, P., Slama, H., Caspar, E., Gevers, W., and Cleeremans, A. (2013). Placebo-suggestion modulates conflict resolution in the Stroop Task. *PLos One*, *8*(10): e75701.

Maia, T. V. and Cleeremans, A. (2005). Consciousness: converging insights from connectionist modeling and neuroscience. *Trends in Cognitive Sciences*, *9*(8), 397–404.

Mathis, W. D. and Mozer, M. C. (1996). Conscious and unconscious perception: a computational theory. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 324–328). Hillsdale, N.J.: Lawrence Erlbaum Associates.

McClelland, J. L. (1979). On the time-relations of mental processes: an examination of systems in cascade. *Psychological Review*, *86*, 287–330.

McClelland, J. L. (1981). Retrieving general and specific information from stored knowledge of specifics. *Proceedings of the Third Annual Cognitive Science Society*, 170–172.

McClelland, J. L. (2010). Emergence in cognitive science. *Topics in Cognitive Science*, *2*(4), 751–770.

McClelland, J. L. and Rumelhart, D. E. (1986). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.

Metzinger, T. (2003). *Being No One: the Self-Model Theory of Subjectivity*. Cambridge, MA: Bradford Books, MIT Press.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*(7), 309–315.

Munakata, Y., McClelland, J. L., Johnson, M. H., and Siegler, R. S. (1997). Rethinking infant knowledge: toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, *10*(4), 686–713.

Nelson, T. O. and Narens, L. (1990). Metamemory: a theoretical framework and new findings. *The Psychology of Learning and Motivation*, *26*, 125–173.

Newell, B. R. and Shanks, D. (2014). Unconscious influences on decision making: a critical review. *Behavioral and Brain Sciences*, *37*(1), 1–19.

Nisbett, R. E. and Wilson, T. D. (1977). Telling more than we can do: verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

O'Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 883–917.

O'Reilly, R. C. and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press.

Pacherie, E. (2008). The phenomenology of action: a conceptual framework. *Cognition*, *107*, 179–217.

Pacton, S., Perruchet, P., Fayol, M., and Cleeremans, A. (2001). Implicit learning out of the lab: the case of orthographic regularities. *Journal of Experimental Psychology: General*, *130*(3), 401–426.

Pasquali, A., Timmermans, B., and Cleeremans, A. (2010). Know thyself: metacognitive networks and measures of consciousness. *Cognition*, *117*, 182–190.

Perruchet, P. and Vinter, A. (2002a). The self-organizing consciousness. *Behavioral and Brain Sciences*, *25*(3), 297–330.

Perruchet, P. and Vinter, A. (2002b). The self-organizing consciousness: a framework for implicit learning. In R. M. French and A. Cleeremans (Eds.), *Implicit Learning and Consciousness: an Empirical, Computational and Philosophical Consensus in the Making?* (pp. 41–67). Hove, UK: Psychology Press.

Persaud, N., McLeod, P., and Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*, 257–261.

Ramsoy, T. Z. and Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*, 1–23.

Raz, A., Fan, J., and Posner, M. I. (2005). Hypnotic suggestion reduces conflict in the human brain. *Proceedings of the National Academy of Sciences of the U.S.A.*, *102*(28), 9978–9983.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: an essay on the cognitive unconscious*. Oxford, UK: Oxford University Press.

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

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

Rumelhart, D. E. and McClelland, J. L. (1986a). On learning the past tense of English verbs. In J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models* (pp. 216–271). Cambridge, MA: MIT Press.

Rumelhart, D. E. and McClelland, J. L. (1986b). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Sandberg, K., Timmermans, B., Overgaard, M., and Cleeremans, A. (2010). Measuring consciousness: is one measure better than the other? *Consciousness and Cognition*, *19*, 1069–1078.

Schapiro, A. C. and McClelland, J. L. (2009). A connectionist model of a continuous developmental transition in the balance scale task. *Cognition*, *110*(1), 395–411.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., and Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414.

Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge, MA.: MIT Press.

Sergent, C. and Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science*, *15*(11), 720–728.

Seth, A. (2016). The real problem. *Aeon*. https://aeon.co/essays/the-hard-problem-of-consciousness-is-a-distraction-from-the-real-one.

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

Shapiro, K. L., Arnell, K. M., and Raymond, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*, 291–295.

Simons, D. J. and Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*, 261–267.

Timmermans, B., Schilbach, L., Pasquali, A., and Cleeremans, A. (2012). Higher order thoughts in action: consciousness as an unconscious re-description process. *Philosophical Transactions of the Royal Society B*, *367*, 1412–1423.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*(5), 1–14.

Vermeiren, A. and Cleeremans, A. (2012). The validity of d' measures. *PLoS One*, 7(2), e31595.

Weiskrantz, L. (1986). *Blindsight: a Case Study and Implications*. Oxford, England: Oxford University Press.

Windey, B., Gevers, W., and Cleeremans, A. (2013). Subjective visibility depends on level of processing. *Cognition*, *129*, 404–409.

Windey, B., Vermeiren, A., Atas, A., and Cleeremans, A. (2014). The graded and dichotomous nature of visual awareness. *Philosophical Transactions of the Royal Society of London, B., 369*.

Wolpert, D. M., Doya, K., and Kawato, M. (2004). A unifying computational framework for motor control and social interaction. In C. D. Frith and D. M. Wolpert (Eds.), *The Neuroscience of Social Interaction* (pp. 305–322). Oxford, UK: Oxford University Press.

3

CONSCIOUSNESS, LEARNING, AND CONTROL

On the path to a theory

Viktor Allakhverdov, Margarita G. Filippova, Valeria A. Gershkovich, Valeriia Yu. Karpinskaia, Tanya V. Scott, and Natalia P. Vladykina

1. Epistemological introduction

The process of learning is enigmatic. As a result of training, the simplest sensory and motor tasks performance improves: e.g. sensory thresholds decrease, stimulus discrimination improves, reaction time response to a light flash decreases, the time it takes to add single digit numbers decreases as well, the movements become more precise, and so on. Gottlieb and colleagues (1988) summarized these observations: ". . . there is no human movement too simple to be improved through practice" (p. 439). One of the authors of this article, for instance, completed a simple sensory–motor task (playing a game similar to "Tetris") over several years: even after a hundred and ten thousand trials, he kept observing improvements. In figure skating, athletes learn to distinguish micro changes of the skate angle with the sole of the foot (in measurements of several millimetres – known as "the feeling of the skate edge"). This skill usually develops within seven and a half years of daily practice. In all of the learning scenarios, people are unable to explain either what exactly leads to the enhancement of results or what they do for the improvement to occur.

The key question for the theory of learning is: why do people performing the same trivial action several times in a row start to perform it better and faster? Why are we able to do something at the end of the learning process that we had no idea how to do before, without any realization of what it is that we have done to learn it? Mere repetition of actions should not improve the efficacy of action because on the one hand, if we repeat non-effective actions they remain non-efficient, and on the other, if the action changes from one performance to another, then it is not a *repetition of the same* action. (This reminds us of a well-known paradox of scientific knowledge acquisition from the times of antiquity: if a scientist knows what s/he is looking for, then this knowledge cannot be new, and if s/he does not know what s/he is looking for, then what is *it* that s/he is looking for?) Another question is: why can a person attain a result that appears randomly high in the process of learning only to see this result becoming not attainable during many subsequent repetitions?

There exists a myriad of theories that attempt to answer how learning is possible. These theories use hypothetical physiological mechanisms that start to work faster and with more precision as a result of training. Each of such theories has its helpful insights, but each one has flaws, as well. Thus, a widely accepted theory of learning does not exist. Consider the Tetris example from above: sometimes a once obtained high result was impossible to repeat even after 2000 following trials, and then, in the process of learning, this once obtained high result becomes rather average. This most likely means that the physiological mechanisms already know how to act accurately and fast. If this is true, why does one need a thousand repetitions to reach such a result consistently? Moreover, in extreme situations and in altered states of consciousness, a person is able to achieve results that are usually obtained with long and tedious training without any learning at all (e.g. a person is able to walk the wire without losing balance in a somnambulistic state).

Let us try to posit the problem differently. We will start with the statement that even before any learning starts a person is able to do what s/he needs to learn to do. Then, the question is not how and why a person learns but why a person does not exhibit the skills without prior learning. The reverse formulation of a question can be fruitful, as we know from the history of physics. Before Galileo, physicists attempted to understand reasons behind movements of bodies and the forces that made bodies move. Galileo's principle of inertia turned the question upside down: bodies always move when no force is applied. He stated that it is necessary to explain why bodies stop moving, not why and how they move. This was possible when Galileo had to revert to an idealization: an ideal non-existent construct: he had to introduce a material point (bodies do not have measurements: size and weight) and a zero force (no forces are applied to the body). Idealization allowed him to disregard all chance outcomes that muddle and refract the essence of the event. Every idealization is a priori an incorrect description of the reality, but without them no theory can be built in natural sciences.

We are going to follow this example. Here we introduce an idealization: human cognitive abilities have no limits except for those that are limited by the logics of cognition. It is given that cognitive abilities are restricted by certain physical and physiological limitations. Nevertheless, while building a psychological theory, we can neglect them because limitations observable in experiences far exceed them. (Thus the speed of light and speed of the nervous impulse are limited, but these limitations do not define the speed of reading a book.) It is the logics of cognition that determine a complex architecture of brain transformations.

The process of learning is exceptionally complex. On the one hand, some of the algorithms of cognition (and learning as a part of cognition) have to precede the process of cognition, otherwise neither cognition nor learning would ever start. On the other hand, there is no universal algorithm that can optimally resolve all the cognitive problems. It is also impossible to prove this thesis strictly due to the lack of an operational definition of the term "cognition". Most attempts to formalize even some of the aspects of the process lead to the abyss of insolubility. Gödel's incompleteness theorems demonstrate that any formal system in which natural numbers can be defined is always incomplete for it contains statements, truth values of which cannot be determined within the same formal system. That is to say that if we use arithmetic in our system (what system does not use it?!), we will come across problems that we cannot solve. Church and Turing's theorem demonstrates that we cannot predict which problems we would be unable to solve. The universal algorithm to maximally compress the incoming information does not exist either (computation of the complexity of information using Kolmogorov's index is also impossible, it depends on the algorithm used). The list can go on.

Essentially, the problem is that no cognitive system has access to reality in its entirety. Only partial description of reality is possible. Therefore, such description can be erroneous. In the process of cognition, one should store the correct knowledge and modify the incorrect one. If one knew what should come as a result of cognitive action, then one would know what to store and what to modify. But the cognizer cannot match today's knowledge with future knowledge, for the latter is not yet known to him. There is always a choice: either to protect the existing knowledge, or to happily reject it. And there is no means of evaluating the choice in advance.

The goal of cognition is to give a correct description of reality in the known language. But the result itself does not only depend on reality but also on the means of cognition (non-universal), and on the language of the description that has to be chosen before one knows whether it is appropriate to the goal. That is why all obtained results have to be meticulously verified. How does that happen? The philosophy of science postulates several crucial ways of verifying scientific constructs. First, all results should be checked using a different way from the way by which they were obtained (the principle of independent verification). Secondly, conclusions made by one person should be confirmed by another person (intersubjective verification method). Third, all data should be checked for consistency. We believe that such methods of verification should be used by humans within all cognitive processes (after all, science is the best model of cognitive actions). The brain builds psyche and consciousness exactly for the purpose of such verifications. Within the introduced idealization, the brain immediately processes all the incoming information (this statement is consistent with the data obtained in implicit learning experiments), performs the most complex computations, catches patterns, creates hypotheses, manipulates motor commands, and much more. But a person starts to act only after consciousness verifies all these automatic constructs, confirms or corrects them, and only after that would sanction an action. So using Baars' metaphor, consciousness is the brain's project manager.

Thus, mental and conscious acts are determined by the tasks of verification, and physiological mechanisms just ensure the fulfilment of these tasks. Such an approach allows us to resolve the learning paradox postulated earlier. It is not the brain or an organism that learns, but consciousness learns to manage the brain (the organism) in order to use their abilities. In this chapter, we show how such an approach works using empirical data support.

2. Learning: problem definition

If we take a closer look at the process of learning, we see that it is paradoxical. For instance, sensory–motor learning is commonly described in the following way: a person completes the task better by repeating the same action over and over again. But this is absurd! How can one increase the efficiency of an action by merely repeating the same action over and over again? If actions are not effective to begin with their repetition would not make them effective, and if actions are modified, then, they are not merely a repetition. Perhaps, one improves and perfects each action rather than just repeats it. Then, the question arises: how do we know that one ineffective action is better than the next one? And if we do know this, then, why do we start by performing an ineffective action? In order to perform an action, a person needs to apply a conscious effort but sensory–motor learning succeeds because sensory–motor actions are automatic (happening without conscious awareness). In this chapter, we put forward an explanation of the logics behind the process of learning and show that our explanation is barren without understanding of the mechanisms of consciousness.

If a person is able to respond faster, why is s/he not able to do so right away? And if s/he was able to do so by implicit learning, why cannot s/he use this learning explicitly? Why is it that after a warm-up period the actual learning curve (not the averaged and fitted curve shown in textbooks) does not show either a smooth change (as it is shown in behavioristic models), or an abrupt increase in effectiveness (as described in Gestalt models) but rather a series of ups and downs?

The process of memorization that is in many ways similar to the process of learning is no less mysterious. Ebbinghaus (1913 [1885]) established that we store much more information than we can recall. Modern studies claim exceptional capacities of the visual long-term memory. For example, Brady, Konkle, Alvarez, and Oliva (2008) have shown that participants were able to memorize more than 2,500 images of objects after just one exposure. And this is not the limit! If this is true, then why is it necessary to learn and memorize information of smaller volumes so it can be recalled? Why do mnemonic devices (e.g. method of loci) that require more memorization facilitate recall?

Let us turn to the studies of implicit learning that demonstrate that participants are able to find rather complex patterns in the stimuli unconsciously that they cannot report on consciously. Moreover, they determine these patterns between some features of the stimuli that they do not consider valid or noticeable consciously (see Moroshkina et al. in the current volume).

This leads to even more questions: why does behavior connected with stimulus pattern change even when participants are not aware of the pattern? What prevents them from becoming aware of it? We strongly believe that giving merely an isolated description of the nature of the implicit learning in order to describe the nature of these paradoxes would be futile. The key to answering these questions lies in the theories that explain the most central and most mysterious psychological phenomenon – consciousness. As it is consciousness that makes the most critical decisions but it is unclear how it does it. To explain means to take something unclear and unknown and make it something well defined and familiar. When talking about consciousness it is challenging to establish well defined and familiar. Different schools proposed theories of how consciousness works but did not come to an agreement on how to explain consciousness.

In order to talk about the nature of implicit learning, we have to start with the functioning of consciousness. Here, we put forward a model of consciousness proposed by Allakhverdov (1993, 2000) and further developed by his followers at the department of psychology at the State University of Saint Petersburg, Russia. In this model, consciousness is regarded as a pivotal verification tool for actions during the process of cognition. We are convinced that without any learning occurring, a human is able to press a button almost momentarily and perform complex actions, and that our brain is able to perform complex computations. Our research shows that participants can transform dates into the days of the week almost momentarily without performing any conscious calculations (Allakhverdov, 1993) and to distinguish between correct and incorrect responses (Naumenko, 2010). But at the same time humans are not robots who process stimuli and not just calculators that perform complex computations. Any task uses consciousness, and it impacts the performance of this task. Using a metaphor, we can say that consciousness is the brain's general manager that controls the actions of all the departments that report to it. Not all information reaches this general manager but without its resolution no decision is made. We assume: learning does not happen by the organism or the brain but rather it is consciousness that learns to manage both of them so that they can improve performance and act more efficiently.

In the following sections, we discuss these theoretical statements in order to formulate the pathway to understanding the nature of implicit learning, using the logics of how cognition functions introduced above. We believe that enveloping the mechanisms of cognition allows us to understand the nature of consciousness. The theory of learning we consider to be a case of cognition that heavily relies on the theory of consciousness. Thus, here we envelop the understanding of how consciousness operates in order to provide a sound theory of learning.

3. Humans as ideal cognitive systems

3.1. Theoretical statement

Allakhverdov (1993) proposed an idealization that considers humans as ideal cognitive systems with no cognition ability limitations. The brain (as well as the rest of the body) should be considered as perfect (with no limits on the amount of input: speed of its processing, time of its storage, and so on). This, of course, is not entirely true; both the human brain and body have certain limitations. However, accepting such an idealization only means that cognitive limitations that we are faced with in empirical studies come first and foremost from the logics of cognition and are not the result of physical, biological, physiological, or sociological restrictions. This idealization allows us to pose a crucial question: why does a flawless cognitive system require the existence of special mechanisms, namely, the mind and consciousness?

3.2. Discussion

Any idealization is useful because it allows us to consider the actual process in its "pure" form without scattering our attention to some irrelevant and muddled details. And even though an idealization is not true a priori, it permits us to create logical conclusions that can be confirmed empirically. This is exactly why any theory is always an abstract model that describes behavior of imaginary non-existent ideal objects. Examples can be taken from science: physics uses ideal gasses, a mathematical pendulum, a black body, the center of mass that has no dimensions; biology refers to idealized population, and social sciences use Weber's pure types, abstract works of Karl Marx, etc. . . .

Idealizations are sometimes used in psychology, for instance the theory of the brain's radical plasticity (Cleeremans, 2008), or the idea that the brain is the "Bayesian ideal observer" (Geisler and Diehl, 2003; Frith, 2007). Many researchers claimed that information, once received, is stored in memory forever – this is, of course, an idealization (Ebbinghaus, Korsakoff, Janet, Freud, Leontyev, and Penfield among many others). Baars (1997) supposes that limitations of the observable possibilities of consciousness with practically unlimited possibilities of the brain should be explained functionally, in other words, by the logic of how the entire cognitive system works. This idea is similar to our theoretical approach. It is that our idealization includes all of the other idealizations.

3.3. Empirical evidence

While idealization cannot be empirically established, its existence is consistent with the results of many studies. Research strongly confirms that people process information much faster and more precisely unconsciously, and are able to react to a stimulus that is presented with a speed at which sometimes they are not even sure if anything was presented. Even actions that are traditionally considered the prerogative of the work of consciousness (complex decision making, semantic transformations, goal setting, social and moral judgments, among others) are very likely to happen unconsciously first and only later become consciously available (for more discussion see: Moreland and Zajonc, 1982; Kihlstrom, 1990; Lewicki, Hill, and Czyzewska, 1992; Allakhverdov, 1993; Murphy and Zajonc, 1993; Winkielman, Berridge, and Wilbrager, 2005; Allakhverdov et al., 2006; Pessiglione et al., 2007; Hebart, Schriever, Donner, and Haynes, 2014).

In one the studies, Karpinskaia and Agafonov asked participants to solve anagrams after presenting participants with single-image random-dot stereograms (SIRDS) – specifically organized images that look like a random distribution of dots but actually constitute to a 3-D image if looked at under a special angle and refocusing of the gaze (Karpinskaia and Agafonov, 2010). Before an anagram would appear participants were primed by either (a) a correct answer, (b) SIRDS with the correct answer, (c) SIRDS with information irrelevant to the anagram solution, or (d) an empty field. The results indicated that anagrams that were primed with the correct answer (a) were solved the fastest; anagrams that were primed with SIRDS with the correct answer (b) demonstrated second-fastest reaction times results. It took the longest to solve an anagram when participants were primed with SIRDS with irrelevant information (c). In all cases, participants were not aware of what SIRDS represented because they did not know that they could have contained answers, or that one's visual focus had to be changed to decipher the SIRDS.

How is it possible that participants perceive the covert image? It is rather unlikely that they "see" a stereo image unconsciously. However, it is possible to figure out what the image is by the distribution of separate dots (their local density), which allows us to create an envelope curve. Mathematics solves this problem by one of two methods: Gaussian filter and wavelet compression with Gabor's elements (Karpinskaia and Shelepin, 2010). The results of this experiment confirm that participants can perform complex transformations which can be described with such mathematical methods quickly and unconsciously. Sklar et al. (2012) claim that there is evidence that even very complex math operations could be performed unconsciously. There are, however, skeptics of such possibilities (Moors and Hesselmann, 2017).

4. The impact of the mode of cognition on the result of cognition

4.1. Theoretical statement

The mode of cognition precedes the process of cognition. Many methods (in other words, possible algorithms) of cognition exist, but the one and only correct one, conceivably, does not exist. In any case, whichever method of cognition we choose, it is impossible to prove that it is the best one for the unique criterion of effectiveness of cognition does not exist. The result of cognition, then, depends not only on what is being learned but on the chosen method of learning.

In genuine science, it is advisable to eliminate the impact of the method of cognizing of the result. We propose to rely on three means of knowledge verification. First, all statements should be confirmed in a way different from the way they were obtained (independent verification). Secondly, all knowledge should be checked for non-contradiction (consistency). And lastly, scientific models proposed by some people should be checked by other people (intersubjective verification). Since such means of theory confirmation are used in science, it would be logical to assume that human beings as ideal cognitive systems should use similar ways of checking the input. We claim that mind and consciousness are essential for these operations.

4.2. Discussion

Not only was every attempt to formalize a perfect method of cognition futile but it also proved that such a method does not exist. The idea of proof, in truth, usually is based on some kind of a logical trick. It is assumed, for example, that the most perfect algorithm of cognition is able to complete any task and resolve any problem. But if this is true, then, it should be able to solve the following problem: to create a problem that is impossible to solve. Whether or not such an unsolvable problem can be created, there must be a problem that cognitive algorithm cannot solve, and thus would be imperfect. Gödel, the genius of mathematics, provided such a logical trick using the language of arithmetic. He demonstrated that one could use the language of arithmetic to create problems that cannot be solved using the methods of arithmetic.

A conceptual problem lies within this logical dead end. In the process of cognition one acquires both correct and erroneous knowledge. The correct knowledge needs to be stored, while the inaccurate knowledge needs to be either corrected or modified. If one would know the result of cognition at onset, then, one would know what knowledge to store and what knowledge to modify. And only then would it be possible to evaluate the success of the cognition process. But the result of the cognitive process is not known beforehand. There is always a choice: to continue to correct one's notions protecting them from rejection, or on the contrary, to continue rejecting them. In each case, there is no logical means of defining which choice is a better one. One cannot give an unambiguous evaluation of the accuracy of the cognition process results.

The philosophers of the Enlightenment formulated this problem from a different angle (Locke among others). The philosophical premise was that the final results of cognition are already given to consciousness. However, one cannot evaluate the accuracy of these results because one cannot compare notions existing in consciousness with the existing reality since consciousness has only notions but not the reality itself. For instance, how can a person compare his/her understanding of him/herself with him/herself? One only knows what one thinks of him/herself but not who s/he really is. "It is impossible", philosophers exclaimed in despair, "to match an image of an object with an object itself!" Thus, one cannot compare something that is present in consciousness with something that is not.

4.3. Empirical evidence

The historians of science assert that a scientific theory contributes to the scientific progress if (a) it felicitously describes data and (b) prompts new studies and predicts new phenomena. But these are two different criteria that do not allow for evaluation of the success of the theory at the moment of its inception. Therefore, some theories that are actually descriptive of the data are worse than some other theories that become more promising for different reasons. For example, Copernicus strived to describe the motion of planets around the Sun as circular (not what data was suggesting). That led to serious deviations from the elliptical orbits. Earlier theory of Ptolemy's described the movement of the Sun and the planets around the Earth as epicycles which was very approximate but seemed much more data-based. Nevertheless, it was Copernicus' theory that was heuristic and led to the establishment of physics as a science. Yet, during Copernicus' times, how could one evaluate his theory? Nobody realized how brilliant it was.

5. Independent cognition schemas: the basic level

5.1. Theoretical statement

There are at least two independent schemas of cognition that lie at the core of cognition. They act in parallel receiving different inputs, utilizing different algorithms of processing, having different feedback channels, and verifying the results of each of these schema processing. We consider this level of cognition as basic and exclusively physiological.

Let us imagine that one of the schemas receives extraceptive input (from outside stimulation); using its own algorithms it determines all possible patterns of the input, predicts new inputs, and performs verification of the adequacy of the observations to the subsequent observations. Let us say that this schema uses inductive processing. Now imagine that the second schema builds hypotheses (guesses) about the outside world and attempts to check whether somatic changes or motor actions compatible with the hypotheses are possible. Thus, it only receives proprioceptive and intraceptive information. Let us say that this schema is using deductive processing. Even though the proposed schemas seem very probable, they are only used here as an illustration of the key idea: a possibility of existence of two independent schemas of a perfect idealized cognitive system (Allakhverdov and Gershkovich, 2010).

If results of cognition obtained through two independent sources are compatible with each other, then there is hope that these results depend on reality and not on the chosen method of cognition (although even Kant understood that it cannot be a certainty). The independent compatibility check is necessary but is not enough.

5.2. Discussion

Different authors expressed the idea that independent systems of input processing exist. Fodor (1983) insisted that the architecture of cognition consists of a myriad of parallel independent processes that are "cognitively impermeable", and therefore, other systems cannot influence their functioning. Baars (1991) described cognitive architecture as a multitude of relatively independent specialized processes that carry out specific processes. We consider differently organized independent schemas of cognition as non-specialized and available to perform any task.

The proposed approach of two different schemas of cognition opposes approaches that connect sensory cognition and activity regulation into a sequential chain. Following Sechenov, many physiologists and psychologists believed that cognition starts from stimuli, and they thought that the objective is to describe the architecture of the stimulus transformation on the route from a sensory input to a motor output. These ideas were sustained by the fact that the life-support system has a program of innate reactions to certain stimuli. And, of course, it is methodologically convenient to observe an organism's response to stimuli. Is it the only possible route of cognition? A number of researchers (Dewey, Piaget, and others) claimed the opposite: in the beginning was an action. There are many proponents of this approach. In our proposal we combine both approaches: the process of creating sensory models and hypotheses about the world are checked both simultaneously and independently; cognition starts both from reacting to stimuli and by creating hypotheses from these observations.

Pavlov found out that any connections between stimuli could be built within the "organism activity units". He considered this to be his main discovery (Pavlov, 1927). This would have been impossible if sensory processes and motor processes had been connected from the beginning. Only by assuming the independence of sensory and motor processes on the basic level can one explain the arbitrariness of the majority of the higher level sensory–motor connections. A specialist in child psychology Averin (1998) backs our hypothesis of the duality of the sensory and motor systems at the basic level. He notes that in the process of ontogenesis, motor and sensory development happen in parallel and gradually create sensory–motor coordination.

When two different cognitive schemas receive two confirmed results, these results need to be matched against each other. But what is the mechanism behind the matching process?

Below, we show how we resolve this issue.

6. Subjectively experienced signal

6.1. Theoretical statement

Cognitive results are obtained by separate means on the basis of different input, in other words, they are expressed in different languages, and no rules should exist for translating the language of one source into the language of another (otherwise, we cannot talk about their independence). On the one hand, independent schemas of cognition at the basic level should not receive any messages about successful matching of the fragments of one system with the fragments of another system; otherwise if the information is interpretable between different systems, the systems lose their independence. On the other hand, if these systems do not receive any feedback about the results of matching, independent checking is senseless; and therefore, nothing can be modified or corrected.

We assume that the qualitative signal is generated in the higher block working on the matching task, in other words, if the results of the two cognitive mechanisms matched the criteria, the signal of success is sent. The signal, however, does not give information about the results themselves but about how the basic-level system is functioning in general. None of the quantitative evaluations can be transferred to the basic level (for instance, what outcomes should be modified and how they should be modified). If the signal is affirmative, the work of the basic level continues; if the signal is negative, the system will be modified. Modifications can happen in different ways because the signal does not carry any information on what exactly should be modified.

We prefer to equate this qualitative signal with emotion. No matter how one understands emotion, everyone agrees that it plays an evaluative role and is experienced subjectively. Today, no one can answer the question: "how can anything be experienced subjectively?" We, at least, try to narrow this explicatory gap by addressing the question of the origin of a subjectively experienced emotional evaluation, and by stating that it is important to acknowledge that such a signal is necessary for an ideal cognitive system. This should give us a possibility and hope of extrapolating all the richness of the subjective phenomena: the origin of conscious thoughts, the diversity of desires, and our moral principles.

Let us dub the level of cognition where the matching of results from different schemas of the basic level takes place as the "mental level". The subjectively experienced signals emerge at this level. (Our label is conditional here because mentality and consciousness have such diverse and contradictory meanings that any label cannot possibly reflect all meanings.) The results received at the basic level are merged at the mental level. We call the results of such a merger "mental constructs".

6.2. Discussion

In our view, the dilemma of how the results obtained in different schemas that are not interlegible can be checked for compatibility does not have a successful solution either in the modular theories or in the Baars framework of concept of consciousness.

The signaling system that we are proposing considers these signals as not specialized and not task-specific. The system merely sends notifications about the accuracy of the process in general because the process would take place no matter what. In some sense, it can be construed along the lines of the ideas of humanistic psychologists who believe that humans are destined to realize their capabilities. But how do they know that any action is aimed to reach its target? According to Maslow (1970), people rely on the "internal impulse" motivating them to perform an action. According to Rogers (1961), a person experiences a specific "organismic feeling" informing him/her of the accuracy of the actions. Thus, it is assumed that a person should evaluate his/her actions' accuracy without any awareness, and only the final assessment would enter consciousness.

6.3. Empirical evidence

A signal informs us that a mental construct is available at the mental level, in other words, that a certain problem has been solved. This means that every time a participant finds a solution to the problem (completes the task), a subjective signal "the problem has been solved" is received. Without such a signal participants would not know that the problem was solved or a task was completed. This signal precedes the solution itself. Tikhomirov (1969) demonstrated this effect in his studies. Chess masters were solving complex chess problems discussing the steps out loud. The electrodermal activity (EDA) was measured to identify emotional changes that participants experienced during the creative process. He found out that dermal resistance dropped several seconds prior to the announcement of the final solution. Bechara, Damasio, Tranel and Damasio (1997) discovered a similar phenomenon. Several other authors claim that a change in emotional state is an indicator of the problem solving process, and a person is not aware of the cause of that emotional state change (Forgas, 1995; Schwarz, 1990). Negative emotions reflect dissatisfaction with the solution while positive ones send a signal that everything is fine and on track.

Our claim that such a signal is non-specialized is confirmed empirically (for more discussion and support, see Allakhverdov and colleagues, 2015). It is important to note that the existence of such a signal is described by other researchers but due to the requirements of the experimental logistics, participants usually report on the experiences verbally (e.g. feeling of familiarity reported Jacoby and Dallas, 1981; preferences reported in Chetverikov, 2014; and confidence by Koriat, 2012; among others). For us it is crucial to note that participants are not able to report or explain the feeling for the signal we are talking about is not realized and is not specific, and thus can be used in the performance of different tasks (see Mealor, Dienes, and Scott, 2014).

There are situations in which the signal of the solution of *one* problem can be accepted as a signal of the solution of *another* problem. It appears that in such cases the *other* problem will be solved faster because there is no need to wait for the "the problem had been solved" signal.

In one of her experiments, Naumenko (2010) presented participants with two three-digit numbers to multiply for two seconds. The presentations were primed for 30 ms. The primes were either a correct answer, an incorrect answer, or no answer at all. So when the presentations were primed with the correct answer the reaction time decreased even if the responses were incorrect. We explain these results in the following way: a participant momentarily unconsciously completes the calculation; then, the presentation of the correct prime creates the "problem solved" signal. However, because the participant is not aware of the completed calculations, s/he chooses the answer randomly, but the received signal of the solution confirms it, and therefore, decreases the time of response.

Filippova (2009) asked participants to perform two simultaneous tasks presented on a split screen. Participants had to follow changes in the imagery on the left side of the screen while solving simple cognitive tasks (lexical decision, anagrams, masked object recognition) on the right side of the screen. The images presented on the left were ambiguous. Participants were not informed that the images were ambiguous and the majority of them did not become aware of that. First, participants had to identify which meaning they perceived; based on the response the computer program would change the image gradually toward the meaning that was not realized. Filippova compared the response time for the solution of the cognitive tasks before and after participants had become aware of the second meaning (when it appeared on the left). Results showed that as soon as the participant has become aware of the second meaning the solution time for *all* cognitive problems (related and not related to the ambiguous image) decreased compared to the response times before the second meaning had become known). The explanation of the results unfolds from the idea that when a non-realized meaning is found, consciousness receives a "the problem is solved" signal that allows participants to skip an accuracy check of the task at hand, and this leads to faster response times. This would only be possible if the signal "the problem is solved" is non-specific and is not problem- or task-specific. Sometimes though, the unconscious signal occurs during the performance of consciously attended-to transformations, thus, it might be ignored or vice versa can erroneously support as a reinforcement of the accuracy of the transformations (Mealor, Dienes, and Scott, 2014).

7. The mental level of cognition

7.1. Theoretical statement

During the process of matching results at the mental level, only constant unchanging results can be compared. The key methods of processing the information received from the basic level at the mental level are the division of information into the discrete units (fragments) and an attempt to distinguish fragments received from different schemas. We postulate independent schemas are involved in the result matching. But it is important to understand that verification and matching can only happen between non-changing pieces; changing values cannot be matched or verified. And as any dynamic process with feedback, it would always be iterative and in discrete portions – test > operation, test > operation, and so on. It is highly likely that numerous options of how this happens are at work here: from the simplest one proposed by Gestaltists – to look for isomorphic relationship between fragments of two languages with the following verification that they remain isomorphic in the future events – to the more complex mathematical transformations. We are not going to hypothesize how this happens. The mechanism of the processes is, of course, important. But resolving logical problems that arise here is more critical.

Mental constructs can be checked for compatibility independently. Let us call the cognitive level where such verification happens the "level of consciousness". Mental constructs that are compatible to the ones selected previously can enter at this level (more on this in the following sections). The effectiveness of mental level work is evaluated against the demands of the mental constructs required by consciousness. If consciousness does not accept any of the new constructs, it sends a signal to the mind that all the results are incorrect (and consequently of the inaccuracy of the results to the basic level). The mind stops generating positive emotional signals. Conversely, *if consciousness persistently selects just one mental construct, this construct stops coinciding with the emotional signal, and thus becomes unconscious.*

To create mental constructs is to equate the non-identical: some result *S* (in one language) is identical to some result *P* in another: thus, *S* is *P*. But we know that S should be S, and not P. For instance, consider two statements: (1) Sigmund Freud (*S*) is the founder of psychoanalysis (*P*). (2) Sigmund Freud (*S*) is Sigmund Freud (*S*). The first statement (1) refers to the history of psychology, while the second (2) is different from the first and is simply a tautology. Therefore, "*Sigmund Freud*" and "*the founder of psychoanalysis*" point to the same result but have different meanings (in other words, *S* is *P* and simultaneously *S* is not *P*). Russell (1905) clarifies this thought: a statement that "Walter Scott is the author of *Waverley* is different from a statement that "Walter Scott is Walter Scott".

In order to avoid a contradiction, we would say that *S* and *P* at a given moment belong to the same class (or have the same meaning). This also presumes that both *S* and *P* necessarily belong to different classes. Therefore, the psyche/mentality (and consciousness) is able not only to equate the non-identical, but to distinguish the indistinguishable. Vekker (1974) noted that a solitary object is always realized as a representative of a class, and calls this a phenomenon of generalizing. Volume, speed, and threshold limitations on the input, storage, and processing of the information that we observe in experiments are defined by the class to which certain mental constructs belong and were selected by consciousness as belonging to that class.

In the most general form this can be formalized as the following: all constructs that enter from the mental level into consciousness: (1) are necessarily realized as belonging to a certain class (have an assigned meaning); (2) are matched with other members of the class, and no class can have only one member, that is to say each meaning has synonyms; (3) can be assigned to different classes simultaneously, namely every item has a homonym.

7.2. Discussion

If every symbol only had one meaning different from all other symbols, then all definitions should be considered false, reasoning would be impossible, and logics would be paralyzed – this was noticed by Frege (1975) in *Logical Investigations*. (Consider arithmetic: if "2" would only mean "2", then no transformations would be possible. Arithmetic exists because 2 = 1+1, 2 = 3-1, and 2 = 22:11 and many others.)

A symbol cannot be its meaning. The last requirement allows us to avoid paradoxes of self-circulation. Wartofsky (1979) puts forward an analogous requirement: "anything can be taken as a model of anything else if and only if we can sort out the relevant aspect in which one entity is like another, the relevant properties which both have in common" (p.6).

If different statements denote the same meaning, they do not cease being different statements. Linguists formulated a law: potentially any linguistic symbol is simultaneously a homonym and a synonym (Kartsevski, 1965, p. 85). Vygotsky (1986) stated, "The same thought can be expressed in different phrases, the same way as one phrase can serve as an expression of different thoughts." Miller and Johnson-Laird (1976) found that without synonyms a language could not exist as a structure. Gadamer (1986) made an identical statement about homonyms: the core of language, it seems, is created by an ability of words despite their definite meanings to be open to interpretation, that is to say that every word can be interpreted in an array of ways, and this flexibility allows for such enterprise as speech.

7.3. Empirical evidence

Things that we become consciously aware of belong to a certain class. Back in the 19th century, James Cattell (1886) discovered the *word-superiority effect*: a letter presented within a word is recognized faster and with more accuracy than the same letter presented in a random sequence of the same letters. Therefore, perception of a letter depends on *which class it belongs to*: to a class of letters or to a class of letters that constitutes a word. Almost a hundred years later Weisstein and Harris (1974) discovered an *object-superiority effect*: visual detection of line segments increases if these lines create a three-dimensional object. People memorize about seven items of the class (chunks, in Miller's terms), but not symbols; for example, seven letters versus seven words (five-letter words would contain 35 letters, respectively). Our experiments show that even sensory thresholds change due solely to the illusory change of the size of the stimulus, that is to say by assigning the same stimulus to different classes (Karpinskaia, 2010).

We cease to become aware of the constant unchangeable contents. The contents of consciousness in the words of James (1983 [1890]), is a constant flow. It is impossible to be thinking of the same thing for a long time. Many known phenomena confirm the idea that we stop being aware of content that does not change. Participants stop reacting to imagery constant in brightness and color within 1–3 seconds of exposure to it. A constant audio or a constant tactile stimulus of medium intensity ceases to be noticed very quickly. A color background loses its color after a prolonged fixed gaze and becomes gray (the phenomenon of color adaption). Multiple repetition of the same word leads to the loss of its meaning (the phenomenon of semantic satiation). Multiply repeated actions become automatic and are performed without awareness. If you put participants into a situation of sensory deprivation, they would either fall asleep or start hallucinating. If you present participants with a number of symbols and request them to keep them in mind – very quickly some of these symbols would be forgotten: would leave consciousness. (The last example does not explain the nature of forgetting but leads to the idea that in order to talk about forgetting, it is not required to introduce non-testable hypotheses along the lines of broken traces, memory fragility, and so on.)

8. The conscious level of cognition

8.1. Theoretical statement

The main function of consciousness is to eliminate contradictions within conscious experience (cf. Baars 1988, p. 82: "conscious content is internally consistent"). In order to perform this function, consciousness uses a logical apparatus since the requirement for non-contradiction is the requirement of logics. Consciousness connects a mental construct from the mental level to a uniform case that does not contain any contradictions by using elementary logical operators ("and", "or", "not", "if. . . then. . ."). If these new mental constructs are not in agreement with the situation (are contrary to it, or are not compatible with expectations); then, consciousness can: (a) try to resolve contradictions using logical operations so that new contradictions would not surface; (b) reject these new constructs (suppress them), which would give a signal about errors in the mental constructs for the given situation and would request new constructs to be created; (c) choose another situation (quantization for a different situation is a function of the mental level). Thus, at this conscious level it is impossible to find oneself in contradictory situations, although it is normal that situations themselves can be contradictory.

One can purge a contradiction only in a constant situation. This means that while consciousness is working on resolving a contradiction, it does not work with any newly incoming constructs. It can track new contradictions only if it resolves them one at a time, that is to say that consciousness functions successively. A situation can be considered non-contradictory if it can be presented as consequences inferred from several non-contradicting axioms that would explain the origin of all incoming mental constructs; in other words, it is able to create a determinate description of the situation.

Logical connections of the mental constructs built by consciousness enter the mental level to undergo an independent verification of their compatibility at the basic level. If no contradictions occur, the mental level sends an emotional confirmation signal (and vice versa, if they are not compatible, a negative emotional response is sent). In order to maintain independence between the mental level and consciousness, this signal is not bound to specific results of verification; it does not differ from other signals about the success of cognition. This signal can differ in intensity but it merely informs of the completion of the task and does not clarify which task exactly is completed (from numerous tasks that are being solved at the mental level).

8.2. Discussion

Psychological texts are full of maxims, such as: "we only see things that we understand", "the world in our consciousness is distorted unrecognizably". Many psychologists admit that a conscious being has a need to operate in a rational (non-contradictory) world (see, for instance, Poduska, 1980). Since ancient times philosophers note an inescapable desire to describe a world where everything is interrelated and defined. Hume (1748) called this desire a natural instinct that cannot be generated or suppressed by reasoning. This need is clearly manifested in the thinking of primitive people and children. Levy-Bruhl (1978) assures that primitive people believe that everything is interrelated: that there is a reason for raining and successful or unsuccessful hunting and fishing. According to Piaget (1959), a child needs a reason for everything, "no matter what it is worth". Entire schools of research (reasoning attribution, probability prognosis, and so on) empirically prove that people tend to look for and find reasons to relate events that happen by absolute coincidence.

8.3. Empirical evidence

When input is not consistent with the mental constructs accepted by consciousness, it is often pushed out of consciousness or is unconsciously smoothed. A good example for this statement can be found in Hock (2012): inhabitants of a small mountain village remote from civilization never saw planes that were taking off from a nearby airport. They did not even hear the sound of engines because it could not fit into their worldview. Freud used many examples where contradictions and unconscious ambiguities could be suppressed from consciousness (though, we should probably not take his interpretations of these examples seriously). People perceive multiple meanings when looking at ambiguous images or reading polysemantic texts unconsciously but they are aware of only one of these meanings at any given moment. Festinger (1957) and his followers conducted a series of experiments to show the tenacious eagerness of participants to resolve a contradiction (inconsistency) of two "facts" about themselves or a situation without realizing that such a contradiction existed. Logical inconsistency of information, its discrepancy with previous knowledge, urges people to reduce the dissonance (a term that Festinger coined as "cognitive dissonance").

Let consider examples from our studies. Participants (all were college students) were presented with an excerpt from Lermontov's (1840) poem "The Novice" (Duff, 1919) that is well known to any Russian high school student. In it, the protagonist fought a snow leopard that suddenly appeared. The snow leopard (called "*the waste's eternal guest*") behaved either like a wild cat or like no cat would ever behave at all: "*howling, pawing and furrowing up"* the sand with his paw in anger, standing on his rear paws ("*reared right up as people stand"*), tenderly wagging his tail ("*thumping his tail in friendliest wise")*, and so on. Lermontov built his depiction in such a way that no one would notice a contradiction in patterns of this animal's behavior. So, when asked to recall the plot of the poem, most participants suppressed the contradiction and described snow leopard's behavior in terms consistent with a cat's behavior: they would easily remember fragments that are consistent with cat's behavior: only 21% stated that the leopard was standing on his hind paws; . . . . but most suppressed the oxymoron "the eternal guest" (recalled only by 8%), and only 12% recalled it as a locale "the eternal guest of wastes" (Allakhverdov, 2000).

We can provoke participants to make errors. If we ask participants to solve a series of problems simultaneously making confusing statements: "A three gallon jar can fit more than 17 maple leaves. How many leaves can one fit into a three gallon jar?" they tend to respond with the numbers matching the given number no matter how ridiculous this number might be. Then, we compare the answers of these participants with those of the control group that was only asked "How many leaves can one fit into a three gallon jar?" These participants give various answers not related to the number 17, of course (Tversky and Kahneman, 1974; Strach and Mussweiler, 1997 among many others). When we repeat the procedure after three weeks, the experimental group participants' responses match the previously given numbers even more apparently, and they report much higher confidence in their responses (a tendency known when resolving cognitive dissonance).

We observe the same effect when participants make their own spontaneous errors. Once an error is made and a person is unaware of this error, s/he tends to repeat this error. These recurrent errors are made faster and with more confidence than the ones that do not reoccur. Three well-educated but not too experienced typists (in the era of typewriters) were asked to type a text "as fast as possible without paying attention to the quality". In the sixteen thousand typed words errors constituted 3%. The probability of an error appearing in the same words over again was six times higher (Allakhverdov, 1993). In other words, the errors were consistent and protected by an earlier mark of correctness. In another study, Vladykina (2008, 2010) asked participants to differentiate visual and audio stimuli in the interval of uncertainty, where all stimuli seemed practically identical, and participants could only distinguish them by chance. Nevertheless, they were able to distinguish them without even being aware of that. When stimuli were presented to them again participants identified them with the same rate of errors (this is only possible when participants unconsciously distinguish identical and non-identical stimuli).

We observe recurrent errors in everyday experiences. At school in the arithmetic class, we learned that when adding multi-digit numbers we are prone to making errors. Thus, it is important to check the steps, and moreover, the checking should happen in another manner. So for instance, if we were moving from top to bottom, we should reverse the action. But why did we have to do so? Because this is considered an obvious way to detect errors. However, what it means is that if a student added 2 plus 3 and got 6, and did not notice this error, s/he will repeat this error in the same spot. (For more on the subject of recurrent errors, read Kuvaldina's chapter in the current volume.)

9. Positive and negative choice

9.1. Theoretical statement

Identification of a class assumes the existence of class boundaries. A boundary can only be drawn when one knows what lies on both sides of that boundary.

Therefore, a class can be equally determined by members that belong to that class and by the ones that remained outside the boundary, the rejected ones. Consciousness strives not only to confirm existing knowledge but also to continuously reject things that were rejected in the same situation. Let us label what consciousness chose to be aware of (accept) a "positive choice", and use the term "negative choice" for things that are rejected, things that we decided not to be aware of.

Both the positive and the negative choice have aftereffects. Things that were selected to enter consciousness are activated and have a tendency to be processed faster. Things that were rejected by consciousness (or pushed out of consciousness) are also in the state of activation but have different indexation and are marked as things *not* to be aware of in a given context. When context changes, indexation changes; the things that remained unconscious have a higher than chance likelihood to enter consciousness.

Positively chosen items are unstable, because the contents of consciousness cannot remain constant and stable (this point is at length discussed above). Linguists would say that any meaning is intangible. Content is displaced moving along logical transitions from one member of the class to another. Within the same context only negatively chosen items (the rejected ones) remain constant. Therefore, we would say that the negatively chosen (rejected) meanings define the meaning of a given context.

9.2. Discussion

Readiness to repeat a previously performed action or use of an existing solution is characteristic of any cognitive device. It is not specific only to consciousness. We can observe an effect of predictive coding even in neurons. It is particularly imperative in the process of signification: a connection between a sign and a meaning that has been once created has an aftereffect; otherwise, this connection would cease to exist.

Readiness to reject (not choose) an action or a solution repeatedly is, in all likelihood, a feature specific to consciousness' functioning. This readiness is closely linked to the process of signification. A person understands meaning only in opposition to any other meaning. James (1983 [1890]) introduced the law of dissociation by changing the accompanying elements: consciousness first identifies those qualities of an object that distinguish it from similar elements accompanying its presentation. Linguists entertain similar ideas. Things and events are named when they are viewed in opposition to other things and phenomena; when, as linguistics would say, they become elements of the contrastive set. The term "acoustic guitar" appeared only after an electric guitar was created, and WWI received its ordinal number only after the start of the second one, WWII (Fillmore, 1985). We say: the choice to realize a certain meaning of a symbol requires an unconscious rejection of some other meanings, or in other words, it performs an operation that we are calling the *negative choice*.

Within this framework, we are able to describe creative processes as well. A scientist who diligently searches for a solution to a problem and cannot find it would provide us with a good illustration of the logics put forward here. Let us assume that at some point at the mental level a mental construct is being formed that is the solution to the problem. However, this mental construct is not ready to be accepted for a number of reasons, e.g. construct is in contradiction with previous knowledge, or in contradiction with certain norms and ideals (Ptolemy's vs. Copernicus' theory example above). The mental level indexes this negatively chosen solution as an erroneous one in the given context. The more the scientist spends time on the solution of the problem (continues to exist in the same context), the harder it is to overcome the aftereffect of the negative choice (the rejection of this solution). When the context changes, the solution enters consciousness with ease. This process allows for an explanation of the incubation phase and the phenomenon of insight.

Let us turn to one popular illustration of these phenomena. In the 19th century, chemists did not realize the existence of cyclic molecules (a molecule of benzene, for instance). As one of the legends has it, once Kekulé, a chemist, went for a walk while thinking of the composition of benzene (incubation phase) and saw monkeys on their way to the circus. The monkeys were in a circle and in order not to fall they were holding on to each other. "Aha!" – atoms in benzene are connected into a ring – came an insight. We assume that Kekulé already had a mental construct that was consistent with the empirical data: the molecule of benzene is structured as a ring but the construct was rejected as inconsistent with the concepts accepted in chemistry at that time. The more he tried to solve the problem, the harder it was for him to accept a solution that was once rejected in accordance with the aftereffect of the negative choice phenomenon. As soon as he stopped solving the problem and started walking and looking around, the context had changed but the mental construct: "composition – ring" was already active and was ready to enter consciousness. When he looked at the monkeys, this mental construct quickly entered consciousness. It also received a strong confirmation signal that "the problem is solved". The scientist smacked his lips in agitation experiencing a powerful emotional lift but he did not know yet what problem he had solved. But since the problem of benzene composition was of current interest to him, he quickly came to the solution. Such an approach to the creative processes explains the results of experiments conducted by Ponomarev (1967) (see Allakhverdov et al., 2015). His participants had to connect four dots (corners of a square) with three straight continuous lines without taking their pencil off the paper and so that the pencil would return to the original spot. Many participants failed to complete the task in the given time limit. Then, Ponomarev introduced a task that served as a hint to the solution. He, first, taught the participants a simplified checkers game: four pieces were placed on the board. Participants had to capture pieces by starting in one of the corners. In order to do so, they used the exact movement necessary for the solution of the drawing problem. Then, a sheet of tracing paper was placed on top of the same board, and four dots were drawn in place of the pieces. This hint worked only if it followed the attempt at the main drawing problem. A hint, thus, works only if it is given in the process of the solution of the main problem. The construct for the solution must be activated in order for the hint to work. Monkeys holding paws in a circle would not lead Kekulé to the composition of benzene if he were not already working on a solution to that problem (but the solution was negatively chosen).

Following this interpretation of the creative process, Allakhverdov (2001) construes an emotional impact of art. Artworks, according to him, are constructed as follows: an artist creates a purposefully masked contradiction that one cannot be aware of (recall Lermontov's description of the snow leopard). At the same there is a way offered to resolve this contradiction (Lermontov's protagonist is fighting for his freedom not with a particular animal but with a terrible monster that incorporated characteristics of a variety of animals). As soon as the contradiction is resolved, an emotional signal that "the problem is solved" is received. But we do not know which problem is solved because we were not aware of the contradiction initially. Hence, a scientific discovery provides a single emotional arousal, while the works of art provide multiple emotional reactions.

9.3. Empirical evidence

Bardin (1969) demonstrated the impact of accompanying elements on sensory thresholds. He illustrated that a class is defined by those members that belong to it as well as by the elements of information that were rejected. Participants looked at horizontal lines and lines that were inclined to .5, 1, 2, and 3 degrees in the first series of trials, and to 1, 4, 5, and 10 degrees in the second series of trials. In the first series of trials, participants identified a line that was inclined at 1 degree as inclined and the line inclined to .5 degrees as horizontal, and in the second series they identified a line inclined at 1 degree as horizontal.

Allakhverdov first discovered the aftereffects of the negative choice in 1974. He asked three professional musicians with perfect pitch to identify and name six notes played simultaneously in an atonal chord. The notes that were not recognized in one chord had a tendency to remain unrecognized in the following chords if they were presented in the chord immediately following the one in which they were not identified. In contrast, these notes were named erroneously when they were not presented in the chord immediately following the one in which they were not identified but in the later chords (because they were already activated but negatively chosen). Allakhverdov (1993) discovered that when a participant memorized a list of words containing homonyms but was aware only of one meaning of the word (negatively chose the other meaning), then, the probability of recall of these words was lower compared to unambiguous words. In a series of elegant studies by Filippova, it was shown that unawareness of one of the interpretations of an ambiguous figure impairs the completion of the cognitive tasks semantically related to that rejected interpretation (Filippova, 2009; Filippova, 2011; Filippova and Chernov, 2013). The result obtained in these experiments resembles the phenomenon of negative priming (Milliken et al, 1998; Tipper, 2001; Frings, Schneider, and Fox, 2015). It also corresponds to ideas of the necessity to make an unconscious decision.

Tal and Bar (2014) consider the negative priming effect as a marker of the hypothesis about an object. This approach is close to the approach we are postulating here. However, we understand the reasoning behind negative inhibition differently. Tal and Bar (2014) claim that the decision is inhibited not to create interference with the selected solution, while we claim that inhibition is consistent with suppression of the solution not chosen earlier and maintaining the chosen one because when context is changed this can lead to errors. This allows us to give a simple interpretation of the relationship between positive and negative priming (Dehaene et al., 1998; Ortells et al., 2016; Brocher, Koenig, 2016). If a participant receives a rather simple task (lexical decision, recognition, or trivial arithmetic operations), then a prime presented for a brief time would decrease reaction time in the case of correspondence of the prime to the solution (positive priming) and would increase reaction time when the prime does not correspond to the solution (negative priming). We interpret this effect as follows: if a participant solves the problem after the decision about the prime was made (to ignore it – negative priming), and if the problem is resolved before then a positive priming effect would occur (see Kostina and Allakhverdov, 2017 for empirical evidence).

10. Highest levels of consciousness

10.1. Theoretical statement

Since consciousness works on self-confirmation, the constructs accepted by it must be additionally verified. Let us introduce one type of such verification – intersubjective: the consciousness checks its constructions against those of other people. Let us call this level of cognition "a social level". And once again, we immediately run into a logical problem. Such verification would be already needed at the dawn of humanity when language did not yet exist. We need to interact with another person and evaluate whether his/her behavior is in accordance with our assumptions. But if a partner in response behaves in strict compliance with the laws of physics (he was pushed > he falls, etc. . . .) or in compliance with physiological needs (he got food > he eats, etc. . . .), then, such behavior allows only for checking hypotheses about the laws of physics or physiological needs. Such behaviors say very little about a person's inner world. On the one hand, when people verify their hypotheses, they should behave in a way that would not physically influence behaviors of another person; and on another hand, they have to initiate actions that would trigger the feedback of the others.

However, any partner not only is an object but also is an agent in a cognitive process. Partners have to mutually verify hypotheses about the inner world of each other. And they should choose to act in such a way that would not influence their partner's actions. Imagine the following scenario: two people simultaneously start verifying hypotheses about each other and perform some bizarre actions that do not require any response from the other (for instance, start making some nonsense sounds). Since the consciousness of each of them looks for a reason for any action they would think that the absurdity of the partner's behavior is triggered by their own actions. This hypothesis about the reasons is erroneous but as soon as it emerges it becomes true because both persons form the same hypothesis about the relationship of actions, and each one of them would try to confirm it. So, one of them continues his/her bizarre actions so that s/he can confirm the hypothesis about the relationship between the actions while the other one would continue his/her bizarre actions. This, of course, does not give the participants any insights into understanding each other. When these related actions become more sophisticated, we end up with social patterns, e.g. rituals, social norms, and last but not least languages, cultures, and the entire human history.

Confirmation of one's constructs about the world with the constructs of other people happens within and by the virtue of the work of consciousness. Assumptions about other people's worldview must be distinct from other conscious hypotheses (have different indexation).

The emergence of language opens new and very powerful capabilities for verification of one's own hypotheses. Language structures help consciousness to create new descriptions of the world and self and to attempt to create linguistic hypotheses about oneself in the surrounding world. The concept of oneself (self-concept) presupposes a uniform consistent representation. (Naturally, "self" behaves differently in different contexts, but it is the same constant "self" nevertheless.) All of these constructs must be independently verified as well. The final validation of the accuracy of these constructions is a comparison of these language descriptions with one's own behavior and decisions. Thus emerge the highest level of consciousness functioning – the personal level. Ancient sages referred to this level as the pinnacle of cognition – the cognition of self, resulting in self-knowledge.

10.2. Discussion

Social stimuli are processed differently from other stimuli. Likely, mirror neurons play a role here. Koffka (1928) brought attention to the fact that an infant recognizes his/her mother's face at the age of 2 months but is unable to distinguish colors. Meltzoff and Moore (1977) demonstrated that infants at 12–21 days old are able to imitate facial and manual movements of adults (for example, they are able to stick out their tongues in response to an adult sticking their tongue out). At the same time, it is believed that they are unable to distinguish between objects.

If others participate in a cognitive process of problem solving, it impacts all cognitive processes because other people's opinions influence our perception, memory, and thinking. There are many examples in psychological literature that our sensory impressions can be distorted based on the opinions of other people (Moscovici, 1985; Geen, 1989; Chartrand and Bargh, 1999; Nihei et al., 2002).

The personal level is manifested during cognitive problem solving in conscious usage of different strategies, metacognitive techniques, level of ambition, and in the prognosis of the effectiveness of the solution (both in terms of speed and accuracy) as well as in the desire to prove this prediction of efficacy. The prognosis of performance defines the system of quantification of the speed and accuracy that consciousness is selecting from the prepared mental constructs. The prognosis itself is not realized in a quantitative measure (it happens on the mental level, but the expectations for success or failure are realized). The personal level of control is also manifested when participants unconsciously set the error rate for their performance, and thus will have to confirm it in the process of problem solving (see more discussion of this issue in the current volume).

In one of our studies that we described above (Naumenko, 2010), we asked participants to multiply two three-digit numbers in two seconds, and then, choose the correct answer among three answer choices. Participants chose the correct answer in 30% of cases, guessing by chance. However, when the same examples were given with the answer options, participants tended to repeat the choice of the correct answer even though they thought that they were still guessing. Thus the results indicated that participants unconsciously were able to distinguish a correct answer from an incorrect one. Why did they act as if they were guessing the answers? All of them knew that they could not really perform such complex calculations (they were not aware of it). So as soon as consciousness would choose the correct answer, it would lead to contradicting their prediction on their ability of performance, and thus would create a signal for error. As a result, when participants choose from three answer options, the probability of choosing the correct one returns to one of three (probability of chance in a three-option answer choice is .33).

In experimental situations during completion of tasks, participants' personal level of control is manifested in different behaviors: by using jokes (for instance, "it is not enough to know how to solve problems, one has to love it, too"), by refusal to continue working on the task ("I am tired", "I am bored", "let's take a break", etc.), by introducing additional self-instruction ("let me try to solve this problem faster or in an unexpected way"), by requests to the experimenter to compare their performance with the performance of others, and so on. Sometimes it can lead to critical failures in the performance of the task at hand. We observed cases where it would suddenly take adult educated participants 6–7 seconds to add "2+3", and even give an incorrect answer to such a trivial problem. No one knows what exactly participants are thinking about during these moments. Customarily, in real empirical settings psychologists consider such behaviors as statistical noise of the data.

10.3. Empirical evidence

When reporting empirical results, one should keep in mind that any participant specifically controls whether his/her behavior conforms to the experimenter's expectations. Filjaeva and Korovkin (2015) report that when participants solve cognitive tasks involving direct communication with an experimenter, they demonstrate communicative behavioral patterns, using communicative gestures and facial expressions more often than when such communication is absent.

More importantly – the solutions of the cognitive tasks can vary. Vladykina (2010) showed that performance effectiveness in visual sensory tasks increases when an experimenter provides feedback ("correct" – "incorrect") after each answer compared to the situation where the same feedback is given on a computer screen. In her experiment, the task was to compare lines of different lengths to the standard and choose the one that matches the length of the standard. The experimenter was performing a technical task of recording answers and just informing the participants whether their answer was correct or incorrect. However, interaction with a real person (even if it was non-involved) increased not only the proportion of correct answers but also boosted the aftereffect manifestations.

When solving simple cognitive tasks, participants evaluate the effectiveness of their performance beforehand. If one makes a decision to make errors, then errors will be made. To validate the idea that at the personal level of control participants perform tasks in accordance with their pre-set error rate level, consider the following empirical example. Participants that were briefly exposed to a reading on a pointer instrument made a decision on which readings they were going to identify erroneously beforehand. As a result, the reaction time of the very first response on a specific reading indicated how many errors on this particular reading would be made: if a decision was made to identify this reading erroneously through the following trials, the reaction time of the very first response increased (Allakhverdov, 1993; Andrijanova, 2014).

11. Learning and cognitive control

11.1. Theoretical statement

Instructions to a problem always create a contradiction between what is given (the initial state) and what result should be obtained (the solution). There are four types of control that allow moving from the initial state of the problem to the solution. After certain actions/operations are performed on the lower levels, consciousness checks for consistency of the result of these operations with the ones that already exist – *operational control*. Then it checks whether the obtained result is in fact the solution to the given problem – *task control*. This is followed by the verification of consistency of the result to the expectations of the experimenter and/or other people – *social control*. And finally, it checks whether the solution is reached with the level of effectiveness that participants set beforehand – *personal control*. Each type of control happens involuntarily, and only one type of control functions at any given moment (consciousness works successively), and we are aware only of the results of these controls.

11.2. Discussion

Operational control is the slowest of all: there can be many operations. Each operation is successively compared with the result of every completed operation. In case of success, the signal "the problem is solved" occurs. In the absence of that signal, consciousness cannot guarantee accuracy of the solution. We confirm that such comparison does happen by demonstrating that participants tend to reproduce not only the correct answers but their errors as well, without even being aware that errors were made. If operational control confirms the accuracy of a decision, we start using it less often, and possibly do not need to use it at all: as we observe in cases of phenomenal calculations (savant-like) or memorizing ability (also referred to as eidetic memory). In the 1960s, Soviet psychologists were amazed by the chess grandmaster Tolush. He was asked to memorize the position of chess pieces on the board after a brief tachistoscopic exposure. After such exposure Tolush stated that it was absolutely futile to ask him to indicate where pieces were and even how many there were but he was absolutely confident that White were winning in the current position (in our terminology, he knew it without any operational control) (Bekhtereva, 1978 citing Nebylitsyn).

The signal that "a problem is solved" is not problem-specific; and therefore, periodically it is important to control which problem is being solved. This control happens beyond participants' control. If a problem is being multiply solved during a certain period of time participants stop being aware of that problem (for constant material exits consciousness); and consequently, control over that problems happens more and more infrequently which leads to a decrease in the time needed to reach a solution. (This problem would come back to consciousness from time to time for the control itself does not stop existing). Ach (in Liper, 1963) asked participants to perform a simple adding operation. Participants prone to introspection reported that at the beginning of the experiment they were very aware of what their task was, and the more they did it the less they were aware of what exactly they were doing but they continued to perform the adding task with a consistent degree of accuracy.

If we ask participants to perform a task that should *not* be performed (for example, "not to think of Paris"), its completion without errors is unequivocally impossible. As soon as one starts controlling whether the task is carried out correctly ("what is it that I am *not* thinking about?" and "Is it what I am *not* thinking about?"); the thought of Paris, the capital of France, enters consciousness (task control). Now let us ask our participants to perform two tasks simultaneously: something to think about and something *not* to think about (to ignore). For instance, "Please, think about a monkey and do not think about Paris". The more complex the main task, the higher is the involvement of the operational control. Since movement from one type of control to another is successive, then more infrequently the other type of control would take over (task control). Thus the more complex the thought process about the problem in hand, the less thought would go to the things that should be ignored.

A sound example of what role social control plays is a change in how a problem is being solved under a hypnotic suggestion. It appears that under hypnosis it is possible to reduce all types of control except for the social one. As a result participants are able to act automatically without controlling operations and tasks, and therefore, they are able to calculate, memorize, and even play chess better (see Rajkov's 1976 discussion, for instance).

The personal control checks the effectiveness of task completion against the level that consciousness is ready to provide based on the first trials or previous experience. Confirmation of prognosis is only possible for the series of actions. Since the content of consciousness constantly changes, compatibility check results cannot remain constant. If a participant accidentally performs considerably better than the pre-set effectiveness level, consciousness is forced to go over several failed probes in order to match its own prognosis. Learning curves support these ideas. Gradual accumulation of successful probes allows consciousness to confirm that a higher level of performance is possible and adjust the performance prognosis. A change in prognosis allows people to act more effectively.

11.3. Empirical evidence

In our view, the phenomenon of interference illustrates that task control is involuntary. Let us consider the Stroop's effect phenomenon. Participants had to name colors of the font (the *main task*) without reading the color word itself (a *task of ignoring*): when the word "red" was presented printed in "blue", participants' response should have been "blue". The difficulties in performing this task are usually attributed to the fact that a strong automatic reading skill resource dominates a weaker skill of color naming. This explanation does not hold though. Protopapas, Archonti, and Skaloumbakas (2007) reported that children that hardly know how to read experience more difficulties with this task than adults. These difficulties, as we see it, come from a different source: task control. Participants spontaneously verify what task they are performing ("I am not reading the color word 'red', am I?"). According to this logic, the more complex the task at hand the more it would take to control the result and less frequently would task control happen. Thus interference should weaken. In our experiments, the increase of complexity of the main task (e.g. when participants have to state shapes or color tinges, and thus work on an additional task along with the main task), interference does not strengthen but on the contrary weakens (strengthening would follow from all the accounts of the divided resource usage) (Allakhverdov and Allakhverdov, 2015).

Complicating the task control process leads to a less frequent operational control and leads to a more successful learning. Moroshkina (2010) in an experiment using the switch-task paradigm asked participants to alternate addition and subtraction of consecutively presented pairs of one digit numbers (1 to 9). Each number pair would appear without a respective sign of an operation to be performed (either adding or subtracting) so that participants had to keep it in mind. One group of participants (*simple operation alternation* group) had to add the first pair of numbers, and then subtract a smaller number from a bigger number in the following presented pair, then repeat the addition, then the subtraction. The second group (with *complex operation alternation*) had to add the first and the second presented pair of numbers, and then subtract a smaller number from a bigger number in three following presentations, then again add twice and subtract three times consecutively. Evidently, in the group with a *complex operation alternation*, task control on what task is being performed occurred more frequently. As a result, participants of this group (where there is less time for operational control) learned faster and were more effective than participants from the *simple operation alternation* group.

If we modify a certain parameter that is not related to the learning process or to the material to be learned during learning in a predictable regular manner, these predictable modifications register at the basic level and are supported by a positive "a problem is solved" signal. This leads to checking what task is being completed (performing task control) more frequently. This, in turn, leads to a lesser frequency of operational control. In order to confirm this statement, we conducted studies on how modifications to irrelevant predictable patterns affect control processes during completion of the *main task*.

Tukhtieva (2014) formed problem solving sets (*Einstellung*) but modified irrelevant parameters (e.g. changed colors while forming *Einstellung* for the size of Uznadze's circles, and modified presentation method in the Luchin's jar problem by using numbers, texts, drawn jars, and even cartoon effects). She found out that regular modifications to irrelevant parameters reduce the *Einstellung* effect. In our interpretation that would mean a decrease in operational control that is required to check on solutions that were already found. She also discovered that irregular modification to these irrelevant parameters strengthens the set effect.

In the series of experiments, we have demonstrated that changes in parameters irrelevant to the main learning task have an effect on memorization (Allakhverdov et al., 2006). We found out that regular modifications to parameters that were not asked to be recalled (e.g. changes in the color of presented numbers, changes in dashes between numbers, and so on) facilitated memorization of items, and irregular changes resulted in decline in recall. Gershkovich (2010) showed that regular changes to the background on which stimuli for memorization were presented resulted in a smaller number of attempts needed to memorize stimulus items compared to memorization of items that were presented on a non-changing background.

12. Conclusions

Our position is that even before the process of learning, a person implicitly is able to do what s/he is learning to do but is unable to explicitly realize this skill. The proposed paradigm uses an idealization according to which a person's brain momentarily can make complex computations and recognize patterns in the incoming information but does not always realize it and sometimes does not even use these abilities. This idealization supposes that the limitations of cognition that we come across in experiments are not due to physiological or psychological reasons but exclusively to the logics of cognition.

Introduction of such idealization allows us to pose a crucial question: why does an ideal cognitive system require special mechanisms such as psyche and consciousness? We propose that consciousness performs the functions of control of the verification operations and as a result of them sanctions of action. In order to describe independent verification processes, we introduced two independent cognitive schemas. One is sensory and the other one is regulatory, one of which only receives extraceptive information that uses information received from the environment and uses induction; the other receives only intraceptive information and uses deduction. Here, it is important to accept the existence of the two independent systems rather than figure out their workings. Thus, such systems cannot be paralleled to the largely accepted top-down and bottom-up processing. We concentrate on the fact that the systems are not talking to each other but rather provide a set of independent verification results. Moreover, here we talk about a very early process of matching that we call "the basic" level of verification. Such an approach is not identical to O'Regan and Noë's (2001) sensory–motor approach to perception, who state that perception relies on mastery of sensory–motor dependencies. On the contrary, we propose a strictly independent sensory system of formulating hypotheses about the world verified by actions. Thus, sensory and motor schemas are built simultaneously and independently from each other. We consider these differently organized systems of cognition to function in parallel and independently in performing all tasks and not only in visual perception.

The level of cognition where the basic verification takes place, we called the "psychic" level because subjectively realized signals are being created at this level.

Ideas that higher levels of information processing are slower and more discrete acting under "all or nothing" principles, and that the lower levels are faster and probabilistic, can be found in other contemporary works (see Charles et al., 2013, Tal and Bar 2014, among others). Dennett (1991), for instance, mentions discreteness of consciousness in its seemingly continuous actions.

The next step of working with the information is verification for consistency that happens after the basic verification level is complete. We use the term "conscious" level of verification. The mental constructs that do not contradict previously selected ones are at work on this level. Here, we can find parallels with other theories that state that cognition happens as a result of stating hypotheses and/or predictions (see Bartlett, 1932; Bruner 1957, Enns and Lleras, 2008; Gregory, 1997; Hohwy, Roepstorff, and Friston, 2008; Hohwy, 2013; Neisser and Becklen, 1975; and Panichello, Cheung, and Bar, 2013). According to Friston (2012), only contradictory information moves to the next level, such contradictory information is referred to as "erroneous predictions". Thus, it is required so that the inconsistencies are explained. Our approach is similar in this sense. If the construct is selected, it stops receiving emotional indexing and thus stops being realized.

In our approach, not only does consciousness strive to confirm the existing knowledge but also to reject the knowledge that was once rejected in the same circumstances. The *positive choice* is coined for items accepted by consciousness, and the *negative choice* is everything that has been rejected. The logics behind the latter are the following: choosing a class supposes boundaries of the class to exist.

The boundary can only be determined by stating what lies on both sides of it: thus the boundary is defined by everything that belongs to the class as well as by everything that does not (in some sense the process of inhibition can be used as a parallel here). This view is somewhat similar to Dehaene and Changeux (2011) and Tal and Bar (2014), who consider consciousness as a mechanism of focusing on the most appropriate interpretation of reality obtained after multiple computations. They also connect inhibition to the single interpretation of the conscious experience.

The highest levels of cognition in our approach are social and personal. On the social level the created constructs of reality are matched to the constructs of others. We suppose that understanding of how others perceive the world has a different marking compared to other hypotheses about the world. And the last key verification happens on the personal level where verification is based on the matching of created descriptions about the world with the understanding of oneself and compared to one's behavior. Such verification happens on the highest level of cognition.

Thus consciousness relies on different levels of controls that perform level-specific operations and, in turn, uses the positive and negative choice to keep information in check and assign meanings, and learn about the world around us. In multiple theories that explain the work of consciousness, the conclusions are similar. The main difference with the current approach is to present the logics of the appearance of the signal of the correct vs. incorrect result of information processing at the unconscious level, and the necessity of marking such a signal in a special way to move it into the conscious level. Baars (2011) not only justifies the necessity of functional explanation of limitations observed during conscious work with the information but also the idea that consciousness works on eliminating contradictions. He also concludes that information from consciousness can be spread among all structures. But why conscious decision is necessary remains unclear. Baars asserts that a state is realized if it can globally impact memory or other cognitive activity. We postulate something different: information that is supported by the positive emotional signal becomes available to consciousness, and after that consciousness verifies the information for consistency and contradictions, and performs intersubjective verification. Cleeremans (2008, 2011) and Pasquali et al. (2010) using Rosenthal's (2004) model use the first level neural network mechanism that explains the unconscious cognition well. Cleeremans considers consciousness as a second level neural network that studies and describes the work of the first level network. We find this idea to be profound but the question still remains: where does awareness in the second level network come from? Let us assume that a person knows something but also is aware of how and why this knowledge is available. The key idea, in our view, is that of an independent verification by means of which knowledge is acquired autonomously of the means of cognition. Therefore, where Cleeremans considers that one first level learning loop is sufficient, we are postulating the existence and necessity of two loops. We are convinced that the question of the role of consciousness and awareness is key to explaining the phenomenon of implicit learning.

13. Consequences of the proposed model for implicit learning

I All the patterns and rules of the stimuli, i.e. grammar rules, structures, sequences, are detected on the basic level. The idealization converts the question of "why does implicit learning occur?" to the question: "why are there cases where we do not observe it?" We assume that implicit learning does not occur when conscious control interferes with the work of unconscious processes.

II Patterns that are determined on the basic level can generate a mental construct that leads to the "problem is solved" signal sent to consciousness.

III When participants have to become aware of patterns at the social level, the patterns are often perceived as undistinguishable, thus effectiveness of the completion of the task is set relatively low.

IV Consciousness tries to guess what mental construct is compatible with the problem's solution (since signals are not problem-specific); however, it is not able to perform operational control (unless it finds a solution obtained earlier, which it can compare this mental construct with).

V By performing a time-consuming gradual search, consciousness can find logical relationships and patterns but due to the personal level-set limitations and instructions to a problem, consciousness forms random connections and relies on them heavily rather than relying on the emotional signal about the solution.

VI The more a given situation requires verbalizations and explanations of consciousness' decisions, the more consciousness would rely on its formations rather than rely on the "problem is solved" signal. (See a more detailed discussion of this issue in the current volume by Moroshkina and others.)

VII To see the functioning of consciousness in the process of implicit learning we have to base our findings not on the subjective criteria of awareness but to show the effects of the negative choice. As it was demonstrated above, the manifestations of these effects are distinctive of consciousness.

Therefore, the model of consciousness proposed here states that explanation of the function of consciousness is detrimental to understanding the nature of implicit learning. Thus, coming from the ideas expressed here, implicit learning itself should not be surprising or mysterious. This type of learning is called "implicit" because we are unaware of the fact that learning is happening – and this fact *is* surprising. An explanation of the learning effects is first and foremost an explanation about the role of consciousness. According to the approach introduced here, it is not the brain or the organism that is learning, but consciousness. Consciousness learns to manage the brain and the organism in order to extract the information generated by them. This statement should not be interpreted as suggesting that consciousness is an obstacle to cognition and learning. Quite the opposite is true. It is only due to consciousness that participants are able to solve problems, perform tasks, and react to stimuli. Otherwise, they would be staring at the screen with a nonsense sequence of symbols, would not press buttons in response to the flash of light, would not state their confidence of the accuracy of the solution to the problem, and so on. And since consciousness sanctions the solution of the given tasks, it uses the full power of its control.

Acknowledgements: Preparation of this chapter was facilitated by a grant from the Russian Foundation for Basic Research (RFBR) # 17-06-00473

References

Allakhverdov M. V. and Allakhverdov V. M. (2015). O chem prosche ne dumat'?: O prirode Strup-interferencii. *Shagi/Steps*, 1, 122–137.

Allakhverdov V. M. (1993). *Opyt teoreticheskoi psikhologii*. Saint Petersburg: Pechatnyj dvor.

Allakhverdov V. M. (2000). *Soznanie kak paradox*. Saint Petersburg: DNK.

Allakhverdov V. M. (2001). *Psikhologiya iskusstva*. Saint Petersburg: DNK.

Allakhverdov V. M. and Allakhverdov M. V. (2014). Fenomen Strupa: interferentsiya kak logicheskii paradox. *Vestnik Sankt-Peterburgskogo universiteta*. 16(4), 90–102.

Allakhverdov V., and Gershkovich, V. (2010). Does Consciousness Exist? In What Sense? *Integrative Psychological and Behavioral Science*, 44, 340–347.

Allakhverdov V. M., Agafnov A. J., Vishniakova E. A., Volokhonskii V. L., Voskresenskaia E. J., Gershkovich V. A., Ivanov M. V., Ivanova E. N., Ivanova N. A., Karpinskaia V. J., Kuvaldna M. B., Ledovaya J. A., Moroshkina N. V., Naumenko O. V., Sergeev S. F., and Filippova M. G. (2006). *Eksperimental'naya psikhologiya poznaniya: kognitivnaya logika soznatel'nogo i bessoznatel'nogo*. Saint Petersburg: Izdatel'stvo Sankt-Peterburgskogo universiteta.

Allakhverdov V. M, Gershkovich, V. A., Karpinskaia V. J., Morohkina N. V., Naumenko O. V., Tukhtieva N. H., Filippova M. G. (2015). Evristicheskii potentsial koncepcii J.A.Ponomareva. *Psykhologichekii journal/Psychological Journal*. 36(6), 24–34.

Ananiev B. G. (1962). Formirovanie odarennosti. In V. N. Mjasishhev (Ed.). *Sklonnosti i sposobnosti* (pp.15–37). Saint Petersburg: Izdatelstvo Sankt-Peterburgskogo universiteta.

Andrijanova N. V. (2014). Ustojchivye oshibki v processe nauchenija: osobennosti i vozmozhnosti prognozirovanija. *Vestnik Sankt-Peterburgskogo universiteta*, 16(4), 124–131.

Averin V. A. (1998). *Psikhologija detej i podrostkov*, second ed. Saint Petersburg: Izdatel'stvo Mikhajlova V. A.

Baars B. (1988). *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.

Baars B. J. (1991). A curious coincidence? Consciousness as an object of scientific scrutiny fits our personal experience remarkably well. *Behavioral and Brain Sciences* 14(4), 669–670.

Baars B. J. (1997). *In the Theater of Consciousness*. New York, NY: Oxford University Press.

Baars B. J. (2011). *Cognitive Theory of Consciousness*. San Diego: The Neurosciences Institute.

Bardin K. V. (1969). Struktura priporogovoj oblasti. *Voprosy psixologii*, 4, 34–44.

Bartlett F. (1932). *Remembering: a Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Bechara A., Damasio H., Tranel D., and Damasio A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275, 1293–1295.

Bekhtereva N. P. (1978). *The neurophysiological aspects of human mental activity*. New York: Oxford University Press.

Brady T. F., Konkle T., Alvarez G. A., and Oliva A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Science*, 105(38), 14325–14329.

Brocher A. and Koenig J. P. (2016). Word meaning frequencies affect negative compatibility effects in masked priming. *Advances in Cognitive Psychology*, 12(1), 50–67.

Bruner Jerome S. (1957). On perceptual readiness. *Psychological Review*, 64(2), 123–152. http://dx.doi.org/10.1037/h0043805.

Cattell J. (1886). The time taken up by cerebral operations. *Mind*, 11, 277–282, 524–538.

Chartrand T. and Bargh J. (1999). The Chameleon Effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910.

Charles, L., Van Opstal, F., Marti, S., and Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* 73, 80–94. doi: 10.1016/j.neuroimage.2013.01.054.

Chetverikov A. (2014). Warmth of familiarity and chill of error: affective consequences of recognition decisions. *Cognition and Emotion*, 28(3), 385–415. http://doi.org/10.1080/02699931.*2013*.833085.

Cleeremans A. (2008). Consciousness: the radical plasticity thesis. *Progress in Brain Research*, 168, 19–33.

Cleeremans A. (2011). The Radical Plasticity Thesis: how the brain learns to be conscious. *Frontiers in Psychology*, 2, 1–12.

Dehaene S. and Changeux J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.

Dehaene S., Naccache L., Le Clec H. G., Koechlin E., Mueller M., Dehaene Lambertz G., et al. (1998). Imaging unconscious semantic priming. *Nature*, 395(6702), 597–600.

Dennett, Daniel (1991). *Consciousness Explained*. London: Allen Lane, The Penguin Press.

Duff J. D. (Ed.) (1919). *Lermontov's Novice (Mt͡syri)* (Russian text, accented) 1839. Russian text with introduction, notes, and vocabulary by J. D. Duff. (Translated as The Novice by E. W. Morgan. Scotland: E. W. Morgan, 1975.) Cambridge: Cambridge University Press.

Ebbinghaus H. (1913 [1885]). *Memory: a Contribution to Experimental Psychology*. New York: Teachers College, Columbia University.

Enns J. T. and Lleras A. (2008). What's next? New evidence for prediction in human vision. *Trends Cogn Sci*. 12(9), 327–333. doi: 10.1016/j.tics.2008.06.001.

Festinger L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Filippova M. G. (2009). Osoznavaemye i neosoznavaemye komponenty vospriyatiya mnogoznachnykh izobrazhenii. *Psikhologicheskie issledovaniya: Sbornik nauchnykh trudov*, 7, 73–91. Samara: University Group.

Filippova M. G. (2011). Does unconscious information affect cognitive activity? A study using experimental priming. *The Spanish Journal of Psychology*. 14(1), 20–36.

Filippova M. G. and Chernov R. V. (2013). Psikhologicheskie i psikhofiziologicheskie korreljaty vospriyatiya dvoistvennykh izobrazhenii. *Vestnik Sankt-Peterburgskogo universiteta*, 12(2), 21–33.

Filjaeva O. V. and Korovkin S. J. (2015). Obektivnye markery insaitnogo resheniya. In S. S. Belova, A. A. Grigorjev, A. L. Zhuravlev, E. A. Lapteva, D. V. Ushakova, M. A. Holodnaya (Eds.), *Tvorchestvo: nauka, iskusstvo, zhizn': Materialy Vserossiiskoi nauchnoi konferencii, posvyashchennoi 95-letiiu so dnya rozhdeniya Ja. A. Ponomareva*. (pp. 367–370). IP RAN, 24–25 September 2015. Moscow: IP RAN.

Fillmore C. J. (1985). Frames and the semantics of understanding. *Quaderni di semántica*, 6(2), 222–254.

Fodor J. A. (1983). *The Modularity of Mind: an Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press.

Forgas J. P. (1995). Mood and judgment: the affect infusion model (AIM). *Psychological Bulletin*, 117(1), 39–66.

Frege G. (1975). *Logical Investigations*. Oxford: Blackwell.

Frings C., Schneider K. K., and Fox E. (2015). The negative priming paradigm: an update and implications for selective attention. *Psychonomic Bulletin & Review*, 22, 1577–1597. doi:10.3758/s13423-015-0841-4.

Friston K. J. (2012). Prediction, perception and agency. *Int. J. Psychophysiol.* 83(2), 248–252.

Frith K. (2007). *Making up the Mind: How the Brain Creates Our Mental World*, first edition. Hoboken, NJ: Wiley-Blackwell.

Gadamer H.-G. (1986). *The Relevance of the Beautiful and Other Essays*. Cambridge: Cambridge University Press.

Geen R. G. (1989). Alternative conceptions of social facilitation. In P. B. Paulus (Ed.), *Psychology of Group Influence*, second edition (pp. 15–51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Geisler W. S., and Diehl R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, 27(3), 379–402.

Gershkovich V. A. (2010). Vlijanie uslozhnenij figuro-fonovykh otnoshenij pri pred-javlenii celevogo stimula na process ego zauchivanij. In V. A. Barabanshhikov (Ed.), *Jeksperimental'naja psikhologija v Rossii: tradicii i perspektivy* (pp. 368–372). Moscow: IP, Russian Academy of Sciences.

Gottlieb G. L., Corcos D. M., Jaric S., and Agarwal G. C. (1988). Practice improves even the simplest movements. *Experimental Brain Research*, 73(2), 436–440.

Gregory R. L. (1997 [1966]). *Eye and Brain: the Psychology of Seeing*. London: Weidenfeld and Nicolson; 5th edition 1997, Oxford University Press/Princeton University Press.

Hebart M. N., Schriever Y., Donner T. H., and Haynes J. -D. (2014). Fueling doubt and openness: experiencing the unconscious, constructed nature of perception induces uncertainty and openness to change. *Cognition*, 137, 1–8.

Hinde R. A. (1966). *Animal Behavior*. New York: McGraw-Hill.

Hock R. (2012). *Forty Studies That Changed Psychology*, seventh edition. New York, NY: Pearson.

Hohwy J. (2013). *The Predictive Mind*. New York: Oxford University Press.

Hohwy J., Roepstorff A., and Friston K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, 108(3), 687–701

Hume D. (1748). *Philosophical Essays Concerning Human Understanding*. London: A. Millar.

James W. (1983 [1890]). *The Principles of Psychology*. Cambridge, MA: Harvard University Press.

Jacoby L. L. and Dallas, M. (1981). On the relationship between autobiographical and perceptual learning. *Journal of Experimental Psychology: General*. 110: 306–340. doi:10.1037/0096-3445.110.3.306 нас срр 10 .

Karpinskaia (Karpinskaya) V. J. and Agafonov. A. J. (2010). Pomogaet li podskazka, esli ona ne osoznaetsa? Rezul'taty issledovaniya praiming-effektov. *Izvestiya Samarskogo nauch-nogo centra RAN*, 12(3), 90–93.

Karpinskaia (Karpinskaya) V. J. and Shelepin J. E. (2010). Neosoznavaemoe vosprijatie avtostereografishcheskikh izobrazhenii. *Eksperimental'naya Psikhologiya*, 3(3), 57–65.

Karpinskaia (Karpinskaya), V. Y. and Vladykina, N. P. (2010). Decision making regarding conscious and unconscious perception in detection and discrimination tasks. *Journal of Russian and East European Psychology*, 48(3), 33–51.

Kartsevskij, S. (1965). Ob asimmetrichnom dualizme lingvisticheskogo znaka. In V. A. Zveginceva (Ed.), *Istoriya yazykoznaniya XIX–XX vv. v ocherkakh i izvlecheniyakh*. Moscow: Prosveshchenie.

Kihlstrom J. F. (1990). The psychological unconscious. In O. P. John, R. W. Robins, and L. Pervin (Eds.), *Handbook of Personality: Theory and Research* (pp. 445–464). New York: Guilford.

Koffka K. (1928). *The Growth of the Mind: an Introduction to Child Psychology*. London: Kegan Paul, Trench, Trübner & Co.

Koriat, A. (2012).The self-consistency model of subjective confidence. *Psychological Review*. 119(1), pp. 80–113.

Kostina D. I. and Allakhverdov V. M. (2017). Negativnyi prayming-effekt kak proyavlenie posledeistviya negativnogo vybora. *Peterburgskiy psikhologicheskiy zhurnal*, 17, 69–103.

Levy-Bruhl L. (1978 [1922]). *Primitive Mentality*. New York, NY: AMS Press.

Lewicki P., Hill T., and Czyzewska M. (1992). Nonconscious acquisition of information. *American Psychologist*, 47(6), 796–801.

Liper P. (1963). Poznavatel'nye process. In S. S. Stevens (Ed.), Jeksperimental'naja *Psikhologija*, 2, pp. 301–302. Moscow: Inostrannaja literatura.

Maslow A. H. (1954/1970). *Motivation and Personality*, second edition. New York: Harper.

Mealor A., Dienes Z., and Scott R. B. (2014). Unconscious sources of familiarity can be strategically excluded in support of conscious task demands. *Psychology of Consciousness: Theory, Research, and Practice*, 1(3), 229–242.

Meltzoff A. N. and Moore M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, New Series, 198(4312), 75–78.

Miller J. and Johnson-Laird P. (1976). *Language and Perception*. Cambridge, MA: Harvard University Press.

Milliken B., Joordens S., Merikle P. M., and Seiffert A. E. (1998). Selective attention: a reevaluation of the implications of negative priming. *Psychological Review*. 105(2), 203–229.

Moors P. and Hesselmann G. (2017). A critical reexamination of doing arithmetic nonconsciously. *G. Psychon Bull Rev*. doi:10.3758/s13423-017-1292-x.

Moreland R. L. and Zajonc R. B. (1982). Exposure effects in person perception: familiarity, similarity, and attraction. *Journal of Experimental Social Psychology*, 18, 395–415.

Moroshkina N. V. (2010). Conscious control of inattention in task-Switching. *Journal of Russian and East European Psychology*, 48(3), 81–95.

Moscovici S. (1985). Social influence and conformity. In G. Lindary and E. Aronson (Eds). *The Handbook of Social psychology*, third edition. Volume 2, pp. 347–412. New York: Random House.

Murphy S. T. and Zajonc R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 64(5), 723–39.

Naumenko O. V. (2010). Proyavlenie kognitivnogo bessoznatel'nogo pri reshenii vychislitel'nykh zadach [Cognitive unconsciousness in solving calculating tasks]. PhD thesis. Saint Petersburg.

Neisser U., Becklen R. (1975). Selective looking: attending to visually specified events. *Cognitive Psychology*, 7, 480–494.

Nihei Y., Terashima M., Suzuki I., and Morikawa S. (2002). Why are four eyes better than two? Effects of collaboration on the detection of errors in proofreading. *Japanese Psychological Research*, 44 (3), 173–179.

O'Regan J. K. and Noë A. (2001). The sensorimotor contingency theory: toward a new account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939–1031.

Ortells, J. J., Noguera, C., Álvarez, D., Carmona, E., and Houghton, G. (2016). Individual differences in working memory capacity modulates semantic negative priming from single prime words. *Front. Psychol*. 7, 1286.

Pasquali A., Timmermans B., and Cleeremans A. (2010). Know thyself: metacognitive networks and measures of consciousness. *Cognition*, 117, 82–190.

Panichello N., Cheung O., and Bar M., (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, 3, 620. doi: 0.3389/fpsyg.2012.00620.

Pavlov I. P. (1927). Conditioned Reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex. London: Oxford University Press.

Pessiglione M., Schmidt L., Draganski B., Kalisch R., Lau H., Dolan R. J., and Frith C. D. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science*, 316, 904–906.

Piaget J. (1959 [1923]). *The Language and Thought of the Child*, third edition. London, New York: Psychology Press.

Poduska B. (1980). *Understanding Psychology and Dimensions of Adjustment*. New York: Mcgraw-Hill.

Ponomarev J. A. (1967). *Znanie, myshlenie, umstvennoe razvitie*. Moscow: Prosveshchenie.

Protopapas A., Archonti A., and Skaloumbakas C. (2007) Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, 54(3), 251–282.

Psychologia sporta vyshikh dostigenii (1979). Moscow: Fizkultura i Sport.

Rajkov V. L. (1976). O vozmozhnosti uluchsheniya zapominaniya v gipnoze. *Novye issledovaniya v psikhologii*, 1, 15–19.

Rogers C. (1961). *On Becoming a Person: a Therapist's View of Psychotherapy*. Boston: Houghton Mifflin Company.

Rosenthal David M. (2004). Varieties of higher-order theory. In R. J. Gennaro (Ed.) *Higher-Order Theories of Consciousness*, pp.19–44. Amsterdam: John Benjamins Publishers.

Russell B. (1905). On denoting. *Mind*, New Series, 14(56), 479–493.

Schwarz N. (1990). Feelings as information: informational and motivational functions of affective states. In E. T. Higgins and R. M. Sorrentino (Eds.), *Handbook of Motivation and Cognition: Foundations of Social Behavior*, Volume 22, pp. 527–561. New York, NY, US: Guilford Press.

Sklar A.Y., Levy N., Goldstein A., Mandel R., Maril A., and Hassin R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences*, 109(48), 19614–19619.

Strach F. and Mussweiler T. (1997). Explaining the enigmatic anchoring effect: mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73(3), 437–446.

Tal A. and Bar M. (2014). The proactive brain and the fate of dead hypotheses. *Frontiers in Computational Neuroscience*. 8:138. doi: 10.3389/fncom.2014.00138.

Tikhomirov O. K. (1969). *Struktura myslitelnoi deyatelnosti cheloveka*. Moscow: Izdatel'stvo moskovskogo universiteta.

Tipper S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *The Quarterly Journal of Experimental Psychology*: Section A, 54(2), 321–343.

Tukhtieva N. H. (2014). Vliyanie tipov izmeneniya irrelevantnykh parametrov zadach na effekt ustanovki. *Vestnik Sankt-Peterburgskogo universiteta*, 12(3), 41–48.

Tversky A., and Kahneman D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, New Series, 185(4157), 1124–1131.

Vekker L. M. (1974). *Psikhicheskie process, Oshchushchenie i vospriyatie*, 1. Leningrad: Izdatelstvo Leningradskogo universiteta.

Vladykina N. P. (2008). O zakonomernostyakh raboty soznaniya v zone nerazlicheniya. *Vestnik Sankt-Peterburgskogo universiteta*, 12(2), 117–122.

Vladykina N. P. (2010). Reshenie sensornykh zadach v zone nerazlicheniya pri nalichii obratnoi svyazi. Proceedings of the Fourth International Conference on Cognitive Science, 22–26 June 2010, 1, 192–193.

Vygotsky L. (1986 [1934]). *Thought and Language*. Cambridge, MA: MIT Press.

Wartofsky M. W. (1979). Pictures, representation, and the understanding. In R. S. Cohen, and M. W. Wartofsky (Eds.), *Models: Representation and the Scientific Understanding, Boston Studies in the Philosophy of Science*. Volume 48. pp. 175–187. Dordrecht: Springer Science & Business Media.

Weisstein N., and Harris C. S. (1974). Visual detection of line segments. *Science*, 186, 427–435.

Winkielman P., Berridge K. C., and Wilbrager, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, 31(1), 121–135.

4

IMPLICIT LEARNING FROM ONE'S MISTAKES

The negative choice aftereffect

Maria Kuvaldina, Andrey A. Chetverikov, Alexandr S. Odainic, Margarita G. Filippova, and Natalia V. Andriyanova

Introduction

Imagine you are writing an e-mail to your friend and typing very quickly. While you are checking for the overall meaning of the message you probably become less attentive to the grammar and make a couple of mistakes. They are corrected automatically by the text editor and you might not notice the correction. Next time you type similar words you might notice that you are repeating the same mistake: for example, typing "spychology" instead of "psychology". You smile, correct yourself and forget about it. To your great surprise you might find yourself repeating the "spychology" mistake long after you thought you initially caught it, which makes it a recurring error. Somehow a certain stimulus and a certain erroneous response to this stimulus stick together and prevent us from correcting the error. How does this situation come to be? Does it depend on overall performance, that is, do recurring mistakes happen only when we perform poorly in the task, or is it possible for such errors to continue to occur even when overall performance is good?

Dunlap (according to Yates, 1959) described three models that could be used to explain the occurrence of recurring mistakes even in the presence of correct answers. The "Alpha" model states that a response to a given stimulus pattern increases the probability that on the recurrence of the same stimulus pattern, the same response will occur. The "Beta" model states that there is no connection between the probability of stimulus occurrence and the response to it. The "Gamma" model, in contrast to the Alpha model, states that certain responses decrease the probability of future co-occurrence of the stimulus–response association (Peak, 1941; Yates, 1959). If we consider these three models in an attempt to account for recurring mistakes, then the Gamma model tells us that every time we make a mistake, we are trying to correct our behavior and switch to the other response. In this case no recurring mistakes are possible. The Beta model disconnects the stimulus from the response, so any type of co-occurrence might be coincidental. Taking this into account, the proportion of recurring mistakes must be equivalent to the proportion of non-recurring mistakes in the task. The Alpha model describes a situation where a correct response, once given, will be repeated. At the same time, a mistake, once made, will also tend to be repeated, as shown in the example above. The Alpha model essentially states that we memorize not only correct responses, but also erroneous responses. The mandatory association of both errors and correct responses to a given stimulus allows for a high level of performance together with a certain level of recurring errors in the same task. For example, the Alpha model makes it possible to explain why one would persistently misspell some word like "psychology" over and over again even though one is otherwise proficient at writing. This is the perspective we will use in this chapter.

With the Alpha model in mind, we suggest that *recurring mistakes happen because we learn to make them*. One might consistently misspell "spychology" for "psychology"[1] because one has probably learned a certain manner of typing, or an irrelevant representation of the task, an "undesirable pattern of response" (Yates, 1959), or a "nonrepresentative rule" (Reber, 1989). Reber points out that "if subjects emerge from the learning phase with rules (either explicit or implicit) that are not accurate reflections of the grammar, this knowledge base will consistently lead them to misclassify particular items" (Reber, 1989, p.227). Recurring errors result from explicit or, which is more plausible, implicit learning of irrelevant or distracting information that might either be present or not in the stimulus sequence.

This is, while somewhat paradoxical or ironic, overall not surprising. There are many examples showing that people automatically memorize distracting or irrelevant information. For example, in a visual search, observers remember the features of distracting stimuli (Chetverikov, Campana, and Kristjánsson, 2016; Kristjánsson and Campana, 2010; Maljkovic and Nakayama, 1996) resulting in changes in search efficiency. Importantly, the repetition of task-irrelevant features also affects future trials (Huang, Holcombe, and Pashler, 2004; Burnham, 2015), and repetition effects are moderated by task context (Thomson and Milliken, 2013). These results indicate that even when people do not *have* to learn some aspects of the task, they nevertheless do it, even if it harms future performance. Such automatic learning could exist in parallel to more controlled or explicit learning (Maljkovic and Nakayama, 1996; 2000; Perruchet, 1985; 2015) and become a source of recurring errors.

On the other hand, the erroneous response pattern may not seem erroneous or irrelevant from a subjective perspective. "Error", as a term, is always relative to a particular frame of reference. The action that is considered a mistake from the experimenter's point of view could be a correct behavior from the subject's viewpoint. Recurring errors often appear in experimental designs that do not provide any feedback to a subject (Reber, 1989). Thus, a subject who has made a mistake might be quite confident that she is doing everything right and keep her behavior consistent. Dienes and Scott (2005) illustrated that consistency of errors might be associated with conscious efforts to do the task. In their experiment, the participants were asked to report on the source of their decisions in an artificial grammar learning task. Judgments of responses to be based on "guess" and on "intuition" did not correlate with high level of recurring errors ("consistent errors" in the authors' terms). It is only when subjects attributed their decisions to "rules" or "memory", i.e. made a special effort to control the task performance, that the proportion of recurring errors in comparison to non-recurring errors increased.

To summarize briefly described evidence, recurring errors may appear (1) when we make a conscious effort, (2) repeat the task several times and (3) fail to receive feedback about the accuracy of our previous judgments. Even so, some questions arise. If we consider the Alpha model, we need to consider the fact that our memory stores correct and incorrect responses together. How do we differentiate between an actually correct response and a response that seems to be correct? If recurring mistakes are stored in memory together with correct responses they might have an origin different from non-recurring mistakes. Is it so?

The "negative choice" framework

V. M. Allakhverdov, working in the former Soviet Union in 1974, not only developed a framework that accounts for the differences between correct and erroneous responses, but also described recurring errors as a consequence of a purported "negative choice" mechanism. In this chapter, we will first describe this approach and the main findings obtained. We will then compare studies carried out by V. M. Allakhverdov's group to similar effects observed in a variety of studies conducted elsewhere.

"Negative choice" is a hypothetical mechanism based on the idea of sustained unawareness. Allakhverdov (1993, 2000) reasoned that, just as is the case for all human information processing, the selection of information available for responses is subject to learning. Hence, it is possible that the "choice" to ignore some information might be implicitly remembered and later negatively affect the likelihood of the same information being used on the next encounters (see description of the theoretical context in Chapter 3 of this volume). This mechanism was used to explain a number of effects, such as recurring errors, the inability to recognize the second meaning of a reversible figure or difficulties in the retrieval of well-known information.

Crucially for the present treatment of recurring errors, the information necessary for a correct decision can be "negatively chosen". Allakhverdov (1993, 2000) states that, for example, if we fail to retrieve some information in a memory task, we still remember it, but with a tag stating "not to be retrieved". This tag keeps the information in memory, although it prevents us from reporting it until the context of the task is changed and predictions about the stimulus pattern and response are changed as well. This "negative choice" framework explains the Alpha model (Peak, 1941) discussed above: a response to a given stimulus pattern increases the probability that on the recurrence of the same stimulus pattern, the same response will occur because it is stored in memory with an extra information about its previous usage.

Thus, according to the "negative choice" hypothesis, our memory stores previously chosen and non-chosen responses to the stimulus together with their identification as "to be retrieved" or "not to be retrieved" or, if we take a wider approach, with an identification as "to be aware" or "not to be aware". This label or tag makes recurring mistakes different from non-recurring mistakes because it assumes that the element is activated in memory together with its tag.

Allakhverdov (1993) further suggested that a pattern of responses is kept unchanged only as long as the context of the cognitive task or the task itself remains the same. Once the context changes, the tag/label structure of stimulus–response associations also changes. Moreover, the tag "not to be retrieved" is not applied anymore and the stimulus remains activated in memory. That is why the information that is now stored without a tag starts to pop up in your mind and may become a source of reminiscence-like effects as well as a source of non-recurring mistakes.

Let us consider our example about misspelling the word "psychology" again. According to the idea of "negative choice", when we first encounter the task of spelling the word, we implicitly form an association between the representation of the word and our particular response. The first spelling mistake may occur because of some external/internal factors. Once it is made, however, the erroneous response is stored in memory together with the representation of the word. The correct response is also stored in memory, but with the tag "not to be retrieved" or "not to be used". Because the association is formed between the erroneous response and a representation of the word, we tend to repeat this erroneous response each time we encounter the word "psychology" (as in the Alpha model). We also tend to "negatively choose" the correct spelling. It is important to mention that "negative choice" is a result of unconscious implicit learning and does not imply conscious control over task performance. We can continuously make the same mistake until the task or the context changes, and this change breaks the former association of a word representation and the erroneous response. We may also correct the mistake once feedback gives us an opportunity to become aware of such an association.

In summary, the "negative choice" framework proposes two main statements about recurring mistakes:

1) A previously "negatively chosen" item can become continuously "negatively chosen", i.e. every mistake can become a recurring mistake if the task includes repetitive actions with the same set of stimuli.[2] Due to the fact that the recurring mistake is stored in memory it elicits faster responses and higher levels of confidence ratings in comparison with non-recurring mistakes.
2) A continuously "negatively chosen" item may lose its tag "not to be retrieved" when the task or the context of the task changes. After that, information about the "negatively chosen" item starts to pop up in mind in the form of a non-recurring mistake or reminiscence-like association.

These two statements were coined as the "negative choice aftereffect" (Allakhverdov, 1993).

Further in this chapter we present experimental evidence of this phenomenon.

Experiments on "negative choice aftereffect"

How to measure a "negative choice aftereffect"

The very first experiments on the "negative choice aftereffect" were conducted in 1974 as a part of doctoral dissertation by V. M. Allakhverdov (1974). Here we present some of the data from a 1977 study that was later partially described in a book (Allakhverdov, 1993).

On each trial (out of 700), participants were briefly shown a semicircle scale that looks like a speedometer (Figure 4.1) with an arrow that points in a certain position. This scale ranged from 0.1 to 5.0 and was divided into fifty increments. The time of presentation ranged around 300 msec. After presentation participants tried to manually match the position of the arrow with the one they had previously seen. For example, when the arrow pointed to 3.4, participants had to position a test arrow into the same location. The error was calculated as the absolute deviation of the response as compared to the original position. The task required a lot of practice and only three subjects carried out all 700 trials.

The goal of the study was to test whether any given response of the subject is dependent on the previous responses that were associated with this very stimulus. In other words if I made a mistake when trying to position an arrow into the "3.4" location, will I tend to make the same mistake again when I am presented with the same position a second time? If yes, then this will be a recurring mistake.

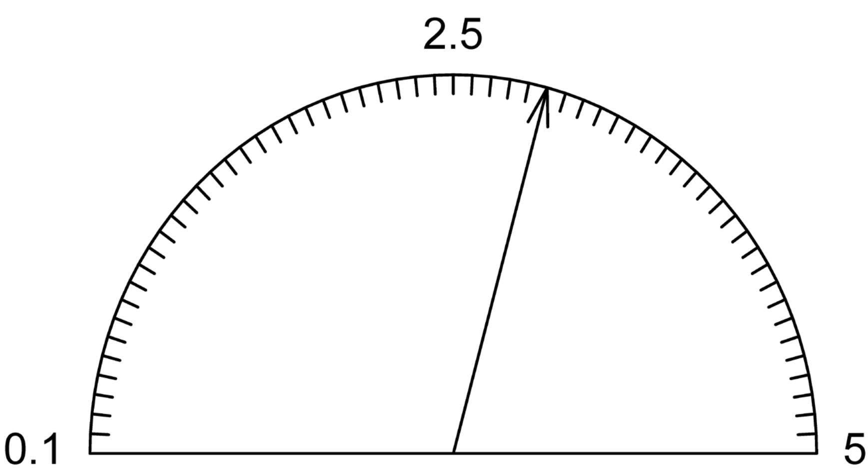

FIGURE 4.1 Semicircle scale that was used in the 1977 experiment of V. M. Allakhverdov in the study of the negative choice aftereffect.

In this experiment, the probability of a mistake across the subjects was 0.29.

The probability of making a mistake the next time the same stimulus was shown was found to be 0.43.[3] This result was confirmed when recalculated per each subject independently. The probability of a recurring mistake was calculated as the quotient of the number of recurring mistakes for this stimulus and the sum of recurring and non-recurring mistakes for the same stimulus. Table 4.1 presents a piece of raw data from the experiment where “0” stands for a correct response and any other number stands for an error measured as a deviation from the correct response. On the first trial, as we can see, the probability of a mistake is 0.5 (6 mistakes out of 12 presented arrow locations). On the second trial the probability of making a mistake is 0.6 (8 mistakes out of 12 presented arrow locations) but the probability of a recurring mistake is 0.5 (4 recurring mistakes out of 8 mistakes of all kinds). The third trial shows an increase of the recurring mistakes probability. Now it is 0.85 (6 recurring mistakes out of seven mistakes of all kinds) whereas the probability of any mistake becomes smaller (0.58).

A skeptical reader may consider such a result an artifact. Reading a set of briefly presented numbers on a semicircle scale can be quite hard. Subjects tend to make most of their mistakes on certain positions, i.e. close to the end of the scale, whereas the center of the scale and increments marked with numbers are usually

TABLE 4.1 Deviations from the reading on the scale pooled for subject and type of stimuli. “x” shows that the protocol marked the response as an incorrect but didn’t specify the deviation from the reading on the scale.

Marks on the scale	*0.1*	*0.2*	*1.7*	*1.7*	*1.7*	*2.4*	*3.1*	*3.3*	*3.7*	*4.4*	*4.6*	*4.6*
Subjects’ ID Trial number	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
1	0	0	0	0	0	0	−1	−1	−1	−1	+1	+2
2	+1	0	x	+1	+1	+1	−1	−1	0	+1	−1	0
3	+1	0	+1	+1	+1	+1	−1	0	0	−4	0	0
4	+1	+1	+1	0	x	+1	0	0	0	+2	0	0
5	+1	+2	1	+1	0	0	−1	−1	0	+2	0	+2
6	+1	+1	+1	+1	0	+1	0	−1	+1	−1	0	+2
7	+1	+1	0	+1	0	0	0	−1	0	+2	+2	0
8	+1	+1	0	0	0	0	0	0	0	+2	+2	+2
9	+2	+1	0	0	x	0	−1	−1	−1	−1	0	+2
10	+1	+1	0	0	0	0	0	−1	−1	+3	+2	0
11	+1	+1	0	0	0	0	−1	−1	0	0	0	0
12	0	+1	0	+1	+1	−1	0	0	0	+1	x	0
13	0	0	x	+1	+1	−1	−1	−1	0	0	x	+1
14	0	0	0	+1	0	0	0	−1	+1	0	x	0
15	0	0	0	0	0	0	x	0	−2	0	x	+1

remembered better. Thus we might suspect that errors recur only in locations where lots of mistakes are made and do not recur when the task is easy.

If this assumption were true, we should observe a tendency to repeat one's own mistakes as a result of a systematic bias rather than any other reason. Then if we take cases with overall poor performance, we might find more recurring errors there. To check for this assumption, only the cases (locations of the arrow pointer) with more than 50% of incorrect responses were analyzed.

If a skeptical point of view is correct, then these cases should show an increased probability of making a recurring mistake in comparison with a probability of making a non-recurring mistake. The probability of making a mistake in low performance cases was 0.61, while the probability of making a recurring mistake was 0.58. We do not see an increase of recurring mistakes in this case. On the other hand, if we take only the cases in which performance accuracy varies from 50% to 80% of correct responses, we can observe a different picture. The probability of making a mistake in high performance cases was 0.37, while the probability of making a recurring mistake was 0.53. Thus this comparison showed that we cannot associate a recurring mistake's effect with poor quality of task performance.

We can also consider the amplitude of mistakes, i.e. the deviation from the reading on the scale. Most of the time a subject's response deviates not more than a couple of units (increments). Table 4.1 shows these response deviations pooled for subjects and stimuli. This table illustrates that mistakes go together in some pattern of a similar deviation. One can assume that an observed pattern is a result of a systematic shift. If a subject tends to make a mistake of a certain type (for example a deviation is always larger than the reading on the scale) then a recurring mistake should be of the same type. Table 4.1 shows that the most common deviation for all three subjects is an increase in the reading on the scale.

To test this we can take only the pairs of mistakes that deviate from the correct response by one unit (those marked by + 1 in Table 4.1). For example, Subject 1 in the second and third trials made the same mistake by recognizing "0.1" reading on the scale as "0.2". Such mistakes were pretty common; they amount to more than 80% of all mistakes. At first, we calculated a theoretical and empirical probability of making two mistakes in a row with the same deviation. Then we calculated the probability of two mistakes with the same deviation but with some number of correct responses between them. The result can be seen in Table 4.2. When the same stimulus is shown, there is a strong tendency to repeat one's own mistakes. This tendency is clear for the very next presentation of the stimulus and becomes less clear for subsequent presentations. Thus we can conclude that a systematic shift, if any, happens for certain stimuli and depends on the previous response to the stimuli. It's important to mention that this systematic shift is not caused by overall poor performance on the task but rather happens when the overall accuracy is above 50%.

The data presented so far serve the purpose of illustration of the first statement of V. M. Allakhverdov's framework, i.e. that i.e. every mistake can become a recurring mistake if the task includes repetitive actions with the same set of stimuli.

TABLE 4.2 Probability of making a mistake when it happens right after the first mistake or with some gap. All mistakes are calculated for the same stimulus. The last column represents data pooled across three subjects.

	Subject 1	*Subject 2*	*Subject 3*	*Across subjects*
Empirical probability of a mistake	0.56	0.54	0.51	0.54
Empirical probability of two mistakes in a row	0.84	0.85	0.79	0.83
Empirical probability when two mistakes are separated by some number of correct responses	0.63	0.67	0.6	0.63

This statement assumes that recurring mistakes unlike single non-recurring mistakes are stored in memory for a longer period of time. If they are stored in memory we might find some evidence of it, for example, facilitated reactions to mistakes that repeat over several trials.

Negative choice aftereffect processing: analysis of reaction times and confidence ratings of recurring mistakes

Erroneous responses are often accompanied by uncertainty and failures of control over the task. In general, mistakes increase response time as compared to correct answers, although results for the tasks with the time pressure seem to be different (see Pleskac and Busemeyer, 2010).

If the subject hesitates to respond then the hesitation may increase the time of processing and the probability of producing a mistake. The change in response time might also correlate with overall response frequency. In simple cognitive tasks, correct responses outnumber mistakes and hence become a more frequent type of answer, which elicits shorter response times (RTs) (Notebaert et. al., 2009; Nunez Castellar et. al., 2010). For example, Rabbitt and Rodgers (1977) showed that post-error slowing is not observed when subjects are making the same type of response which they should have made on a previous trial. This means that the more frequent the error is, the shorter the response latency can be.

In the 1977 experiment, V. M. Allakhverdov noticed that the response latency in the case of a correct response was much shorter than in the case of a single mistake. But this was not the case with recurring mistakes. If the subject repeated her/his own mistake, the response latency decreased in 52% of cases out of all recurring error cases.

Another example of a change in response latencies was shown in one of the experiments of V. M. Allakhverdov's students – N. Andriyanova – conducted in 2015. We will present a more formal description of her experiment here.

Clock faces experiment

Participants

Sixty observers (43 females, 18–28 years old) at Saint Petersburg State University voluntarily participated in the experiment. They were not paid for participation. All reported normal or corrected-to-normal visual acuity.

Procedure and design

A number of analogue clock faces were presented (Figure 4.2). All stimuli were shown on the 19-inch computer screen using PsychoPy (Peirce, 2007). Every clock face appeared on a screen for 200 ms. The task was to memorize the time shown and write a response in a response box once the stimulus had disappeared. A total of 120 trials were grouped into 10 blocks. In each series, the same 12 stimuli were presented in a different order. During the course of the tasks, responses and response latencies were recorded. A recurring mistake was considered to be a mistake made in response to the same stimulus. For example, a case when a subject saw 1:30 p.m. on a clock face and repeatedly mistook the reading for something else was considered a recurring mistake.

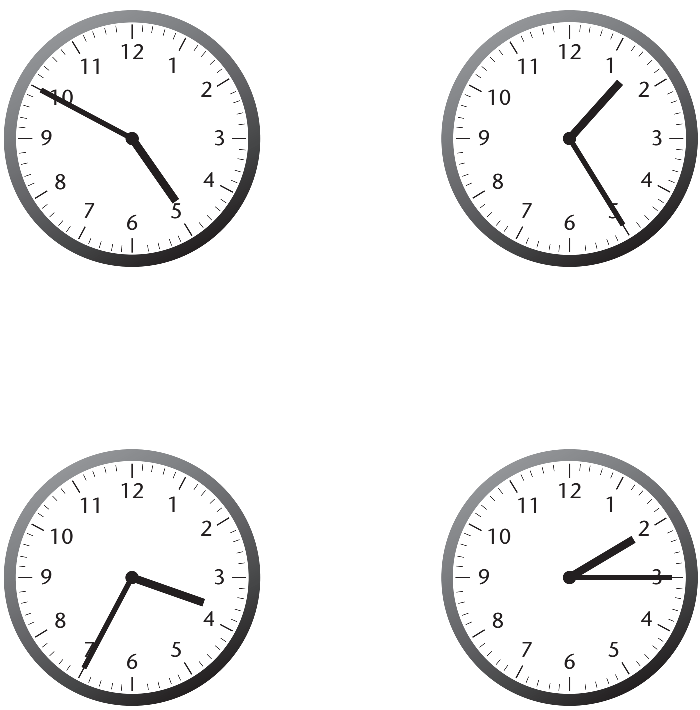

FIGURE 4.2 Analogue clock faces that were presented one per trial in the experiment of N. Andriyanova.

We hypothesized that error latencies might decrease over the course of learning the association between a certain stimulus and response in the task. Thus, reaction latency to a single error (which will become a recurring error) at the beginning of the task should be longer than the average reaction latency to a recurring error.

In our analysis we compared (i) latencies of correct responses, (ii) latencies of single errors, (iii) latencies of single errors (which will become recurring errors) and (iv) latencies of recurring errors. Half of the participants in each experiment were required to estimate confidence in their answers after each trial. We used a one to five confidence rating scale ("one" identified the least confident response and "five" identified the highest level of confidence in a given response).

Results

Mean RTs for different types of answers are presented in Table 4.3. Our data showed that RTs for single errors (M = 7077 ms, SD = 1872) were significantly longer than for correct answers (M = 5298 ms, SD = 1378) and for recurring errors (M = 5564 ms, SD = 1448), $t(49) = 9.506$. $p < .01$. Surprisingly, there was no significant difference between RTs for correct answers and recurring errors, $t(49) = 1.148$; $p > .1$ assuming that there is probably no post-error adjustment in recurring errors (as in Danielmeier and Ullsperger, 2011) (Table 4.3).

RT for a single error (which will become a recurring error) (M = 6092 ms, SD = 1794) was significantly longer than for a repeated mistake, $t(49) = 2.95$; $p < .05$, but significantly shorter when compared to a single error reaction, $t(49) = 7.202$; $p < .01$.

The analysis of the confidence data showed statistically significant differences in evaluation of correct and incorrect responses, Wilcoxon $V(29) = 450$; $p < .01$, i.e. participants were more confident in their correct answers, than they were in the incorrect ones. Confidence levels for repeated errors were nevertheless higher than for single errors, $V(29) = 400$; $p < .01$ (Table 4.4) which might be evidence of recurring errors being processed differently from single ones.

RTs correlated with confidence of the response: RTs were significantly shorter for the confident answers, $t(27) = 4.759$; $p < .01$. However, even when responses were the least confident (from one to three on a scale) we still observed the difference between repeated errors (M = 6134 ms, SD = 2254) and single errors (M = 7746 ms, SD = 2038), $t(27) = 4.007$; $p < .01$.

TABLE 4.3 Mean response times (SD in parenthesis).

Type of answer	*Mean response times (ms)*
Correct response	5298 (1378)
Recurring error	5564 (1448)
Single error (that will become a recurring error)	6092 (1794)
Single error	7077 (1872)

TABLE 4.4 Medians of confidence ratings (range in parentheses).

Type of response	*Medians of confidence ratings*
Correct response	4.3 (1.9)
Recurring error	4 (2.3)
Single error (which will become a recurring error)	3.8 (2.5)
Single error	3.5 (2.3)

Discussion

The experiment described above shows that recurring errors might elicit shorter reaction latencies as well as higher confidence ratings. This result can be interpreted in terms of learning an association between the stimulus and response. When the association is not established (when a mistake is made for the first time) the reaction time reflects an error processing/procedural control operations. When the same stimulus elicits the same erroneous response multiple times the established association between a stimulus and response makes the time of processing shorter and comparable to correct responses. On the other hand the absence of a visible post-error adjustment in recurring mistakes indicates that subjectively they are not perceived as errors.

The status of a single mistake that will become a recurring error is vague. Our data shows that it is processed faster than the single mistake which could be an artifact result. To interpret this we need to assume that the association between the particular stimulus and particular response might be established at the very first presentation of the stimulus. This association might distinguish recurring errors from the single ones and lead to impaired procedural control over former responses and shorten the time of processing.

According to V. M. Allakhverdov, response latency depends on the diversity of the responses (Allakhverdov, 1993). This effect reminds one of the Hick's law (Hyman, 1953) that states that response latency depends on the number and diversity of the signal. Thus, predominant correct responses should elicit the shortest reaction and single errors as the more rare response should elicit longer reaction (Notebaert et. al., 2009; Nunez Castellar et. al., 2010). Recurring errors as a more frequent type of error then should be somewhat between correct responses and single error responses. This is exactly what was confirmed in the experiment of N. Andriyanova.

Robustness of negative choice aftereffect with time and context

As was stated before, the negative choice aftereffect manifests itself in two forms:

1. Every mistake may become a recurring mistake if the task includes repetitive actions with the same set of stimuli.

2. A mistake may cease to be recurring mistakes when the task itself or the context of the task changes. After that the previously stored information starts to pop up in mind in a form of a non-recurring mistake (Allakhverdov, 2000).

Experiments described above illustrated the first statement. Here we want to present two experiments that illustrate the second statement. The first study was conducted by one of the members of V. M. Allakhverdov's research group – M. Filippova. Her experiment aimed at testing robustness of recurring mistakes with time. The second study was conducted by one of the students of V. M. Allakhverdov – A. Odainic. His experiment was aimed at testing the robustness of recurring mistakes with change of task context.

Toy car experiment

Participants

Eighteen observers (11 females, 18–56 years old, age Mdn = 29) at Saint Petersburg State University voluntarily participated in the experiment. They were not paid for participation. All reported normal or corrected-to-normal visual acuity. Group 1 (immediate recognition) consisted of 7 participants (5 females); Group 2 (recognition with delay of one day) – 11 participants (6 females).

Procedure and design

For the purpose of the experiment, we used toy cars no larger than 7 cm each. These toy cars were presented in a row of 14 items (Figure 4.3). Each trial consisted of a different set of 14 toy cars. Subjects were presented with 84 different toy cars throughout the experiment.

In the first part of the experiment, subjects were presented with a "retention set #1", which consisted of 14 toy cars for 15 sec. After a pause, a recognition set consisting of 14 just presented and 14 new toy cars was shown. The recognition task was limited to 25 sec. This procedure was repeated three times, each time with the same set presented for memorization.

Then a new "retention set #2", also consisting of 14 toy cars, was presented for 15 sec. Its recognition was either delayed for one day (Group 2), or followed immediately the first part of the experiment (Group 1). This recognition procedure comprised the second part of the experiment. The recognition set (28 toy cars) for the second part of the experiment, aside from cars from "retention set #2", also contained the items from "retention set #1". These items were either recognized correctly or not recognized by an individual participant during all three trials in the first part of the experiment. On average, subjects were able to correctly recognize about three items.

Altogether 84 model toys cars were used, 2 sets of 14 items used for retention and 4 sets of 14 items used for recognition. New items that were presented for

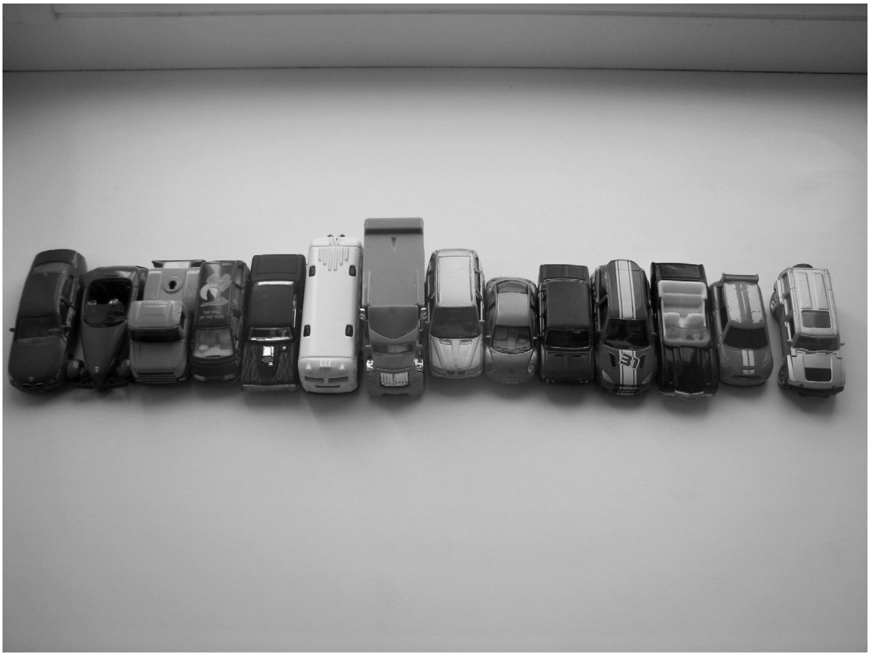

FIGURE 4.3 An example of a set of the toy cars used in the experiment of M. Filippova.

recognition were not presented anywhere else. We hypothesized that the first part of the experiment could create a possibility for repeated mistakes to occur. After that we would be able to test the robustness of recurring mistakes in another task. Immediate presentation of the other task ("retention set #2"), in our opinion, could not override the memory tag "not to be retrieved" for the recurring mistakes that appeared in "retention set #1". On the other hand a delay for a day could have affected the memory traces and loosened the connections between certain stimuli and responses. If recurring mistakes were dependent on a current memory for stimulus–response association then we might find former recurring mistakes as false alarm mistakes in a different task.

Results

At first, we tested whether any of the toy cars in the subsets was salient enough to be recognized better or worse than any other item. We did not find any differences between the recognized and unrecognized set of items in terms of their stimuli specificity ($\chi^2(13) = 12.214$, $p = .510$). Thus, the stimuli set happened to be more or less homogenous and recognition effects could be attributed to some reason other than the saliency of the items.

Then, we compared frequencies of recurring and non-recurring responses. If the item wasn't recognized in three consecutive trials it was considered to be a

TABLE 4.5 Probability of different response types in first part of experiment (N = 18).

All types of mistakes	*Recurring mistakes*	*Non-recurring mistakes*
0.22	0.15	0.07

recurring mistake. Otherwise recognition in at least one trial made a response a non-recurring mistake. This procedure is slightly different from the one that was offered by V. M. Allakhverdov and described above. It was chosen to guarantee that the subject had some consistency in her/his responses.

Altogether the number of recurring (34) and non-recurring (16) mistakes was 22% of all possible recognitions on the last trial (14 items to recognize per 18 subjects). Table 4.5 represents the frequency of different types of response measured as a proportion of these answers to the overall amount of items to be recognized (252).

The number of recurring mistakes was significantly different from that of non-recurring mistakes ($\chi^2(1) = 6.480$, $p < .05$).

The previously described procedure included only one possible response (correct or incorrect) per trial. In the current experiment, the number of participants' answers in each recognition test was not restricted and varied from 7 to 15. When given such a possibility, subjects can chose a strategy to "recognize" as many or as few items as possible. When the subject chooses a large number of responses, such responses have a higher chance of having all correct responses than the choice of a small number of responses. In other words, if I chose 14 out of 28 toy cars, the chance that I picked up correct items is 0.5. If I chose 7 out of 28 toy cars, the chance that I picked up the correct items is 0.4.

The generalized linear mixed model was used (see summary statistics table, Table 4.6) to examine the predictors of the number of responses together with the previous history of erroneous recognitions. The history of previous erroneous recognitions was calculated as a cumulative sum of the number of mistakes made before the current trial. We assumed that the more recurring mistakes a subject made before, the higher is the probability of making such a mistake again. A dependent variable was a response in the current trial which could be either correct recognition ("0") or incorrect recognition ("1").

As can be seen from Table 4.6 the erroneous recognition in the current trial is dependent on the history of previous erroneous recognitions although the number of responses can become a confound.

The second part of the experiment consisted of one recognition trial of "retention set #2". Items that the subject worked with during the first part (previously recognized or not recognized) became fillers for the second part. Thus participants didn't have to report on them and once recognized the stimuli were considered to be false alarm mistakes (Table 4.7). If recurring mistakes are kept in the memory then they will pop up in the mind during the second part of the experiment in the form of false alarm mistakes.

TABLE 4.6 Generalized linear mixed model summary for the experiment of M. Filippova.

	Erroneous response		
	Odds Ratio	*CI*	*p*
Fixed Parts			
(Intercept)	0.25	0.00 – 17.55	.525
Previous recurring errors	11776.31	283.14 – 489796.19	<.001
Number of responses	0.73	0.45 – 1.16	.181
Previous recurring errors: Number of responses	0.57	0.41 – 0.79	<.001
Random Parts			
$\tau_{00.\ \text{Subject}}$		1.063	
$\tau_{00.\ \text{Stimulus}}$		0.000	
N_{Subject}		17	
N_{Stimulus}		7	
ICC_{Subject}		0.244	
ICC_{Stimulus}		0.000	
Observations		325	
Tjur's D		.690	
AIC		179.652	
Deviance		143.255	

Mean accuracy for recognition in Group 1 where the second part was presented immediately after the first one was 54%. Most responses were correct rejections and not false alarms. Subjects still remembered items from the previous "retention set #1" and tried not to retrieve them.

Mean accuracy for recognition in Group 2 where the second part was delayed for 24 hours was 37%. Previously unrecognized items were chosen with a similar frequency as previously recognized ($\chi^2(1) = .578$, $p = .81$) and numerically more often than previously not presented items ($\chi^2(1) = 2.8$, $p = .09$).

TABLE 4.7 Possible responses for the stimuli in the second part of the experiment based on the first part of the experiment.

	Responses in the first part of the experiment	*Responses in the second part of the experiment*
Item is presented in the first part of the experiment	miss	false alarm
		correct rejection
	hit	false alarm
		correct rejection
Item is presented in the second part of the experiment		miss
		hit

TABLE 4.8 Probability of recognition in the second part of the experiment based on the history of responses in the first part of the experiment.

First part of Experiment	*Recurring misses*	*Recurring hits*	*Item is not presented in the first part*
Second part of experiment	*False alarm mistakes*		*Misses*
	0.39	0.42	0.19

With a time delay (Group 2) items from "retention set #1" became more salient and were retrieved in the second part as if they were the actual stimuli (Table 4.8).

Discussion

This experiment was aimed at illustrating robustness of recurring mistakes with time. The difference between the first and the second part of the experiment showed that a 24-hour delay is enough to break the association between the stimuli and a hypothetical tag "not to be retrieved". According to V. M. Allakhverdov's approach the stimulus stays activated in the memory even when this tag disappears. That is why items that were not correctly recognized in the first part of the experiment started to appear in the second part of the experiment in a form of false alarm mistakes. It's important to notice that the false alarm rate for the items that were previously correctly and incorrectly recognized was similar. This evidence supports an idea of equal memory storage for correct responses and recurring mistakes.

Another type of context change was investigated in the study of A. Odainic. His experiment was aimed at testing the robustness of recurring mistakes with a change of task context.

Matrices experiment

Participants

Sixty observers (38 females, 16–30 years old, age mdn = 19) at Saint Petersburg State University voluntarily participated in the experiment. They were not paid for participation. All reported normal or corrected-to-normal visual acuity. Participants were randomly allocated to one of two conditions (see design).

Procedure and design

Observers memorized a series of matrices. Each element of the matrix consisted of a single digit number and a capital letter of the Roman alphabet (for example "8Q"). The following letters – B, C, D, F, G, H, J, K, L, M, N, P, Q, R, S, T, V, W, Y, Z – and digits – from 1 to 9 – were used. There were five trials; each consisted of the presentation of two matrices. Each matrix consisted of 16 elements (four by four size). All elements were generated through an automatic randomization procedure. Four elements out of 16 were repeated, so that each subject on each trial observed a matrix with 12 new and 4 old elements. The trial began with a matrix presentation for 20 sec. The task was to memorize all the elements and to recognize those that were shown again on the second matrix, which was presented immediately after the 20-sec presentation of the first matrix. No time limit was established for the recognition task, which participants carried out by clicking on the items they thought were repetitions. No feedback was provided. The procedure was repeated three times with short time intervals between repetitions. The entire experiment consisted of two conditions randomized between subjects. In the "different place" condition, repeated elements kept their identity, but not the location where they were presented (Figure 4.4).

In the "same place" condition, repeated elements were presented at the same locations they had been shown before. Thus, we checked for the influence of the factor of location in the recognition of the repeated elements. The manipulation aimed to test whether memory for recurring errors contains information about location together with information about identity. We expected that recurring errors would show up in the "same place" condition more frequently than in the "different place" condition. Response in the current trial, i.e. the number of recognized items, could vary from 1 to 16 (there were 16 elements in a matrix). A participant could prefer different strategies of answer: she could decide to trust her memory, be more confident about her response and try to name as many items as possible, or she could decide to distrust her memory and name only the items that

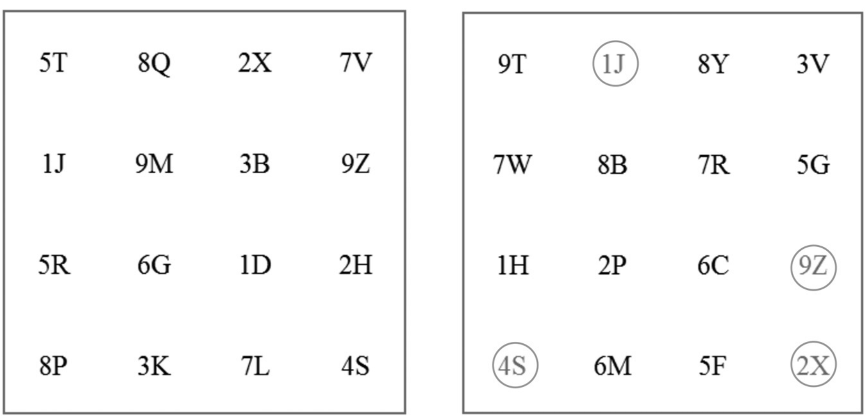

FIGURE 4.4 Design of one trial in the "different location" condition. On the left panel there is a set of elements to remember. The set is presented for 20 sec. On the right panel there is a recognition set. Elements that were repeated from the previous set are highlighted. Repeated elements keep their identity but not their place. In the "same location" condition repeated elements keep their place as well as their identity.

she was sure about. Such reduced response strategy could influence the calculation of recurring mistakes. For example, out of 16 elements presented in every trial, a participant could choose to name only two in every recognition trial. In this case there is a 50% chance level that two elements out of four will be missed and will be treated as a recurring mistake.

Results

We used the same approach to the analysis of mistakes as in the previous experiment.

The generalized linear mixed model was used (see summary Table 4.9) to examine the predictors of number of responses together with the previous history of erroneous recognitions. The history of previous erroneous recognitions was calculated as a cumulative sum of the number of mistakes made before the current trial. We assumed that the more recurring mistakes a subject made before, the higher is the probability of making such a mistake again. A dependent variable was

TABLE 4.9 Generalized linear mixed model summary for the experiment of A. Odainic.

	Erroneous response in the current trial			
	Estimate	*CI*	*SE*	*p*
Fixed Parts				
(Intercept)	2.54	1.35 – 3.72	0.61	.040
Previous recurring errors	5.55	3.99 – 7.11	0.80	<.001
Number of responses	−1.45	−1.83 – −1.08	0.19	<.001
Location	0.22	0.01 – 0.43	0.11	.042
Previous recurring errors: Number of responses	0.21	−1.73 – 2.15	0.99	.834
Previous recurring errors: Location	1.25	0.24 – 2.27	0.52	.016
Number of responses: Location	0.64	0.16 – 1.12	0.24	.009
Previous recurring errors:Number of responses:Location	2.26	−0.28 – 4.80	1.30	.082
Random Parts				
σ^2		0.428		
$\tau_{00,\ Subject}$		0.025		
$\tau_{00,\ Stimulus}$		0.097		
$\tau_{00,\ Repetition}$		0.990		
$N_{Subject}$		26		
$N_{Stimulus}$		4		
$N_{Repetition}$		3		
$ICC_{Subject}$		0.016		
$ICC_{Stimulus}$		0.063		
$ICC_{Repetition}$		0.643		
Observations		345		
R^2 / Ω_0^2		.599 / .599		
AIC		743.371		

a response in the current trial i.e. the amount of repeated elements misses (ranging from 0 to 4).

As can be seen from the results the erroneous response is dependent on the history of previous erroneous recognitions, number of responses and location of the elements, and, partially, on their interactions. Placing elements onto the same positions over the course of trials results in more mistakes in comparison with placing elements on randomized positions ("Random places group"–"Same places group" contrast estimate = – 0.3 (0.15), df = 76.36, t-ratio = –.207, p = 0.04)

Discussion of experiments of A. Odainic and M. Filippova

Both experiments presented above tested the idea of the relevant robustness of the recurring mistakes effect. We hypothesized that a recurring mistake is more likely to happen in presence of the same context, i.e. when the binding between identity of the item and its location is kept the same throughout trials. If the location changes, then the recurring mistake loses its location-identity binding and, possibly, also loses the stimulus–response association in memory.

Both experiments showed that the previous history of recurring errors, i.e. the cumulative sum of the omissions per item, influences the current response. This happened only when the place of the element was not changed throughout the trials (in the experiment of A. Odainic) and when there was a time delay between the retention and recognition procedures (in the experiment of M. Filippova). When the place of the element is changing from trial to trial a factor of previous erroneous responses doesn't influence the current response. We can conclude that binding of location and identity plays an important role in the recurring mistakes effect. If context doesn't change then we are more likely to find recurring mistakes in the sequence of trials.

In both experiments we also see the influence of the subject's number of responses, which goes along with our predictions. The strategy to report a smaller number of elements increases the occurrence of repeated mistakes. Such strategy works because both experiments used free recognition procedure instead of forced-choice recognition. Free recognition allows us to vary the amount of items recognized in particular trial and may depend not only on a subject's memory but also on a subject's choice to respond to more/less items. If these experiments had used a forced-choice procedure, we could have expected no influence of a "number of responses" factor.

Conclusions

Summary of the "negative choice aftereffect" findings

Several experiments described above show that mistakes tend to be repeated. This effect was dubbed the "negative choice aftereffect". It was shown that recurring mistakes may happen in response to the same stimulus when this stimulus is presented several times. Recurring mistakes occur in different cognitive tasks involving either retrieval or recognition processes. This type of error leaves traces in memory that include information not only about the identity of the item that was associated with the erroneous response but also information about the location of and the previous response to this item. Recurring mistakes are different from the set of single mistakes: they elicit shorter reaction times than single errors. There seems to be no post-error adjustment after recurring errors. In addition, recurring errors correlate with higher confidence ratings in comparison with single errors. These results taken together with the ones obtained in implicit learning studies (Reber, 1989; Dienes and Scott, 2005) describe a stable "erroneous" (from the experimenter's point of view) association of response and stimulus that is stored in memory for quite some time. Integrating over the results of the experiments discussed above, those of the experiments by Allakhverdov (1993), and those of the experiments of Dienes and Scott (2005) and Reber (1989), we can list some conditions in which recurring mistakes are more likely to be found:

- The stimulus set must be homogenous, that is, stimuli should have similar features, although they also should be easily named and identified.
- There should be no feedback on accuracy of performance in the task.
- Efficiency of the performance should vary from 50% to 80% of correct responses.
- The location of the item in the presentation set should be constant.
- The item should be repeated several times in a row to elicit a possible recurring mistake response.
- The memory test should not be delayed. If the delay is substantial (for example, more than 24 hours) then recurring mistakes lose the strength of stimulus–response association in the memory and very likely will pop up in the mind.
- Subjects should make an effort to perform well on the task.

"Negative choice aftereffect" in other studies

As we stated it in the beginning of the chapter, recurring errors result from explicit or, more probably, from implicit learning of irrelevant or distracting information. Other examples of "erroneous" association of response and stimulus can be found in a variety of experiments (VanRullen and Koch, 2003; Warriner and Humphreys, 2008; D'Angelo and Humphreys, 2015; Hajcak and Simons, 2008; Neill, 2007).

For example, VanRullen and Koch (2003) presented several different tasks in the study of competition and selection in visual processing. They used free recognition, forced-choice recognition and priming picture-word matching tasks to test how many items a subject can remember after being briefly presented with a natural scene that contained these items. They showed that items not reported during the free recognition task and subsequently missed in the forced-choice recognition task showed a significant negative priming effect, that is, elicited longer responses and were more associated with errors than novel items. The authors concluded that "the negative priming effect suggests that these objects were in fact represented in the visual system, but that this representation was eventually suppressed"

(VanRullen and Koch, 2003, p.80). We suggest that the effect VanRullen and Koch (2003) observed is related to recurring errors. There is a succession of tasks that use the same stimuli; there are items that are missed in both the first (free recognition) and second tasks (forced-choice recognition); these twice-missed items then elicit more errors in the picture-word naming task. All these features match our description of recurring mistakes.

In a study of the tip-of-the-tongue (TOT) state, Humphreys and colleagues (Warriner and Humphreys, 2008; D'Angelo and Humphreys, 2015) found that making an error once increases the chances that it will be repeated. They presented definitions of words that participants then had to name. If participants were unable to remember the word but felt that it was likely to be remembered (i.e., participants experienced a TOT state), they read the same definition once again but after a time delay. The delay was either 10 or 30 sec. In the latter case, (when the delay was longer) subjects showed difficulty in recalling the same words again the second time when the procedure was repeated. Authors concluded that speakers tend to repeat TOT states for individual words. This case, although very different from that described in VanRullen and Koch (2003), also exhibits some similarities with the "negative choice aftereffect". When we described the hypothetical mechanism of recurring mistakes, we assumed that a correct response for the task would be remembered together with the representation of the task, but with a tag "not to be retrieved". This tag prevents the correct response from being used when we carry out the same task after we made an initial mistake. The TOT state is a good example of how such a "not to be retrieved" tag could be experienced subjectively.

The described experiments share one feature: the items that elicit a recurring mistake response are shown to be processed differently from the novel items that do not have an erroneous association between the stimulus and response.

Hypothetical mechanisms of recurring mistakes

Humphreys and colleagues (Warriner and Humphreys, 2008), in their research on repeated TOT states, suggest that implicit learning of the incorrect activation pattern (particularly the co-activation of a word's lemma and phonology) is a cause for repeated errors. Implicit learning of a TOT state happens when the participant tries to produce the word. Unresolved TOT states (the correct answer was not found and the word remained forgotten) are stored in memory together with the incorrect link to the phonology of the word. This extra information increases the likelihood of the TOT state repetition (Warriner and Humphreys, 2008, p. 540). Similarly to the previous explanation, Allakhverdov (2000) suggests that if we continuously do not retrieve the very same words as before, we still remember them but with extra information – a "not to be retrieved" tag. This tag keeps the words in memory although preventing us from reporting them until the context of the task is changed and predictions about the stimulus pattern and response are changed as well. Neill and Valdes (1992) used similar terms when they described the mechanism of the negative priming effect. In their opinion, task performance is mediated by the retrieval of past instances of the current stimuli. If the retrieved episode indicates that the current stimulus was recently ignored, responding is impaired which is called "inappropriate transfer". Analyses of this effect support the idea that a stimulus and response are associated within a given task/context and their association might alter the performance (Neill, 2007).

Here, we followed the Alpha model perspective on stimulus–response association, assuming that a response to a given stimulus pattern increases the probability that the same response will occur on the recurrence of the same stimulus pattern. According to the Alpha model, simple associative learning might happen after a first few iterations of a stimulus–response co-occurrence and continue throughout an experiment. Does it then mean that all possible responses for all stimuli are stored? The hypothesis of "negative choice aftereffect" leans towards this assumption. Supportive evidence comes from the neural network field. It was shown that a Bidirectional Associative Memory model (based on a Hebbian learning rule) with a Quick Learning algorithm can reproduce a human pattern of recurring mistakes when a participant learns and then repeats a hand movement trajectory (Lyahovetskiy et. al., 2013). According to its authors, the Quick Learning algorithm greatly reduces learning epochs, guarantees the recall of all training pairs and increases memory capacity (Hattori, Hagiwara and Nakagawa, 1994). Interestingly, McClelland emphasizes that Hebbian learning mechanisms that were utilized in Quick Learning for the Bidirectional Associative Memory algorithm tend to strengthen the pattern of neural responses that are highly correlated with each other. In his opinion, the Hebbian approach may even strengthen response when it is incorrect or when there is no feedback given (McClelland, 2006). Thus, a model that assumes a storage of all stimulus–response associations also assumes a possibility of recurring errors.

If the Alpha model implies that our memory stores correct and incorrect responses together, then it should also imply some separation of these responses. Ideas of "incorrect activation pattern" (Warriner and Humphreys, 2008), "negative choice aftereffect" (Allakhverdov, 1993, 2000) and "inappropriate transfer" (Neill and Valdes, 1992) all try to explain how incorrect responses could be distinguished from correct and from incorrect but non-repeating responses. Recurring mistakes are different from the non-recurring errors in the way they are stored in memory. The extra information (the tag "not to be retrieved") that is linked to correct answers makes it clearly distinguishable from other types of responses and prevents it from being used in the task.

Does this mean that the representations that are not retrieved are suppressed as VanRullen and Koch (2003) suggested? "Negative choice aftereffect" and recurring mistakes may happen because the correct response is not only stored in memory with a tag "not to be retrieved" but also is actively suppressed from further processing. To confirm that the suppression mechanism is active, we need to infer the consequences of such suppression. Usually, the suppressed elements elicit longer reaction times than non-suppressed elements. In line with this, in VanRullen and Koch's (2003) study the items that elicited more errors in the priming picture-word matching task also elicited longer reaction times, i.e. a negative priming effect. On the contrary, the results of Allakhverdov's group experiments described above did not show any negative priming effect. Thus, we cannot confirm the suppression hypothesis of recurring mistakes.

The recurring mistakes effect can be shown in a variety of tasks. It implies that the cognitive system stores excessive information about the stimulus–response association. This stimulus–response association can be implicitly learnt as an undesirable pattern of response (Yates, 1959) but still will be stored in memory. We began this chapter with an example of a repeated misspelling ("spychology" instead of "psychology"). Based on our current knowledge, we can conclude that it will be quite hard to get rid of this mistake. Firstly, the mistake recurrence will be supported by the increased probability that on the recurrence of the same stimulus pattern, the same response will occur. Secondly, this stimulus–response association will be stored in memory until the context of the task is not changed. Is it possible to get rid of the recurring mistake? In terms of Allakhverdov's approach, one should try to change the context of the task or the task itself. Dunlap (1942) offered instead to make it so that the information we were unaware of is brought to awareness. This approach involves a technique known as "negative practice". According to this method, one should try to repeat an error explicitly several times to correct it (Dunlap, 1942). Likewise, McClelland suggested increasing the contrast and exaggerating the differences between otherwise undistinguishable stimuli that previously elicited recurring mistakes (McClelland, 2006). Sure enough, this method helps to get rid of spelling and pronunciation errors. However, it is not known whether it works for the recurring mistakes that occur in other situations.

Acknowledgements

Preparation of this chapter was supported by Saint Petersburg State University (research grant # 8.38.287.2014) and a grant from the Russian Foundation for Basic Research (RFBR) # 17-06-01014-ОГН.

Notes

1 We would like to thank our colleague Y. Ledovaya for this example of a negative choice mistake.
2 It's important to notice that we assume that each stimulus in a set is more or less independent of any other and does not form a semantically connected string.
3 These results were obtained in 1974. Unfortunately, raw data files were not stored until 2017. We put here numbers in accordance with results published as a Doctoral dissertation (see Allakhverdov, 1993).

References

Allakhverdov, V. M. (1974). Zakonomernosti vozniknoveniya oshibok pri operativtih preobrazovaniyah informatsii [Regularities of mistakes occurrences in information processing]. Dissertation thesis, St. Petersburg State University, St. Petersburg.

Allakhverdov, V. M. (1993). *Opit teoreticheskoy psihologii* [*The Experience of Theoretical Psychology*]. Saint Petersburg: Pechatny Dvor.

Allakhverdov, V. M. (2000). *Soznanije kak paradoks* [*Consciousness as a Paradox*]. Saint Petersburg: DNK.

Burnham, B. R. (2015). Intertrial priming of popout search on visual prior entry. *Journal of Vision*, *15(14)*, 1–7. doi:10.1167/15.14.8.

Chetverikov, A., Campana, G. and Kristjánsson, Á. (2016). (In press). Building ensemble representations: how the shape of preceding distractor distributions affects visual search. *Cognition*. doi:10.1016/j.cognition.2016.04.018.

Chetverikov, A. and Kristjánsson, Á. (2015). History effects in visual search for monsters: search times, choice biases, and liking. *Attention, Perception, & Psychophysics*, *77(2)*, 402–412. doi:10.3758/s13414-014-0782-4.

D'Angelo, M. C. and Humphreys, K. R. (2015). Tip-of-the-tongue states reoccur because of implicit learning, but resolving them helps. *Cognition*, *142*, 166–190. doi:10.1016/j.cognition.2015.05.019

Danielmeier, C. and Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology*, *2(233)*. doi:10.3389/fpsyg.2011.00233.

Dienes, Z. and Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. Psychological Research, *69(5–6)*, 338–351.

Dunlap, K. (1942). The technique of negative practice. *The American Journal of Psychology*, *55(2)*, 270–273.

Hajcak, G. and Simons, R. F. (2008). Oops! I did it again: an ERP and behavioral study of double-errors. *Brain and Cognition*, *68(1)*, 15–21. doi:10.1016/j.bandc.2008.02.118.

Hattori, M., Hagiwara, M. and Nakagawa, M. (1994). Quick learning for bidirectional associative memory. *IEICE Transactions on Information and Systems*, *77(4)*, 385–392.

Huang, L., Holcombe, A. O. and Pashler, H. (2004). Repetition priming in visual search: episodic retrieval, not feature priming. *Memory & Cognition*, *32(1)*, 12–20.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45(3)*, 188.

Kristjánsson, Á. and Campana, G. (2010). Where perception meets memory: a review of repetition priming in visual search tasks. *Attention, Perception & Psychophysics*, *72(1)*, 5–18. doi:10.3758/APP.72.1.5.

Lyahovetskiy, V. A., Potapov, A. S., Bobrova, E. V., Bogacheva, I. N. (2013). Obuchaemaya model zauchivaniya posledovatelnosti dvizheniy na osnove geteroassociativnoy neyronnooy seti [Movements sequence learning based on hetero associative neuronal network]. *Matematicheskaya biologia i informatica*. *8(2)*, 665–678.

Maljkovic, V. and Nakayama, K. (1996). Priming of pop-out: II: the role of position. *Perception & Psychophysics*, *58(7)*, 977–991.

Maljkovic, V. and Nakayama, K. (2000). Priming of popout: III: a short-term implicit memory system beneficial for rapid target selection. *Visual Cognition*, *7(5)*, 571–595. doi:10.1080/135062800407202.

McClelland, J. L. (2006). How far can you go with Hebbian learning, and when does it lead you astray? *Processes of Change in Brain and Cognitive Development: Attention and Performance xxi*, *21*, 33–69.

Neill, W. T. (2007). Mechanisms of transfer-inappropriate processing. In D. S. Gorfein and C. M. MacLeod (Eds.), *Inhibition in Cognition* (pp. 63–78). Washington, DC: American Psychological Association.

Neill, W. T., and Valdes, L. A. (1992). Persistence of negative priming: steady state or decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18(3)*, 565.

Notebaert, W., Houtman, F., Van Opstal, F., Gevers, W., Fias, W. and Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition*, *111(2)*, 275–279. doi:10.1016/j.cognition.2009.02.002.

Nunez Castellar, E., Kuhn, S., Fias, W. and Notebaert, W. (2010). Outcome expectancy and not accuracy determines posterror slowing: ERP support. *Cogn. Affect. Behav. Neurosci.*, *10*, 270–278.

Peak, H. (1941). Negative practice and theories of learning. *Psychological Review*, *48(4)*, 316.

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *J Neurosci Methods*, *162(1–2)*, 8–13.

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *The Pavlovian Journal of Biological Science: Official Journal of the Pavlovian*, *20(4)*, 163–170. doi:10.1007/BF03003653.

Perruchet, P. (2015). Dissociating conscious expectancies from automatic link formation in associative learning: a review on the so-called perruchet effect. *Journal of Experimental Psychology: Animal Learning and Cognition*, *41(2)*, 105–127. doi:10.1037/xan0000060.

Pleskac, T. J., and Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, *117(3)*, 864–901. doi:10.1037/a0019737.

Rabbitt, P. and Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *The Quarterly Journal of Experimental Psychology*, *29(4)*, 727–743.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118(3)*, 219.

Thomson, D. R., and Milliken, B. (2013). Contextual distinctiveness produces long-lasting priming of pop-out. *Journal of Experimental Psychology: Human Perception and Performance*, *39(1)*, 202–215. doi:10.1037/a0028069.

VanRullen, R. and Koch, C. (2003). Competition and selection during visual processing of natural scenes and objects. *Journal of Vision*, *3(1)*, 75–85. doi:10.1167/3.1.8.

Warriner, A. B. and Humphreys, K. R. (2008). Learning to fail: reoccurring tip-of-the-tongue states. *The Quarterly Journal of Experimental Psychology*, *61(4)*, 535–542. doi:10.1080/17470210701728867.

Yates, A. J. (1959). Negative practice – a theoretical interpretation. *Australian Journal of Psychology*, *11(1)*, 126–129.

5

CAN WE PLAY SPACE INVADERS UNCONSCIOUSLY? (A: PROBABLY NOT)

Estibaliz San Anton, Axel Cleeremans, and Arnaud Destrebecqz

Introduction

As shown by a large number of studies, both human and nonhuman animals are able to learn the contingencies between stimuli (Rescorla, 1968; Shanks, 1995). While this ability is crucial for the survival of organisms, the precise nature of the learning mechanism remains unclear. Indeed, two radically different views have been defended in the literature: associative strength theories, and propositional theories.

Proponents of the former argue that learning is based on the acquisition of simple associations between events, through the gradual updating of the associative strength between the mental representations of the stimuli (see Bouton, 2007; Pearce and Bouton, 2001; Shanks, 2005; Shanks and Dickinson, 1987, for reviews). The most influential associative learning model, the Rescorla–Wagner model (Rescorla and Wagner, 1972), assumes that learning is bottom-up and automatic. On this view, conscious awareness of the to-be-learned association is not a necessary condition for learning to occur because the behavioral changes induced by learning take place through low-level mechanisms of neural plasticity that have causal influence regardless of the intentions of the learner.

In contrast, supporters of the latter, propositional, view claim that learning is the result of active inference and reasoning deployed over propositional representations (e.g., Mitchell, De Houwer, and Lovibond, 2009). Learning would therefore result from "*the formation and truth evaluation of propositions about relations in the world*" (De Houwer, 2014). This model takes into account not only the relationship between the stimuli (i.e., their co-occurrence) but also the type of relationship between events (e.g., a causal relation; Lagnado, Waldmann, Hagmayer, and Sloman, 2007). Even though some discrepancies can be found in specific theoretical accounts (see De Houwer, 2014), such inferential models assume that learning is a non-automatic process that requires conscious, effortful, and top-down thinking.

Inferential theorists have further argued that the proponents of associative theories actually advocate a dual-process theory (Sternberg and McClelland, 2012) in which knowledge is expressed as a combination of associative and higher-order knowledge. However, higher-order knowledge is not the focus of associative theories, in which it is rather viewed as the basis of reasoning than as a mandatory component of the learning processes. As a case in point, for associative models (Miller and Matzel, 1988), or algorithmic statistical models (Meltz, Cheng, Holyoak, and Waldmann, 1993), cue-interaction effects (such as blocking) do not occur during the learning phase but at test, when participants are asked to reason about what happened during learning. Such a view supports a dual-process theory that involves both associative and propositional processes. Mitchell et al. (2009), however, have argued that inferential accounts should be preferred based on parsimony, as propositional processes can account for the entire pattern of results.

In the context of this ongoing debate, we focus here on a central difference between the two theoretical frameworks. According to inferential theories, learning is akin to rule learning, such as when one learns if/then relations between stimuli (see e.g., Beckers, De Houwer, Pineño, and Miller, 2005). As a matter of fact, participants who have learned such rules can report them when asked to do so (and both types of theory agree on this point). The central question, however, is the following: does learning always reflect initial rule acquisition, or does it build up gradually through repeated exposure to the two related events? According to this latter associative view, this gradual updating process does not necessarily depend on the ability to verbally express the relationship between events, while not excluding that it may eventually result in such conscious representations. According to inferential models, however, learning occurs *because* of the ability to form a conscious representation of the relationship (see De Houwer, 2014, for a critical overview). In other words, the question amounts to an ordering problem: do changes in behavior reflect the gradual strengthening of the associative link between two represented events that may later be expressed as a propositional rule (associative view) or is it the very existence of a propositional rule that causes changes in behavior in the first place (inferential view)?

Addressing this question requires exploring the dynamics of learning. For instance, we can measure two aspects of the newly acquired knowledge: its influence on behavior, and its availability to consciousness. Learning without awareness is already documented. For instance, in implicit learning tasks, participants are typically able to learn incidentally about the relationships between stimuli, but exhibit little or no ability to verbalize what has been learned (e.g., Destrebecqz and Cleeremans, 2001; Fiser and Aslin, 2001; for a review see Cleeremans, Destrebecqz, and Boyer, 1998). In some cases, changes in behavior can be measured before conscious knowledge can be reported. In the case of the Iowa Gambling task for instance, participants' ability to choose the advantageous decks amongst four improves before their being able to explicitly identify which decks are actually advantageous (Bechara, Damasio, Tranel, and Damasio, 1997, but see Maia and McClelland, 2004).

There is substantial and continuing debate in the literature between researchers who believe that associative learning can occur implicitly (Schultz and Helmstetter, 2010, Alamia et al. 2016) and those who claim that human learning is necessarily conscious (Lovibond and Shanks, 2002). Many studies have used subliminal stimuli that are inaccessible to awareness, for example using subliminal instrumental conditioning (Pessiglione et al., 2008), subliminal sequence learning (Atas, Faivre, Timmermans, Cleeremans, and Kouider, 2014), or fear-conditioning paradigms (Baeyens, Eelen, and van den Bergh, 1990; Schultz and Helmstetter, 2010). Some studies have used both procedures at the same time (Raio, Carmel, Carrasco, and Phelps, 2012; Tabbert, Stark, Kirsch, and Vaitl, 2006). Studies have also shown that the conscious or unconscious nature of the learning process may vary according to the training procedures. For example, delay conditioning might be independent from awareness, while trace conditioning would rely on explicitly noticing the to-be-learned association (e.g., in eyeblink conditioning, Clark and Squire, 1998, 1999; fear conditioning, Asli and Flaten, 2012).

In tasks that use neutral or non-emotional stimuli, the possibility of learning unconsciously has generally been put into question based on follow-up experiments using more sensitive measures of conscious knowledge (Lovibond and Shanks, 2002; Maia and McClelland, 2004, 2005). One possibility is that attentional factors also play a role, such as in neutral tasks using non-emotional stimuli. Thus, it may be the case that the amount of attentional resources available to perform the task is so reduced that participants do not learn the association between the paired stimuli at all. As a consequence, only those participants who consciously notice the association actually learn and show performance improvement. To address this issue and to increase the amount of attentional resources dedicated to the task, we used a more engaging experimental situation involving a video game that resembles the well-known "Space Invaders" arcade game. We reasoned that the increased attention resulting from engagement would result in better learning, and therefore manipulated instructions so as to improve attention.

Specific predictions can also be derived from Cleeremans' model of the relationships between consciousness and control during the course of learning. In his theoretical framework, Cleeremans (2008) argues that the quality of the representations acquired during a learning episode determines (1) the extent to which they influence behavior, (2) their ability to support cognitive control, and (3) their availability to awareness (see Figure 5.1). As a consequence, changes in the quality of a representation modify its influence on behavior, as well as its availability to cognitive control and to access consciousness. An important feature of Cleeremans' model is that these three features of representational quality change in different ways over the course of learning, resulting in different patterns of association or dissociation. For example, when representations are weak in the initial stages of learning, they cannot be the basis of conscious, controlled behavior but they may nevertheless exert (weak) effects on behavior. By contrast, after intensive training, high quality representations exert a strong influence on behavior, but this influence is now automatic in the sense that conscious control is not required anymore for successful performance. Such representations are available to conscious inspection, but their automatic character renders such access optional rather than mandatory. Over intermediate stages of learning, representations are both consciously accessible and susceptible to cognitive control, and thus correspond to "explicit" representations. Following Cleeremans (2008), we hypothesized that behavior and availability to consciousness evolve differently as the quality of the underlying representations improves during training. Such a hypothesis thus predicts the possibility of dissociations at certain points during learning.

Crucially, associative and inferential theories make different predictions within this framework. According to inferential theories, learning can only occur when we are able to consciously make inferences regarding the regularities present in the environment. As a consequence, these models predict that the acquired knowledge is always conscious. From the associative perspective, learning occurs gradually, with the quality of the representation pairing two events being strengthened through the repeated presentation of the association (see e.g., Rescorla and Wagner, 1972). This knowledge may then influence behavior right from the start. By contrast, the propositional representation of the association will reach awareness only once it has accrued sufficient quality. Such knowledge should then not be able

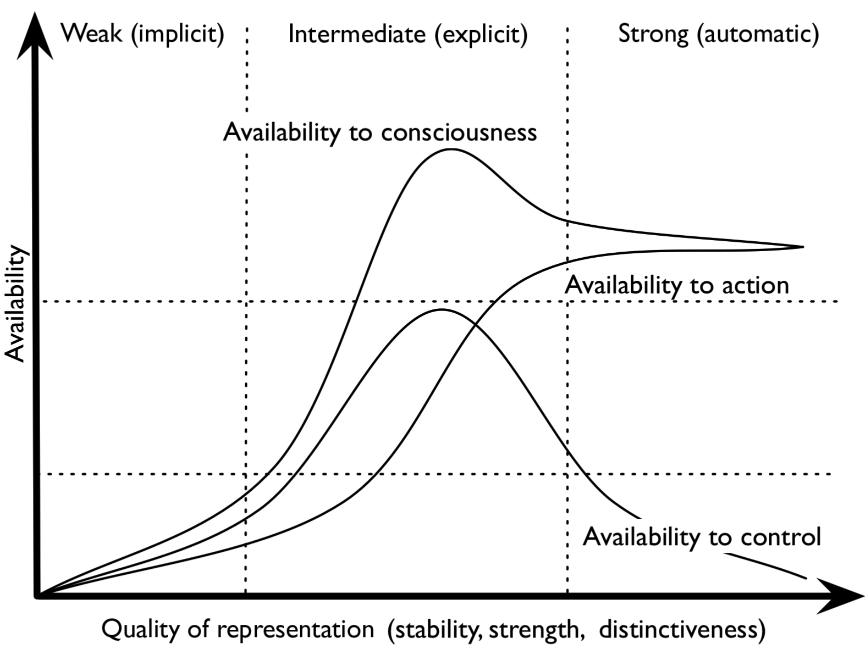

FIGURE 5.1 The Quality of Representation framework.

to influence behavior before it can be expressed. Therefore, it should be possible to find a dissociation between a representation's influence on behavior and explicit learning in the early stages of training, but we expect this dissociation to fade away as learning progresses.

In the present experiments, we used an adaptation of the Martians task (Franssen, Clarysse, Beckers, van Vooren, and Baeyens, 2010) to dissociate both verbal and behavioral measures of associative learning. In the Martians task, participants are asked to prevent Martians from landing on Earth. They have to press the spacebar almost continuously in order to destroy incoming Martian invaders, but the Martians may activate a protective shield. If participants press the spacebar when the shield is on, their shots are returned to them so that human resistance is neutralized and Martians can invade the Earth. It is therefore crucial to learn to predict the occurrence of the shield in order to be able to stop shooting before its onset. Unknown to participants, visual or auditory cues (unintelligible words written in the Martian alphabet or sounds symbolizing the Martian language) are predictive of the onset of the shield: in Experiment 1, a first cue, C+, predicted the shield in 100% of the cases and a second one, C±, preceded the shield in 50% of the cases. Finally, a third cue, C–, was never followed by the shield. In Experiment 2, only C+ and C– were presented. If participants learn the relationship between the occurrence of the cues and the onset of the shield, they should show different bar-pressing patterns for each cue: they should refrain from bar-pressing upon presentation of the perfect predictor of the shield C+, continue bar-pressing during the safe cue, C– (which is never followed by the shield), and they should show an intermediate behavior when the C± is presented.

This situation makes it possible to contrast associative and inferential theories based on the relationship between our continuous measure of performance (bar-pressing) and verbal reports. If participants learn the relation between the cues and the shield implicitly, we would expect a dissociation between behavior and conscious knowledge. That is, we should observe a decrease in bar-pressing rate when the perfect predictor is presented even before participants are able to verbalize the association. Further, this dissociation should fade away with training, as the quality of the representation increases and it progressively becomes available to consciousness. Inferential theories, on the other hand, because they assume that explicit knowledge is mandatory for learning to take place, would predict that changes in response rate should only take place *after* participants became able to report the rule.

In these studies, we manipulated the probability of occurrence of the shield after the preparatory signal. The discrimination between C+ and C– is easier to learn than the difference between C+ and C± as, in this latter case, the shield always follows C+ but the shield only follows C± in 50% of the cases. The shield, by contrast, never follows C–. If learning can take place unconsciously and depends on the strength of the representation, we should observe learning of the C+/shield and C–/no-shield association before learning of the C±/shield association (as C± is less frequently reinforced by the occurrence of the shield than C+).

Experiment 1

Method

Participants

Eighteen students of the first year of psychology in the *Université Libre de Bruxelles* (11 females, 7 males) took part in the experiment. They received course credits for their participation, and were randomly assigned to one of the two experimental groups (N = 9 in each group). None of them was informed of the purpose of the experiment and none of them had previous experience with the task. All participants reported normal or corrected-to-normal vision.

Apparatus and stimuli

The experiment was programmed in Martians V2 software (Franssen et al., 2010) and ran on a Mac Os X 2.6 GHz Intel Core i7. Participants viewed the screen from a distance of 70 cm. Responses were collected through the keyboard. The Unconditioned Stimulus (US) was composed of a metallic sound and the simultaneous presentation of a 0.5 white flashing screen.

FIGURE 5.2 Example of one C, groups of three shapes composing a cue appearing during the task in Experiment 1.

Experimental groups differed in the specific stimuli used as Conditioned Stimuli (C): in half of the participants, stimuli were sounds, whereas in the other half they were sequences of symbols appearing in white on a blue background (Figure 5.2).

Procedure

The experiment lasted about 40 minutes and comprised three parts: a pre-training phase, a US-only phase, and a learning phase.

Pre-training: the experiment was described as a game. Participants were asked to prevent Martians from landing on Earth. Martians appeared on the screen, one by one, every 250ms in rows from the left to the right and from top to bottom. Participants had to press the spacebar to fire a laser gun just before a Martian appeared. Success was indicated by a picture of an explosion instead of the Martian icon. Participants' goal was to destroy as many Martians as possible. During the pre-training phase, participants had the opportunity to learn to produce a regular bar-pressing rate for a period of around 25s, which was used as a baseline in the subsequent phases of the experiment. No C or US were presented during this stage.

US-only phase: in this phase, participants were informed about the protective shield (the US). Participants were told that Martians had developed an anti-laser shield that could appear at unpredictable intervals. The anti-laser shield (i.e., the US) was operationalized as a white flashing screen accompanied by a metallic sound. Within this context, when the shield was active (i.e., during sound and flashes) participants had to stop pressing the spacebar. If participants failed to stop pressing, the laser shot came back at them and a Martian invasion was triggered. In this latter case, invincible Martians (i.e., Martians that were impervious to the laser) appeared on the screen every 50ms (instead of every 250ms) over a period of 5s. However, if participants stopped responding at the shield onset, there was no invasion, but four Martians landed on Earth. During the US-only phase, the shield was presented four times; its onset was not signalled by any other visual or auditory stimulus.

Learning phase: in this phase, the instructions about the US were identical to those in the previous phase. Three different Cs were also presented. Participants were instructed that during the battle, the control station might intercept signals encoded in the Martian language. Participants were told that even though these signals could not be deciphered, they might contain useful information. The participants' goal remained the same in this phase: to prevent Martians from landing on Earth by shooting at them through pressing the spacebar, while avoiding shooting at the shield when it was deployed. The first category of C predicted the shield 100% of the time (C+), the second preceded the shield 50% of the time (C±), and the third was never followed by the shield (C–). The learning phase was divided into 10 blocks, each block consisting of 12 trials (4 trials for each C). Each trial lasted 4, 5, or 10s depending on the C and on participants' performance. After 2s, during which a baseline of bar-pressing was recorded, one of the three Cs was selected and presented during 2s. Depending on the C, the shield then appeared

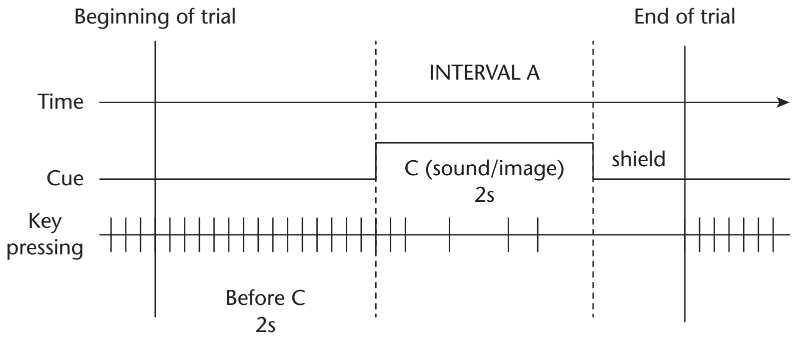

FIGURE 5.3 Schema of the temporal flow of a trial.

for 1s and, depending on participants' performance, an invasion was triggered for a period of 5s (see Figure 5.3). To prevent participants from predicting the onset of the next C, the inter-trial interval varied randomly between 5 and 10s.

After each block of 12 trials, participants also had to indicate whether or not they believed their performance had improved as compared to the previous block. In addition, if participants claimed that their performance had indeed improved, they were also asked to give a subjective written report explaining how they thought they had managed to do better.

Results

Behavioral data

Since the first two phases (pre-training phase and a US-only phase) were only intended to familiarize participants with the task, these data will not be considered for further analysis. In the following, we focus on the learning phase data. To evaluate people's performance, we computed the "suppression ratio" (Arcediano, Ortega, Matute, 1996), that is, the number of presses on the spacebar during the 2s of C presentation divided by the sum of the number of presses during both the C presentation and the 2s immediately before the C presentation (see Figure 5.3).

$$\textit{Suppression Ratio} = \frac{\textit{Responses during } C}{\textit{Responses during } C = \textit{and before } C} \tag{1}$$

During the experiment, participants could learn the C/US relationship but also the duration of the C (i.e., inhibition of delay, Escobar, Suits, Rahn, and Arcediano, 2015; Pavlov 1927; Rescorla, 1967). If they were able to learn how long the stimulus was presented, they could also learn to develop a strategy consisting in pressing the spacebar to hit Martians as long as the C was displayed. In similar previous experiments, learning was measured through test trials in which the duration of the C was increased. In this case, since we measured learning over time, it was not possible to change the duration of presentation of the C. Instead of using the typical suppression ratio described above, we used a more sensitive measure calculated by taking into account the number of key presses during the last 500ms of the C and during the 500ms that preceded C presentation. All Inferential analyses were performed on SPSS Statistics 20 and Bayesian analyses were performed on JASP-software (JASP Team, 2017).

A mixed ANOVA with Group (C sound and C images) as between-subjects factors and Block (Block 1 to 10) and Type of trial (C+, C± and C−) as within-subjects factors was conducted on suppression ratios, yielding a main effect of Type of trial, $F_{(2,\ 32)} = 5.859$, $p<.01$, partial $\eta^2 = .27$. A priori comparisons showed that C− (.31 ± .04) differed from C+ (.21 ± .04; $p = .018$; Cohen's d = −.62) and C± (.095; $p = .029$; Cohen's d = −.56). However, no difference was found between C+ and C± ($p = .12$; Cohen's d = −.39). In order to confirm or disconfirm this result, a Bayesian paired samples t-test was conducted. A BF_{01} of 1.33 was obtained, providing anecdotal evidence for the absence of difference between C− and C±. Analyses failed to show any significant main effect of Group or any interaction involving this factor (all *p*s>.05). A Block × Type of trial interaction was also found, $F_{(18,\ 288)} = 2.671$; $p<.001$, partial $\eta^2 = .143$. Bonferroni-adjusted comparisons revealed that C− trials differed from C+ from the third block until the end of the experiment (all *ps*<.05), with the exception of Blocks 4 and 7. However, the difference between C± and the other two types of trials (C+ and C−) was not consistent (C± vs C− was significant in Blocks 3, 5 8, and 10; and C± vs C+ was significant in Block 9; see Figure 5.4 and Table 5.1 for the descriptive statistics). It seems that participants essentially learned the difference between the two informative types of trials.

Furthermore, a measure of learning can be obtained by comparing C+ trials, in which participants should stop responding, and C− trials, in which participants should continue bar-pressing. Amongst the 18 participants, the learning measure was inconclusive for 6 of them. On the one hand, three of these participants never stopped responding: when any of the three cues were presented, they continued to press the spacebar at the risk of an invasion. On the other hand, the other three participants stopped responding after the presentation of any cue. Hence nothing indicates that these latter participants had learned the difference between the cues (even though they noticed the relationship between a cue onset and the shield appearance) or that they tried to improve their performance.

In the following analysis, we used response rates during the last 500ms of the C as the dependent variable (Figure 5.5). A mixed ANOVA with Block (Blocks 1 to 10) and Type of trial (C+, C± and C−) as within factors was conducted on response rates, yielding a main effect of Type of trial, $F_{(2,\ 34)} = 6.11$, $p = .006$, partial $\eta^2 = .28$, and a Block × Type of trial interaction, $F_{(18,306)} = 2.35$; $p = .002$, partial $\eta^2 = .128$. A priori comparisons showed that C− (1.272 ± .183) differed from C+ (.842 ± .17; $p = .024$; Cohen's d = −.59) and C± (.092; $p = .17$; Cohen's d = −.616). However, no difference was found between C+ and C± ($p = .144$; Cohen's d = −.36). In order

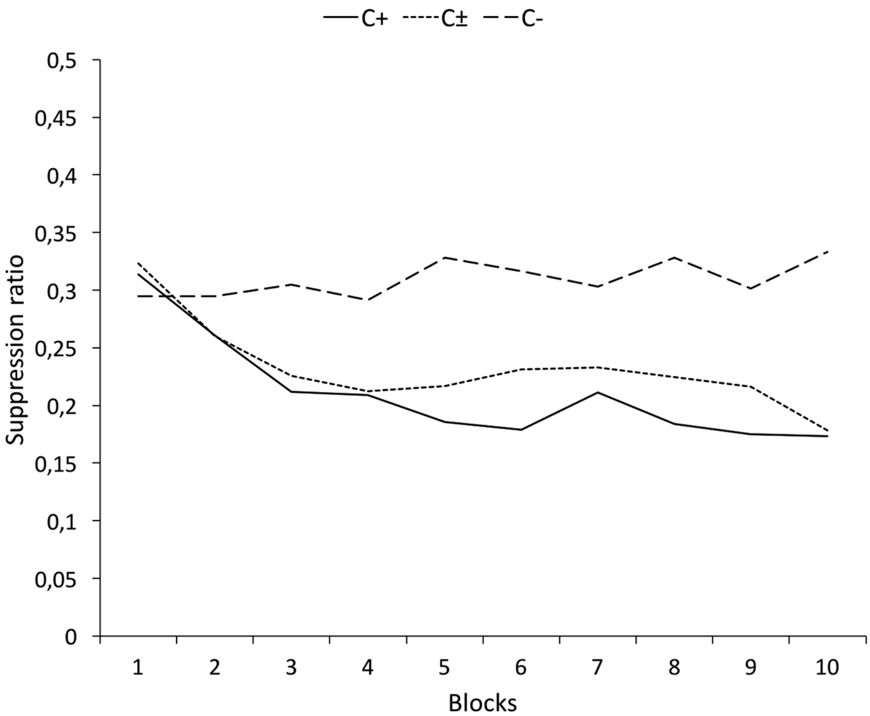

FIGURE 5.4 Mean suppression ratio for each type of C over the course of learning in Experiment 1.

TABLE 5.1 Descriptive statistics of the cues' suppression ratio (means and standard errors) over the blocks in Experiment 1.

	C+	*C–*	*C±*
Block 1	.314 ± .53	.295 ± .048	.323 ± .045
Block 2	.261 ± .05	.295 ± .048	.260 ± .053
Block 3	.212 ± .051	.305 ± .051	.225 ± .051
Block 4	.209 ± .045	.291 ± .048	.212 ± .044
Block 5	.185 ± .052	.328 ± .050	.217 ± .051
Block 6	.179 ± .048	.317 ± .049	.231 ± .048
Block 7	.211 ± .05	.303 ± .049	.233 ± .041
Block 8	.184 ± .044	.328 ± .049	.225 ± .051
Block 9	.174 ± .047	.301 ± .051	.216 ± .048
Block 10	.173 ± .047	.333 ± .049	.178 ± .04

to confirm or disconfirm this result, a Bayesian paired samples t-test was conducted. A BF_{01} of 1.531 was obtained, providing anecdotal evidence for the absence of difference between C– and C±. Analyses failed to show any significant main effect of Group or any interaction involving this factor (all *p*s>.05).

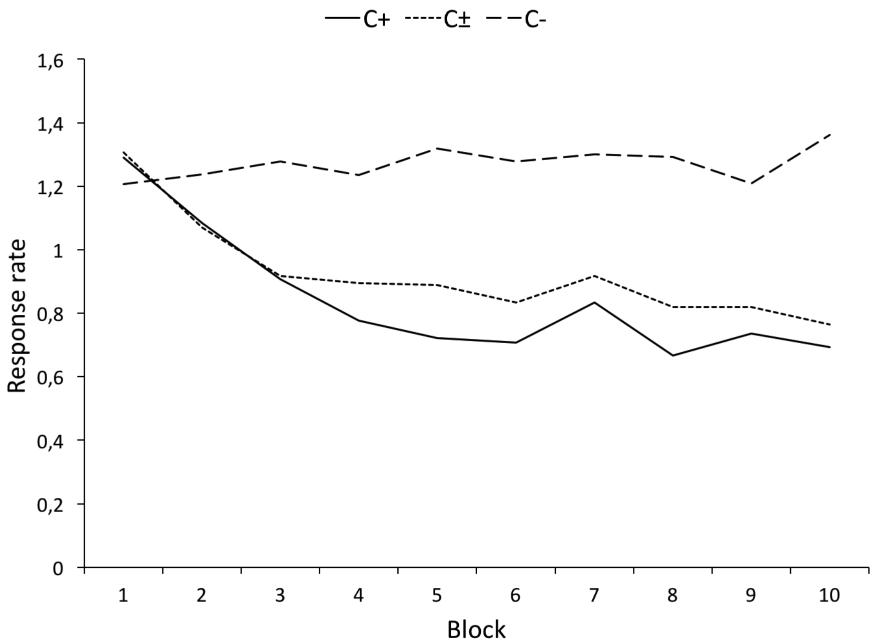

FIGURE 5.5 Mean response rate for each type of C over the course of learning in Experiment 1.

Response rates were also analyzed throughout the C presentation: the duration of the CS was divided into four slices of 500ms and mean response rates were computed for each slice. A mixed ANOVA with Slice (4 levels) and Type of trial (C+, C± and C–) as within-subject factors was conducted on response rate. We obtained a main effect of Slice, $F_{(3,\ 51)} = 26.365$, p<.001, partial $\eta^2 = .61$. Bonferroni-adjusted comparisons revealed that Slices 1 (1.91 ± .069) and 2 (1.7 ± .146) differed from Slices 3 (1.42 ± .181; p = .01 and p = .034 respectively) and 4 (.94 ± .163; all ps<.001). In addition, participants stopped to respond often on Slice 4 than on Slice 3 (p = .004). Furthermore a main effect of the Type of trial was obtained, $F_{(2,\ 34)} = 7.64$, p<.002, partial $\eta^2 = .31$. A priori comparisons revealed that C– (1.68 ± .14) differed from C+ (1.38 ± .13; p = .009; Cohen's d = –.696) and C± (1.41 ± .14; p = .006; Cohen's d = –.74). Again, C+ and C± failed to differ from each other (p=.55; Cohen's d = –.14). A Bayesian paired samples t-test was conducted and revealed moderate evidence for the absence of difference between C– and C± ($BF_{01} = 3.5$). Finally, a Slice × Type of trial interaction was found, $F_{(6,102)} = 4.56$, p<.001, partial $\eta^2 = .21$. Bonferroni-adjusted comparisons revealed that the type of trial did not significantly differ during the two first slices (all ps>.05, see Figure 5.6). However, in the third slice only cue C+ differs from C–, p< .05; participants continued to respond more when C– than when C+ was present. In Slice 4, C– differed from the other two types of trials, C+ (p = .014) and C± (p = .01; see Figure 5.6 and Table 5.2). Overall, this suggests that participants continue to respond in the presence of C– more than when C+ or C± where presented.

Altogether, these results show that participants learned the duration and the identity of the stimulus, as differences between the different stimuli are found from Slice 3. In addition, the data also indicate that even with the secure cue, C–, participants decreased their response rate. This suggests that participants used conservative strategies at cues' onset: whatever the identity of the cue, they tended to decrease their response rate.

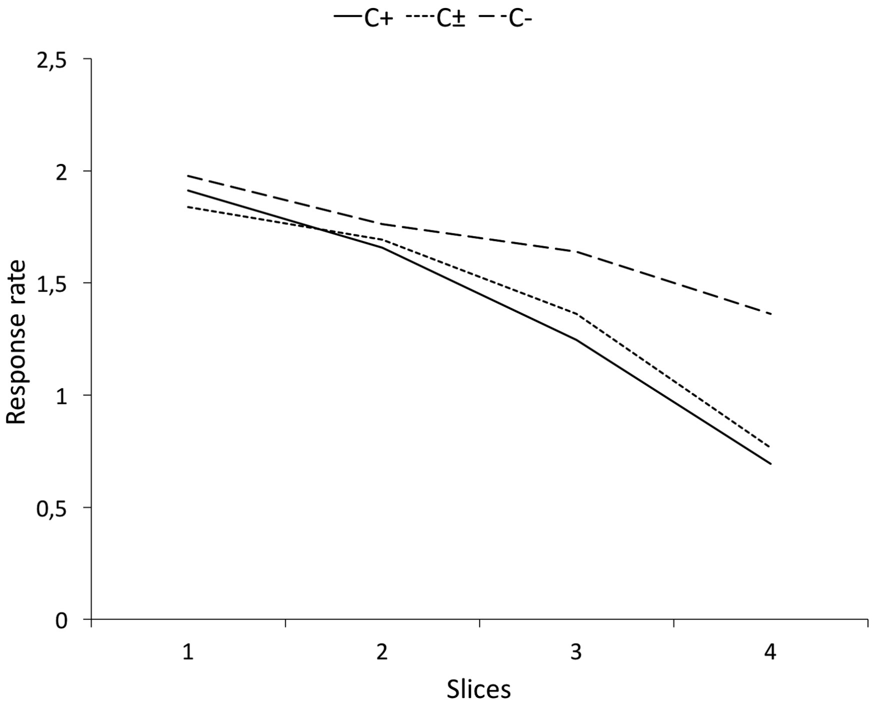

FIGURE 5.6 Mean response rate for each type of C over the slices of 500ms during the presentation of the Cs in Experiment 1.

TABLE 5.2 Descriptive statistics of cues' response rates (means and standard errors) over the slices in Experiment 1.

	C+	*C–*	*C±*
Slices 1	1.91 ± .05	1.98 ± .09	1.84 ± .09
Slices 2	1.66 ± .17	1.76 ± .13	1.69 ± .17
Slices 3	1.25 ± 1.19	1.64 ± .2	1.36 ± .21
Slices 4	.69 ± .19	1.36 ± .21	.76 ± .17

Measures of subjective awareness

Participants' verbal responses were analyzed qualitatively (see Table 5.3). Among the eighteen participants, seven did not show any knowledge about the cues–outcome relationship. Three of these participants failed to express any learning on the behavioral measure (i.e., they did not stop responding after the presentation of the C); their suppression ratios were .5, whatever the identity of the cue. Results of the others four participants did not indicate a significant difference between C+ and C−. Only one participant showed a clear difference in response ratios for the different cues: for C+ (.125), for C− (.5), and for C± (.083). This single participant learned the cue–outcome relationship even if he did not exhibit any explicit knowledge of these contingencies.

Concerning the eleven participants who verbally expressed relevant knowledge of the cues, seven said, starting from the first block, that the cues were useful in helping them improve their performance. Two of them became aware of the relevance of the cues in Block 2, one in Block 5, and one in Block 7.

TABLE 5.3 Description of subjective awareness in Experiment 1.

Subject	*SR C−*	*SR C+*	*Learning?*	*Verbal Report*
1	0.5	0.5	No	/
2	0.167	0	Yes	Strategy: better not respond when a cue is presented
3	0.29	0.33	No	Cues do something, it is random
4	0	0	No	Cues do something
5	0	0	No	Cues do something Strategy: better not respond when a cue is presented
6	0.475	0	Yes	Cues do something and there are different cues
7	0.5	0.08	Yes	Strategy: better not respond when a cue is presented There are different cues
8	0.54	0	Yes	Cues do something
9	0.475	0.5	No	/
10	0.525	0.375	Yes	Strategy: better not respond when a cue is presented There are different cues
11	0.5	0.125	Yes	/
12	0	0.1	No	Strategy
13	0.5	0.5	No	/
14	0.08	0	No	/
15	0.38	0.29	Yes	Strategy
16	0.25	0.08	Yes	/
17	0.5	0	Yes	Cues do something There are different cues
18	0.31	0.225	Yes	/

Nevertheless, only three participants specifically indicated that each cue was different, i.e., that not all the cues predicted a similar outcome. Interestingly, ten of the participants explained their explicit strategies used to perform the task. First, some said that they preferred to stop responding when a cue appeared, whatever the cue identity. Others argued that they tried to finish off bar-pressing until the cue disappeared to destroy as many Martians as possible. Finally, four participants developed superstitions such as "before an invasion the scenery becomes darker"; "After one out of two the shield arrives".

Discussion

The results of Experiment 1 suggest that, on average, participants learn gradually over blocks. Behavioral data cannot determine whether there was an acquisition of complex learning. While participants discriminated between C− vs C+ cues and C− vs C± cues, they seemed to fail to discriminate between C+ and C±. Even if some participants learned that the cues were important in the task, they did not necessarily learn to differentiate between C+ and C±. Another potential explanation would be that participants decided to not respond whenever C± was presented. In other words, a decision made about C± (i.e., predicting the outcome only 50% of the time) would yield more uncertainty than a decision made about C+ or C− (i.e., always predicting the presence or absence of the outcome). The latter two are much more informative in a pre-asymptotic situation. This uncertainty may disappear if participants treated the cue C± as C+. The problem is that if participants decided to take the risk by responding when C± was present, two effects could have occurred: either the outcome did not appear or it did. In the first case, participants continued to respond and kill all the Martians, so they improved their performance. In the latter case, the outcome appears and the losses are much greater than if they had not reacted. Such a conservative strategy is part of the error management theory (Haselton and Buss, 2000; Haselton and Nettle, 2006). To illustrate this point, we need to imagine what happens when someone in a forest sees a large and elongated object hidden in a bush. It could either be a branch or a snake. In such a situation, it is safer to think that it is a snake and run away, even if it is merely a branch. However, to opt for the branch hypothesis implies taking the risk of being bitten if it is in fact a snake. Consequently, for all these reasons, we are unable to discriminate between a failure of discrimination of the cues or an adaptive strategy that appears through learning.

In this study, the use of a verbal report task at different points during the learning phase could influence learning. As this procedure requires participants to reflect about what has happened in the task so far, it cannot be considered as a purely incidental learning task. Each request for participants to report drives attention, and potentially awareness, to the relevant knowledge. In addition, when participants responded to the first question, it is possible that they did not have any knowledge about the relationship between the cue and the outcome during the task itself. However, once the question was asked, they could reason *a posteriori* about the task during the preceding block. Furthermore, as the cues were very salient, the first question would direct the participants' attention towards them in the rest of the experiment. These results are also in line with the artificial grammar literature (Reber, 1989). After a learning phase of complex structures, participants performed more poorly when explicit instructions rather than neutral instructions were presented. Looking for complex rules, i.e., rules participants are not likely to find, produces a lower quality of learning (Reber, 1976). Moreover, participants invented rules trying to explain the underlying structures that were presented. In this study, analysis of verbal reports also indicates that while some participants used explicit strategies, others indicated using superstition and other irrational beliefs (such as those used by sportsmen in the context of sport competitions; see Albas and Albas, 1989 for a review). For example, one participant developed the misconception that the shield appeared after every second sound and only when the second sound was the identical with or higher-pitched than the previous sound.

In Experiment 2, our goal was to minimize the influence of an ambiguous but salient cue (C±) and of the requirement to provide verbal reports in the course of the learning phase. Firstly, only relevant cues (C+ and C−) were presented in order to avoid uncertainty, complexity and the induction of explicit strategies such as breaking off responding when any cue was presented. Likewise, during the learning phase, subjective measures of consciousness were included in the form of a predictive question (asked by the control tower according to the cover story) concerning the possible appearance of the shield. We also asked participants to indicate their confidence in their prediction. In addition, the relevant cues (C+ and C−) were only auditorily presented but, in order to make them less salient, continuous music simulating Martian speech was presented during all the experimental phases. Finally, to test the potential influence of attention to the cues, we included an incidental group, who were not given any specific information about the cues, and an intentional group who were informed that there were some relevant but undeciphered signals encoded into the Martian speech stream. As the relevant cues were totally predictive of the outcome or the absence of it, instructions should focus subjects' attention and it should accelerate learning (Reber, 1976).

Experiment 2

Method

Participants

Twenty-eight participants (8 females, 20 males, mean age 22.75 ± 2.9) voluntarily took part in the experiment. None was informed of the purpose of the experiment and none had previous experience with the task. All participants reported normal or corrected-to-normal vision. Participants were divided into two groups depending on the instructions (14 participants in the incidental and intentional group).

Apparatus and stimuli

The apparatus was identical to that used in Experiment 1. The Conditioned Stimuli were two sounds (246 Hz and 320 Hz) counterbalanced as C+ and C−. During the entire experiment a musical piece composed of a random presentation of sounds between 130Hz and 523Hz was played.

Procedure

The experiment consisted of five phases: a pre-training phase, a US-only phase, a control tower phase, an instruction phase and a learning phase. The music was presented throughout the entire experiment. The pre-training and US-only phases were strictly identical to those used in Experiment 1 and will thus not be described again here.

Control tower phase: participants were instructed that on some occasions during the Martian attack, one of the staff in the control tower may be needing assistance. On such occasions, the control tower asked the participant whether or not he/she believed that the shield would appear. Participants were first required to indicate whether or not they believed the shield would appear next, and then whether or not they were confident in their response. To introduce this phase to the participants, Martians were displayed for a period of 25s, and the questions were then presented.

Instruction phase: the participants' goal remained the same in this phase: to prevent Martians from landing on Earth while refraining from shooting when the shield was active. Half of the participants (the intentional group) was instructed that the control station might receive undeciphered, but potentially informative, audio signals encoded in Martian language. The other half (the incidental group) did not receive any instructions concerning these audio signals.

Learning phase: in this phase, two types of C were embedded in the music. One was a perfect predictor of the shield (C+). The second was never followed by the shield (C−). The learning phase comprised 2 types of randomly sequenced trials: 70 learning trials (35 C+ and 35 C− trials) and 30 control tower trials (10 C+, 10 C− and 10 no C). In each of these 30 control tower trials participants had to indicate whether they expected the shield to occur following either the C+ or the C−, or neither of these two stimuli. This phase was divided into a learning and a stabilization phase, each one composed of 17 trials for each cue.

During any learning trial, we recorded the bar-pressing rate for 2s before presenting C+ or C− for 2s, during which the bar-pressing rate was also recorded. As in Experiment 1, shooting at the shield would cause an alien invasion for 5s. A varying delay between 5 and 10s elapsed between any two trials.

In control tower trials, a question was presented centrally on the screen following either one of the two Cs or none of them. "Do you think the shield will appear now?" Participants had to respond "yes" or "no". Participants were also asked to tell whether they were confident or not in their response. After they responded to the second question, the next trial began.

Results

Behavioral data

As in Experiment 1, we computed the suppression ratio by dividing (a) the number of key presses during the last 500ms of the C by the sum of (a) and (b), the number of key presses during the final 500ms before the onset of the C. We first report results for the learning trials. In order to study the dynamics of performance, the learning phase was divided into two parts in the analysis: a first, learning, phase and a second, stabilization, phase. The first trial was discarded for all participants. All Inferential analyses were performed with SPSS Statistics 20. Bayesian analyses were carried out with JASP- software (JASP Team, 2017).

A mixed ANOVA with Group (intentional vs incidental) as a between factor and Block (learning vs stabilization) and Type of trial (C+ and C–) as within factors was conducted on suppression ratios, yielding a main effect of Type of trial, $F_{(1,26)} = 25.909$, $p<.001$, partial $\eta^2 = .499$. When C– (.359 ± .027) was presented participants pressed the spacebar more often than when C+ (1.8 ± .029) was presented. The interaction Type of trial × Block, $F_{(1,26)} = 16.483$, $p<.001$, partial $\eta^2 = .388$ was significant. Bonferroni-adjusted comparisons showed that participants learned to stop responding when C+ was presented during learning (learning phase: .22 ±.03; stabilization phase: .143; $p = .006$). Nevertheless, participants responded similarly to C– during the learning phase (.35± .03) as during the stabilization phase (.37 ± .03; $p = .572$); $BF_{01} = 4.39$ providing moderate evidence for the absence of difference. Finally, the triple Type of trial × Block × Group interaction was significant $F_{(1,26)} = 4.283$, $p<.05$, partial $\eta^2 = .141$. Bonferroni-adjusted comparisons revealed that the difference between the types of trials (C+ and C–) was significant in the two phases in the intentional group (all *p*s<.001). In contrast, there was no difference in suppression ratios between C+ and C– in the first learning phase ($p = .211$; $BF_{01} = 3.61$) in the incidental group. This difference only reached significance in the stabilization phase ($p<.005$; Figure 5.7).

Measures of subjective awareness

In order to find out whether there was a response bias (i.e., whether participants systematically preferred one option above the other, "guess" or "remember"), we computed the C statistic, developed based on the Signal Detection Theory (Macmillan and Creelman, 1990). This index amounts to the total (standardized) number of trials where participants claimed to remember (hits and false alarms) divided by two (Paredes-Olay, Moreno-Fernández, Rosas, and Ramos-Álvarez, 2010). Values below zero indicate a liberal bias (i.e., participants claimed to remember more often than they claimed to guess), while values above zero indicate a conservative bias (i.e., preference for claiming to guess rather than claiming to remember). To assess the extent to which participants were aware of the acquired knowledge, we focused on three indexes: the guessing criterion (Cheesman and

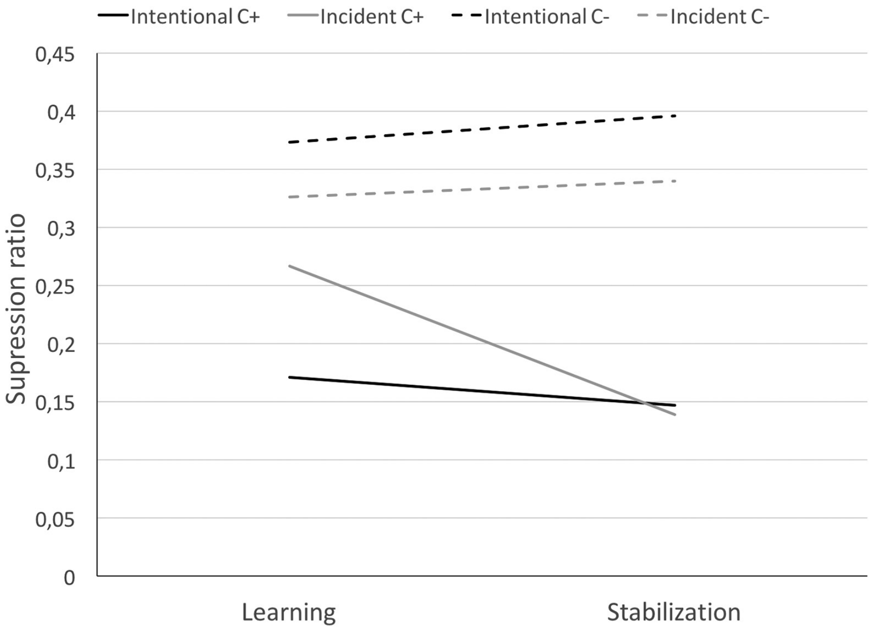

FIGURE 5.7 Mean suppression ratio for each group and for each type of C over the course of learning in Experiment 2.

Merikle, 1984), Type-II *d'* scores (Macmillan and Creelman, 2005) and the zero correlation criterion (Dienes, Altmann, Kwan, and Goode, 1995). The guessing criterion (Cheesman and Merikle, 1984) is met when participants perform above chance level while they claim to guess. To compute the criterion, we calculated the proportion of correct responses when participants reported to guess. A positive result would suggest that learning was at least partly unconscious.

The Type-II *d'* score (Macmillan and Creelman, 2005) and the zero correlation criterion (Dienes, Altmann, Kwan, and Goode, 1995) are measures of the relation between confidence and accuracy. The Type-II *d'* score is based on the Signal Detection Theory. In the context of our study, a "hit" corresponds to a correct response given with confidence, and a "false alarm" corresponds to a confident error. A Type 2 *d'* that represents the difference between these two (standardized) scores may then be computed. Positive values indicate a match between accuracy and confidence and suggest that performance was based on conscious knowledge. The zero correlation criterion (Dienes, Altmann, Kwan, and Goode, 1995) is the difference between the numbers of hits and false alarms. A difference reliably above chance level would indicate at least partly conscious learning.

In the learning phase, the *C* statistic was significantly different from 0 in the intentional group for no C (M = −2.41, SD = 1.40, $t_{(13)} = -3.75$, $p<.01$, Cohen's d = 1), and marginal for C+ (M = −1.62, SD = 2.89, $t_{(13)} = -2.1$, $p = .056$, Cohen's d = .56), indicating that participants exhibited a liberal bias (i.e., preference for claiming to remember rather than claiming to guess). However, a Bayesian one sample t-test revealed anecdotal evidence for the absence of bias for C– (BF_{01} = 2.4).

Intentional participants were more confident in their correct responses than in their errors. This effect was significant for both C+ ($t_{(13)}$ = 3.99, p<.01, Cohen's d = 1.066) and no C ($t_{(13)}$ = 4.58, p<.001, Cohen's d = 1.225). For C–, a Bayesian paired samples t-test provided anecdotal evidence for the absence of participants' confidence in correct responses vs their errors (BF_{01} = 1.29). Similar results were observed with the Type-II scores (C+, $t_{(13)}$ = 3.864, p<.005, Cohen's d = 1.03; no C, $t_{(13)}$ = 4.02, p<.001, Cohen's d = 1.07; C–, BF_{01} = 1.25). However, in the learning phase, performance and confidence were not related to each other in the incidental group (all ps>.05; C–, BF_{01} = 3.7; no C, BF_{01} = 3.59; but C+, BF_{01} =0.88). In addition, in our incidental group, the *C* statistics were not significantly different from 0, *(*all ps>.05; C+, BF_{01} = 2.02; C–, BF_{01} = 3.7; no C, BF_{01} = 3.69*)*, indicating the absence of any response bias. Regarding the Guessing criterion, participants in both groups were not above chance level (.5) when claiming to guess during the learning phase, whatever the identity of the preceding cue (for the ps and *BF* see Table 5.4).

In the stabilization phase, the *C* statistic was significantly different from 0 in the intentional group for all types of cues (no C: M = –2.42, SD = 3.47, $t_{(13)}$ = –2.618, p<.05, Cohen's d = .7; for C+ : M = –3.07, SD = 2.16, $t_{(13)}$ = –5.32, p<.001, Cohen's d = 1.29; and C– : M = –2.44, SD = 3.1, $t_{(13)}$ = –2.94, p<.05, Cohen's d = 1.16) indicating that participants had a liberal bias. Intentional participants were also more confident in their correct than in their incorrect responses, both after C+ and no C ($t_{(13)}$ = 5.99, p<.001, Cohen's d = 1.6 and $t_{(13)}$ = 2.52, p<.05, Cohen's d = .67, respectively). Furthermore, in that latter phase, the intentional group was also able to recognize that C– predicted the absence of the shield ($t_{(13)}$ = 2.59, p<.05, Cohen's d = .69). These results were corroborated by the Type-II scores (no C, $t_{(13)}$ = 2.543, p<.05, Cohen's d = .68; C+, $t_{(13)}$ = 5.949, p<.001, Cohen's d = 1.59; C–, $t_{(13)}$ = 2.619, p<.05, Cohen's d = .7).

TABLE 5.4 Description of *p* values and *BF* of the comparison of guessing criterion in Experiment 2.

Group	Phase		C+	C-	No C
Intentional	Learning	p	945	1	.959
		BF_{01}	8.5	14.61	9.09
	Stabilization	p	.694	.997	.987
		BF_{01}	5.19	11.67	10.62
Incidental	Learning	p	.605	.97	.291
		BF_{01}	4.47	9.34	2.31
	Stabilization	p	.976	.812	.989
		BF_{01}	*9.64*	*6.37*	*10.53*

In the incidental group, the *C* statistic was significantly different from 0 for no C (M = −2.27, SD = 2.65, $t_{(13)}$ = −3.21, p<.01, Cohen's d = .86), and marginal for C+ (M = −1.9, SD = 3.43, $t_{(13)}$ = −2.07, p = .059, Cohen's d = .63), indicating that participants had a liberal bias. However, a Bayesian one sample t-test provided anecdotal evidence for the absence of bias on C− (BF_{01} = 2.11). In the incidental group, participants became more confident in their correct responses than in their errors in the stabilization phase; both after C+ and no C ($t_{(13)}$ = 2.355, p<.05, Cohen's d = .63 and $t_{(13)}$ = 2.882, p<.05, Cohen's d = .77, respectively) but not after C− (BF_{01} = 1.94). The same results were obtained using the Type-II scores (no C, $t_{(13)}$ = 2.984, p<.05, Cohen's d = .8; C+, $t_{(13)}$ = 2.374, p<.05; Cohen's d = .63; C− , $t_{(13)}$ = 1.112, p = .09, Cohen's d = .29; BF_{01} = 2.2). Concerning the guessing criterion, participants were at chance in both groups when claiming to guess whether or not the shield would appear next whatever the identity of the preceding cue (for the *p*s and *BF* see Table 5.4).

Discussion

The results of Experiment 2 suggest that learning is gradual. However, some further factors also influence the dynamics of learning, such as instructions and the inherent complexity of the association, which in turn influences metaknowledge. In particular, we observed differences between the intentional and the incidental groups. During the learning phase, the intentional group exhibited a clear behavioral difference between their sensitivity to C+ and C− conditions: participants kept pressing the spacebar more often when C− was presented than when C+ was presented, thus reflecting their understanding that when the cue was absent (no C), nothing happened. However, when a cue was displayed, they only seemed to have conscious knowledge about the relationship between the occurrence of C+ and the onset of the shield. Nevertheless, they were able to express some metaknowledge about C− and the absence of the shield in the second phase of learning.

Interestingly, the incidental group showed a delayed pattern of behavior in comparison with the intentional group. Without instructions, participants failed to show any behavioral, explicit, or implicit learning in the first phase. In the stabilization phase, they were able to differentiate between the different cues, but they only expressed explicit knowledge about the absence of any outcome in the absence of a cue, and about the relationship between C+ and the appearance of the shield.

Overall, our data suggest that it was easier for participants to represent the link between a cue and an outcome than between a cue and the absence of this outcome. This result could be explained by associative theories, which predict learning only when the two linked events, a cue and an outcome, are actually presented. For example, in the Rescorla and Wagner (1972) model, the more an outcome is unpredictable, the more learning can occur. In this model, the two stimuli are needed to reinforce the association between them (but see Van Hamme and Wasserman, 1994). In our experiment, participants failed to exhibit explicit learning about C− and the absence of the outcome. This is congruent with the predictions of associative theories, since it is not possible to learn when no event is present. However, participants were able to learn and to express explicit knowledge about the circumstances when no cue was presented and nothing would happen, indicating that participants used rules or inferences to perform correctly. Such ruled-based learning is influenced by the amount of learning (one's expertise; Shanks and Darby, 1998). One possible explanation of our results could be that participants had sufficient practice to learn about the fact that only when a cue was present would the outcome appear. Nevertheless, more training is required to develop rules about C− conditions. If learning was driven only by inference and by a rule-based process, participants should not experience difficulties discerning between C+ and C−. In other words, once a participant realized that cues were informative and that in the absence of a cue the shield never appeared, he/she should be able to reason about cues. Thus, the participant could consciously detect that C− is never followed by the shield at the same time that he/she identified that C+ was followed by the shield. This pattern of data fits well with a dual-process theory in which two learning processes influence behavior and consciousness.

The intentional group exhibited improvement in behavior faster than the incidental group, which needed more evidence to learn. This suggests that instructions also modulate the speed of behavioral and conscious learning. Likewise, the intentional group seemed to achieve metaknowledge of the relationship between cues quicker than the non-instructed group, who only demonstrated some explicit knowledge in the stabilization phase. These results are in line with earlier observations showing that instructions influence learning (Lovibond, Been, Mitchell, Bouton, and Frohardt, 2003; Mertens, Kuhn, Raes, Kalisch, De Houwer, and Lonsdorf, 2016; Mitchell and Lovibond, 2002; Raes, De Houwer, De Schryver, Brass, and Kalisch, 2014; Sternberg and McClelland, 2012; Waldmann and Holyoak, 1992, 1997). In addition, our behavioral and consciousness results fit well with the statistical learning literature, showing that instructions affect the amount of participants' knowledge and metaknowledge of a structure. In intentional conditions, participants could use explicit strategies to learn the contingencies of the material, and they acquired more knowledge and more metaknowledge than participants in incidental conditions (Arciuli, Torkildsen, Stevens, and Simpson, 2014; Bertels, Destrebecqz, and Franco, 2015; Turk-Browne, Junge, and Scholl, 2005). Explicit instructions were sufficient to modify participants' behavior (Mitchell and Lovibond, 2002), which represents a challenge to associative learning theories: if learning is only driven by the cumulative presentation of contingencies, we should not observe differences between intentional and incidental groups.

Nevertheless, interpretations of the results of the measures of consciousness need to be taken with caution. During the control tower trials, participants were first required to indicate whether or not they believed the shield would appear next, and then whether or not they were confident in their response. The confidence measures refer to the judgment they had just provided and not about the cause of the appearance of the shield. It is still possible that participants remained unaware about the relation between the cue and the outcome, even if our criteria provided evidence of metaknowledge. For example, a participant could know that the shield would appear and he could be certain of this while remaining unable to explain the reason why he responded he was sure that the shield would appear (Dienes and Scott, 2005). Further experiments are needed to disentangle the dynamics of learning and consciousness in these situations.

General discussion

Across the two experiments, the dynamics of learning was measured to address the question of whether learning always depends on rule acquisition, or whether it can take place gradually through repeated exposure to the paired events. A dual-process theory, where the two processes of learning, associative and inferential, is consistent with the results of our two experiments. Learning initially seems to be gradual and to some extent unconscious. This associative learning process drives behavioral changes that participants end up noticing with training. Noticing the association results in further changes in behavior that are accompanied by conscious awareness. However, this associative process could not explain all the results. Behavior was also influenced by top-down information. For instance, instructions and participants' awareness of the task influenced learning. This evidence predisposes participants' attention to the relevant cues, inducing faster learning. This result suggests that procedural and declarative representations develop in parallel and interact with each other during learning (Shanks, 2007). Humans have the ability to learn on a trial-by-trial error-correction basis and, at the same time, they are able to reason about the situation, so creating expectancies about the process underlying the task.

In addition, in our experiment, participants were asked to produce a predictive judgment that could be biased by reasoning. Participants could correctly encode the relationship between the cues and the shield, as they were sensitive to the contingencies between them. However, when they were asked to produce a predictive judgment, it is possible that they expressed not only the relevant knowledge they had acquired, but they could also manifest a biased reasoning about the task. Actually, participants could track and extract contingencies even if they are also biased when they are asked to express judgments (Allan, Siegel, and Tangen, 2005; Ratliff and Nosek, 2010; Vadillo, Blanco, Yarritu, and Matute, 2016). This possible explanation matched with a dual-process theory. Interestingly, in our studies, dissociations between behavior and explicit learning were found. These dissociations could also constitute evidence for the coexistence of two learning processes (McLaren et al., 2014; McLaren, Green, and Mackintosh, 1994).

Dissociations between behavior and reasoning are already documented in other forms of learning. For instance, it is possible to dissociate the distinct influences of automatic associative learning and conscious expectancies on behavior (Perruchet, 1985). In addition, in the implicit learning literature, it is well known that participants can learn gradually while remaining unable to verbalize the rules underlying the task (for instance, in Serial Reaction Time tasks, Destrebecqz and Cleeremans, 2001; or in associative learning paradigms, Alamia et al. 2016). To sum up, learning in this task seems to be driven by the interactive relationships of automatically learned associations between events and declarative knowledge. Further experiments are needed to measure the separate contributions of each process during learning and to explore the relation between the two processes.

References

Alamia, A., De Xivry, J. J, San Anton, E., Olivier, E., Cleeremans, A., and Zenon, A. (2016). Unconscious associative learning with conscious cues. *Neuroscience of Consciousness*, *1*, 1–10.

Albas, D. and Albas, C. (1989). Modern magic: the case of examinations. *Sociological Quarterly*, *30*, 603–613.

Allan, L. G., Siegel, S., and Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior*, *33*(2), 250–263.

Arcediano, Francisco, Ortega, Nuria, Matute, Helena (1996). A behavioural preparation for the study of human Pavlovian conditioning. *The Quarterly Journal of Experimental Psychology*, *49B* (3), 270–283.

Arciuli, J., Torkildsen, J., Stevens, D. J., and Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, *5*, 747. doi:10.3389/fpsyg.2014.00747.

Asli, O. and Flaten, M. A. (2012). In the blink of an eye: investigating the role of awareness in fear responding by measuring the latency of startle potentiation. *Brain Science*, *2*(1), 61–84. doi:10.3390/brainsci2010061.

Atas, A., Faivre, N., Timmermans, B., Cleeremans, A., and Kouider, S. (2014). Nonconscious learning from crowded sequences. *Psychological Science*, *25*(1), 113–119. doi:10.1177/0956797613499591.

Baeyens, F., Eelen, P., and van den Bergh, O. (1990). Contingency awareness in evaluating conditioning: a case for unaware affective-evaluative learning. *Cognition and Emotion*, *4*, 3–18.

Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Beckers, T., De Houwer, J., Pineño, O., and Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 238–249.

Bertels, J., Destrebecqz, A., and Franco, A. (2015). Interacting effects of instructions and presentation rate on visual statistical learning. *Frontiers in Psychology*, *6*, 1806. doi:10.3389/fpsyg.2015.01806.

Bouton, M. E. (2007). *Learning and Behavior*. Sounderland, MA: Sinauer.

Cheesman, J. and Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36*(4), 387–395.

Clark, R. E. and Squire, L. R. (1998). Classical conditioning and brain systems: the role of awareness. *Science*, *280*, 77–81.

Clark, R. E., Squire, L. R. (1999). Human eyeblink classical conditioning: effects of manipulating awareness of the stimulus contingencies. *Psychological Science*, *10*(1), 14–18.

Cleeremans, A. (2008). Consciousness: the radical plasticity thesis. *Progress in Brain Research*, *168*, 19–33. doi:10.1016/S0079-6123(07)68003-0.

Cleeremans, A., Destrebecqz, A., and Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Science*, *2*(10), 406–416.

De Houwer, J. (2014). Why a propositional single-process model of associative learning deserves to be defended. In B. G. J. W. Serman and Y. Trope (Eds.), *Dual Processes in Social Psychology* (pp. 530–541). New York, NY, USA: Guilford.

Destrebecqz, A. and Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*(2), 343–350.

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

Dienes, Z. and Scott, R. B. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*, 338–351.

Escobar, M., Suits, W. T., Rahn, E. J., Arcediano, F. (2015). Do long delay conditioned stimuli develop inhibitory properties? *Frontiers in Psychology*, *6*, 1606.

Fiser, J. and Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

Franssen, M., Clarysse, J., Beckers, T., van Vooren, P. R., and Baeyens, F. (2010). A free software package for a human online-conditioned suppression preparation. *Behavior Research Methods*, *42*(1), 311–317. doi:10.3758/BRM.42.1.311.

Haselton, M. G. and Buss, D. M. (2000). Error management theory and the evolution of misbeliefs. *Behavioral and Brain Sciences*, *32*, 522–523.

Haselton, M. G. and Nettle, D. (2006). The paranoid optimist: an integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47–66. doi:10.1207/s15327957pspr1001_3.

JASP TEAM (2017). JASP (version 0.8.1.2) [Computer software].

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., and Sloman, S. A. (2007). Beyond covariation: cues to causal structure. In A. Gopnik and L. Schultz (Eds.), *Causal Learning: Psychology, Philosophy, and Computation*. New York, NY, US: Oxford University Press.

Lovibond, P. F., Been, S., Mitchell, C. J., Bouton, M. E., and Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory & Cognition*, *31*, 133–142.

Lovibond, P. F. and Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3–26.

Macmillan, N. A. and Creelman, C. D. (1990). Response bias: characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin*, *107*, 401–413.

Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory: a User's Guide*. Mahwah, NJ, US: Lawrence Erlbaum Associates.

Maia, T. V. and McClelland, J. L. (2004). A re-examination of the evidence for the somatic marker hypothesis: what participants know in the Iowa gambling task. *Proceedings of the National Academy of Sciences*, *101*, 16075–16080.

Maia, T. V. and McClelland, J. L. (2005). The somatic marker hypothesis: still many questions but no answers. *Trends in Cognitive Sciences*, *9*(4), 162–164.

McLaren, I. P., Forrest, C. L., McLaren, R. P., Jones, F. W., Aitken, M. R., and Mackintosh, N. J. (2014). Associations and propositions: the case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory*, *108*, 185–195. doi:10.1016/j.nlm.2013.09.014.

McLaren, I. P., Green, R. E., and Mackintosh, N. (1994). Animal learning and the implicit/explicit distinction. In N. C. Ellis (Ed.), *Implicit and Explicit Learning of Languages*. London, United Kingdom: Academic Press.

Meltz, E. R., Cheng, P. W., Holyoak, K. J., and Waldmann, M. R. (1993). Cue competition in human categorization: contingency or the Rescorla–Wagner learning rule? Comment on Shanks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1398–1410.

Mertens, G., Kuhn, M., Raes, A. K., Kalisch, R., De Houwer, J., and Lonsdorf, T.B. (2016). Fear expression and return of fear following threat instruction with or without direct contingency experience. *Cognition and Emotion*, *30*, 968–984. doi: 10.1080/02699931.2015.1038219.

Miller, R. R., and Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22, pp. 51–92). San Diego, CA: Academic Press.

Mitchell, C. J., De Houwer, J., and Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Science*, *32*(2), 183–198; discussion 198–246. doi:10.1017/S0140525X09000855.

Mitchell, C. J., and Lovibond, P. F. (2002). Backward and forward blocking in human electrodermal conditioning: blocking requires an assumption of outcome additivity. The *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, *55*(4), 311–329. doi:10.1080/02724990244000025.

Paredes-Olay, C., Moreno-Fernández, M. M., Rosas, J. M., and Ramos-Álvarez, M. M. (2010). ROC analysis in olive oil tasting: a signal detection theory approach to tasting tasks. *Food Quality and Preference*, *21*, 562–568.

Pavlov, I. P. (1927). *Conditioned Reflexes*. London: Oxford University Press.

Pearce, J. M. and Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, *52*, 111–139. doi:10.1146/annurev.psych.52.1.111.

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. The *Pavlovian Journal of Biological Science*, *20*(4), 163–170.

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., and Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*, *59*(4), 561–567. doi:10.1016/j.neuron.2008.07.005.

Raes, A. K., De Houwer, J., De Schryver, M., Brass, M., and Kalisch, R. (2014). Do CS-US pairings actually matter? A within-subject comparison of instructed fear conditioning with and without actual CS-US pairings. *PLoS ONE*, *9*(1): e84888. doi:10.1371/journal.pone.0084888.

Raio, C. M., Carmel, D., Carrasco, M., and Phelps, E. A. (2012). Nonconscious fear is quickly acquired but swiftly forgotten. *Current Biology*, *22*(12), R477–479. doi:10.1016/j.cub.2012.04.023.

Ratliff, K. A. and Nosek, B. A. (2010). Creating distinct implicit and explicit attitudes with an illusory correlation paradigm. *Journal of Experimental Social Psychology*, *46*, 721–728.

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

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

Rescorla, R. A. (1967). Inhibition of delay in Pavlovian fear conditioning. *Journal of Comparative & Physiological Psychology*, *64*, 114–120. doi: 10.1037/h0024810.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative & Physiological Psychology*, *66*(1), 1–5.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non reinforcement In A. H. Black and W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Schultz, D. H. and Helmstetter, F. J. (2010). Classical conditioning of autonomic fear responses is independent of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(4), 495–500. doi:10.1037/a0020263.

Shanks, D. R. (1995). Is human learning rational? *The Quarterly Journal of Experimental Psychology*, *48*(2), 257–279.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts and R. L. Goldstone (Eds.), *Handbook of Cognition* (pp. 202–220). London: Sage.

Shanks, D. R. (2007). Associationism and cognition: human contingency learning at 25. *Q J Exp Psychol (Hove)*, *60*(3), 291–309. doi:10.1080/17470210601000581.

Shanks, D. R. and Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *J Exp Psychol Anim Behav Process*, *24*(4), 405–415.

Shanks, D. R. and Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 21, pp. 229–261). San Diego, Calif: Academic Press.

Sternberg, D. A. and McClelland, J. L. (2012). Two mechanisms of human contingency learning. *Psychological Sciences*, *23*(1), 59–68. doi:10.1177/0956797611429577.

Tabbert, K., Stark, R., Kirsch, P., and Vaitl, D. (2006). Dissociation of neural responses and skin conductance reactions during fear conditioning with and without awareness of stimulus contingencies. *Neuroimage*, *32*(2), 761–770. doi:10.1016/j.neuroimage.2006.03.038.

Turk-Browne, N. B., Junge, J., and Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology General*, *134*(4), 552–564. doi:10.1037/0096-3445.134.4.552.

Vadillo, M. A., Blanco, F., Yarritu, I., and Matute, H. (2016). Single- and dual-process models of biased contingency detection. *Experimental Psychology*, *63*(1), 3–19. doi:10.1027/1618-3169/a000309.

Van Hamme, L. J. and Wasserman, E. A. (1994). Cue competition in causality judgments: the role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*(2), 127–151.

Waldmann, M. R. and Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*, 222–236.

Waldmann, M. R. and Holyoak, K. J. (1997). Determining whether causal order affects cue selection in human contingency learning: comments on Shanks and Lopez (1996). *Memory and Cognition*, *25*, 125–134.

6

CAN UNCONSCIOUS STRUCTURAL KNOWLEDGE BE STRATEGICALLY CONTROLLED?

Elisabeth Norman, Ryan B. Scott, Mark C. Price, Emma Jones, and Zoltan Dienes

Introduction

Strategic control refers to the ability to apply knowledge flexibly in an intentional manner according to current situational demands. If knowledge is under strategic control, this has traditionally been taken as evidence that the knowledge in question is consciously available. Examples include Jacoby (1991), who views strategic control as a criterion of consciousness, and Baars' (1988) Global Workspace Model, according to which conscious information is both controllable and available to higher order thought. However, there are also theories that do not regard strategic control as indicative of consciousness. Higher Order Thought theories do not make claims about the relationship between control and consciousness (Lau and Rosenthal, 2011; Rosenthal, 2005), and the "cold control theory" of hypnosis regards unconscious executive control as characteristic of hypnosis (Dienes and Perner, 2007). In addition, there is now a relatively large body of empirical evidence supporting the hypothesis that strategic control may occur for cognitive content that it is itself not available to consciousness. For example, Lau and Passingham (2007) found that unconsciously perceived stimuli interfered with tasks traditionally thought to require conscious control. Similarly, several studies have reported findings indicating that the go-no go network can be activated unconsciously (Hepler and Albarracin, 2013; Van Gaal, Ridderinkhof, Scholte, and Lamme, 2010), and a study by Schmidt, Crump, Cheesman, and Besner (2007) showed that participants were able to strategically control the application of learned contingencies between colour-unrelated words and colours in a contingency learning paradigm. The focus of the current chapter is on whether the application of unconscious knowledge, acquired through implicit learning, may also be strategically controlled.

Procedures for measuring strategic control in implicit learning

A number of experimental procedures have been developed for measuring strategic control in implicit learning, both within the serial reaction time task (SRT; Nissen and Bullemer, 1987) and within artificial grammar learning (AGL; Reber, 1967). Most of these are based on the logic of Jacoby's (1991) Process Dissociation Procedure, which compares performance under conditions where a person "tries to" versus "tries not to" engage in some act, and where a comparison of performance under the two conditions is seen as indicating the relative influence of conscious and unconscious knowledge.

We here give a brief overview of the most important methods for assessing strategic control in SRT and AGL learning (see Norman, 2015, for a more complete overview).

In the *SRT task*, participants are presented with a visual target that moves between positions on a computer screen according to a complex, pre-defined sequence. The instruction is to make fast key-press responses to indicate the position of the moving target, and reaction time differences between target movements that either follow or violate this sequence are taken to indicate learning. In this paradigm, strategic control refers to participants' ability to control their application of sequence knowledge according to task instructions. The most common measurement of this ability is the *generation exclusion task*, in which participants are instructed to generate a sequence that is different from the sequence on which they have been trained (Destrebecqz and Cleeremans, 2001; Fu, Fu, and Dienes, 2008; Goschke, 1998). Strategic control can be assessed by comparing performance under these instructions and under conditions when participants try to generate the trained sequence (i.e., *inclusion* instructions). Two varieties of the inclusion/exclusion generation task are *free generation*, in which the participants freely generate an *n*-element sequence (e.g., Destrebecqz and Cleeremans, 2001), and *cued generation*, where each trial involves generating a continuation response to a short sequence of, e.g., 3–5 sequence elements. An alternative procedure is the *generation rotation task* (Norman, Price, Duff, and Mentzoni, 2007). This is designed to avoid the possibility that successful exclusion performance could be influenced by a global inhibition of the influence of acquired knowledge, rather than by a moment-to-moment monitoring of this knowledge. During training and generation, stimuli are presented in a square layout. In a cued generation task, participants are instructed to predict the next target position. However, the stimulus–response mapping varies between individual trials. More specifically, participants are told to rotate their response, clockwise or anticlockwise, in accordance with a randomly varying cue (−1, +1, −2) indicated on screen. Yet another procedure is the *inclusion/exclusion recognition task* (Mong, McCabe, and Clegg, 2012), where participants are first trained on two different sequences, and then have to classify a series of unseen sequences according to familiarity. Under *inclusion* instructions, items are to be classified as "old" if they follow either regularity. Under *exclusion* instructions, items are to be classified as "old" if they follow their target sequence, and as "new" if not.

In *AGL*, participants are exposed to a series of non-word letter strings that are constructed from a complex, finite-state grammar (Reber, 1967). Learning is measured as the ability to classify unseen letter strings according to grammaticality, and strategic control refers to the ability to apply or withhold grammar knowledge according to instructions. Most methodological procedures for estimating strategic control involve exposure to two different grammars (A vs. B; Dienes, Altmann, Kwan, and Goode, 1995) in two separate training phases. On each trial of a subsequent test phase, participants are presented with letter strings that either follow one of these two grammars or are ungrammatical and follow neither. Participants may be instructed to classify whether new letter strings follow one specified target grammar throughout the test block (Dienes et al., 1995), which can be referred to as a *pure-block procedure* (Norman, Price, and Jones, 2011). Alternatively, one may instruct participants to alternate their classification between the two grammars on a trial-by-trial basis (Norman et al., 2011), referred to as a *mixed-block procedure*; this can be seen as a more demanding test of strategic control in that it requires a moment-by-moment monitoring of both grammars. An alternative procedure developed by Higham, Vokey, and Pritchard (2000) also involves exposure to two grammars. The test phase contains two types of instruction. *In-concert instructions* ask participants to identify strings that are consistent with either grammar as "grammatical", whereas *opposition instructions* ask them to identify only those strings that are consistent with one of the grammars. The assumption here is that opposition, but not in-concert instruction, requires strategic control. The in-concert condition is largely similar to Dienes et al.'s procedure. A final example is from the neighbouring area of statistical learning. Franco, Cleeremans, and Destrebecqz (2011) presented participants with two speech streams generated from two "artificial languages" (L1 and L2). In a discrimination task participants were presented with words from L1, L2, or neither. They either received *inclusion* instructions, which asked them to say "yes" if the word was from either language, or *exclusion instructions*, which asked them to say "yes" if it was from their target language (L1 or L2).

Strategic control and consciousness in implicit learning: theoretical positions

We will here address some theoretical positions on the relationship between strategic control and consciousness in implicit learning.

Strategic control indicates conscious knowledge. Some would regard strategic control as an indicator of conscious knowledge. In line with the theoretical frameworks of Jacoby (1991) and Baars (1988), measures of strategic control in implicit learning experiments have often been argued to show that control increases with consciousness – i.e. with the extent to which learning can be considered conscious rather than unconscious. For example, Destrebecqz and Cleeremans (2001) used strategic control during the SRT exclusion task to argue that knowledge was less conscious at lower response–stimulus intervals (RSIs), and more conscious when the interval was higher. Similarly, Wilkinson and Shanks (2004) argued that acquired knowledge was conscious when RSI was set to zero (i.e., RSI-0) on the basis of successful exclusion performance in this condition. Thus, they concluded that sequence learning was explicit rather than implicit. Higham et al. (2000) also included their measure of strategic control with the aim of separating between conscious/controlled influences, on the one hand, and unconscious/automatic influences on the other. However, there is also a handful of studies which specifically address whether knowledge that is not fully conscious can nevertheless be strategically controlled. These have been respectively inspired by the distinction between judgement/structural knowledge and the fringe consciousness framework.

Strategic control can occur with unconscious structural knowledge. A position which sees strategic control as compatible with unconscious knowledge, builds on a distinction between two types of knowledge hypothesised to result from implicit learning. These are *judgement knowledge* of whether nor not a certain stimulus complies with the acquired rules, and knowledge of the structure of these rules, i.e., *structural knowledge* (Dienes and Scott, 2005; Scott and Dienes, 2008, 2010). The assumption is that either of these varieties of knowledge could be conscious or unconscious. If structural knowledge is conscious, this will lead to conscious judgement knowledge. However, unconscious structural knowledge could be associated with either conscious or unconscious judgement knowledge. One example is the state of knowing *that* a sentence of one's native language is grammatical but without knowing *why* it is grammatical (Dienes and Scott, 2005). To assess the conscious status of each of the two types of knowledge in an implicit learning experiment, one may ask participants which decision strategy they used when making their classification response. Suggested response alternatives include "random choice", "intuition", "familiarity", "memory" or "rules" (Scott and Dienes, 2008). "Random choice", "intuition", and "familiarity" are assumed to reflect unconscious structural knowledge and therefore defined as "implicit" decision strategies, whereas "memory" and "rules" are commonly referred to as "explicit" decision strategies that are assumed to reflect conscious structural knowledge. The difference between "random choice", on the one hand, and "intuition" and "familiarity" on the other, is that the former is also associated with unconscious judgement knowledge whereas the latter two are associated with conscious judgement knowledge. In both AGL and SRT experiments, strategic control has been reported even when structural knowledge is unconscious. For example, Wan, Dienes, and Fu (2008) found that participants were able to strategically control the application of two grammars even when they reported using feelings, intuition, or random choice to arrive at their decision. Similarly, Fu, Dienes, and Fu (2010) found successful exclusion ability in an SRT task, even for trials attributed to intuition.

Along similar lines, Norman et al. (2007, 2011) have addressed whether implicit, unconscious knowledge may give rise to intuitive "fringe" feelings that may be strategically controlled. Using the terminology of Dienes and Scott (2005), this would refer to a situation of conscious judgement knowledge without conscious structural knowledge. These experiments have focused on whether participants who hold incorrect explicit beliefs about the nature/structure of acquired knowledge, thus indicating that structural knowledge is unconscious, can nevertheless strategically control the application of that knowledge. The implicit learning task must then be set up in a way that allows the participant to develop incorrect beliefs about the rules. This can be done by, e.g., introducing additional random variation in colour and shape of target stimuli, and target position indicators, in an SRT task (Norman et al., 2007). Similarly, random variation can be introduced into the colours and fonts of string elements in AGL (Norman et al., 2011). In an SRT task, Norman et al. (2007) found that even participants who misattributed the nature of the target sequence to irrelevant stimulus properties still showed strategic control over the application of sequence knowledge on a generation rotation task. Similarly, Norman et al. (2011) found that even participants who misattributed the nature of AGL letter regularities to irrelevant string elements showed strategic control over the application of two grammars in a mixed-block classification task.

Taken together, there is already evidence to show that the application of implicitly learned knowledge can be strategically controlled, even when it can be demonstrated that structural knowledge is unconscious.

Combining measurement procedures to study strategic control over the application of unconscious structural knowledge

Different studies have applied different measurement procedures for estimating whether structural knowledge of artificial grammars is conscious or unconscious. Some studies have measured strategic control among subsets of participants who claim unawareness of the learned rules (Norman et al., 2007, 2011), whereas others have measured it on subsets of trials on which participants report having used decision strategies involving unconscious structural knowledge (Dienes and Scott, 2008; Fu et al., 2010; Wan et al., 2008). Both forms of measurement assess participants' awareness of structural knowledge. However, whereas post-experimental questions about the nature of sequence or grammar rules ask about participants' representation of the *contents* of rule knowledge, decision strategy judgements can be seen to mainly reflect participants' understanding of the extent to which their response involved the *application* of conscious structural knowledge, without assessing the content itself (Norman, Scott, Price, and Dienes, 2016). Even though it is reasonable to assume that the two measures would most often converge, there might also be exceptions, e.g., when someone reported that they responded on the basis of a conscious rule related to irrelevant stimulus properties. Used in combination the two measures could be seen as a conservative measure of whether conscious structural knowledge is conscious.

One exception is a recent AGL experiment in which we asked participants, in a combined two-step judgement for each classification trial, to indicate (a) their decision strategy (random choice, feelings of intuition/familiarity, or explicit rules/memories) and (b) the relevant stimulus dimension (letter, colour, font) (Norman, et al., 2016). The rationale for combining the procedures was to provide a robust test of whether unconscious knowledge can be strategically controlled. If strategic control could be demonstrated in cases where participants both claimed that conscious structural knowledge was not involved, and also attributed their responses to irrelevant stimulus dimensions, this would go against the traditional view of strategic control being indicative of consciousness. However, we did not find strong evidence of strategic control on trials where feelings of intuition/familiarity were attributed to incorrect stimulus dimensions – the data were not sensitive enough to distinguish reliably between possible presence of strategic control and the null hypothesis of no control. Stronger evidence of strategic control was found on trials where the correct stimulus dimension was reported. We therefore speculated that strategic control may require at least global awareness of the nature of the rules, i.e., which stimulus dimension was relevant to the grammaticality judgement. However, a concern is that trial-by-trial ratings of stimulus dimension may increase participants' conscious hypothesis-testing and prompt their attention toward the correct nature of the rule. Moreover, this procedure may also not necessarily distinguish precisely between *attention to* certain stimulus properties and *awareness of* their importance to the rule. Therefore, these results need to be supplemented by a study in which decision strategies are measured on a trial-by-trial basis (cf. Dienes and Scott, 2005), but where rule awareness is assessed at the end of the experiment (cf. Norman et al., 2011). We now present an experiment that was designed for this purpose.

Method

Participants were 72 Norwegian students (36 females, 36 males) aged 18–33 (M = 21.7, SD = 3.2). All participants took part in two training phases, in each of which they were presented with letter strings from a different finite-state grammar (grammar A versus B, order counterbalanced across participants). Grammars and letter strings were taken from Dienes et al. (1995, see Figure 6.1). The AGL task was programmed in E-prime 2.0 (Schneider, Eschman, and Zuccolotto, 2002a, 2002b) and displayed by a 19" monitor. In each of the two training phases, each of 32 letter strings was presented three times, one at a time, in random order.

Strings consisted of 5–9 letters (X, V, M, R, T), with each letter written on one of five coloured backgrounds (red, purple, blue, green, or black) and in one of five different fonts (bold, italics, normal, outline, underline). The colour and font of each letter varied randomly between letter strings (see Figure 6.2). The instructions were to examine each string closely during its 7500 ms display period. To ensure participants attended all 3 stimulus dimensions, a post-trial cue was given on 24 randomly selected trials in each training phase, asking participants to report either the letter (8 trials), colour (8 trials), or font (8 trials) of a randomly chosen string element.

When both training phases were completed, participants were informed that letter strings had been governed by a different complex rule in each phase. They then proceeded to the test phase, which consisted of 60 classification trials. On each trial, three novel letter strings were presented simultaneously in a vertical column – one grammar A string, one grammar B string, and one ungrammatical string. Each string type occurred equally often in each screen position. Following the procedure by Norman et al. (2011), the classification rule, i.e., whether to select the grammar A or grammar B item, varied randomly between individual trials and was indicated by a written cue ("Rule 1?"/"Rule 2?") displayed above the letter strings (where "Rule 1" referred to the grammar (A or B) that had governed strings during the first training phase, and "Rule 2" to the second grammar (A or B)).

After each classification judgement participants rated their decision confidence on a three-point scale, but these data are not reported. Finally, using the mouse to select from an on-screen list, they indicated whether their response had been based on random choice, intuition, familiarity, rules, or memory (Scott and Dienes, 2008). The "implicit" decision strategies of "random choice, intuition, and familiarity" represent claims by participants that they were unaware of the structural aspects of the stimuli that motivated their decision (i.e., there was unconscious structural knowledge). Trials attributed to intuition or familiarity differ from those attributed to random choice because, in the former case, the participant claims to be aware of knowing whether they categorized correctly, even if they do not know why (i.e., judgement knowledge is conscious in the former but not latter case). The "explicit" decision strategies of "rules and memory" represent claims by participants that they were aware of relevant structural properties and indicate that both judgement knowledge and structural knowledge are conscious. (See Dienes, 2008, 2012 for further explication of structural and judgement knowledge.)

After the test phase, participants received a questionnaire where they allocated 12 points between the three stimulus dimensions (letter, colour, font) to reflect the extent to which they thought each dimension had contributed to the grammar rules. Conservatively, only participants who allocated 0 points to "letter" were classified as unaware of the nature of the rule, and all others were classified as potentially aware. The frequencies with which participants allocated the distribution of points are presented in Table 6.1.

Results

Each participant's degree of strategic control was expressed as a strategic score (Dienes et al., 1995), defined as the proportion of consistent strings chosen out of

TABLE 6.1 Frequencies with which participants allocated 0–12 points to the letter dimension.

Number of points allocated to "letters"													
	0	1	2	3	4	5	6	7	8	9	10	11	12
N	16	0	0	1	5	0	0	0	5	1	2	0	22

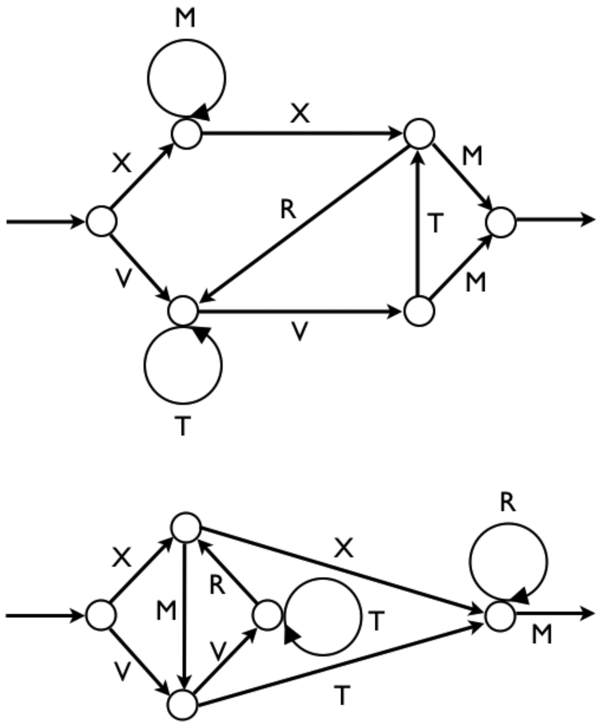

FIGURE 6.1 The two finite-state grammars used in the experiments, grammar A (top), and grammar B (bottom).

all consistent and inconsistent strings. A consistent string is one that follows the target grammar and an inconsistent string is one that follows the non-target grammar.

As analysis of strategic control over grammar knowledge is only meaningful if there is any learning at all, analyses of strategic control only included the 52/72 of participants who chose ungrammatical strings on less than a third of trials. (Note this filter is orthogonal to, and therefore does not artifactually bias, the comparison of the two grammars.) Of these participants, 36 were classified as aware and 16 as unaware of the nature of the rule.

The relationship between strategic control and awareness of the correct rule dimension was examined by comparing strategic scores to a chance level of .5. This was done separately for participants who expressed awareness of the relevance of the letter modality on the post-experimental questionnaire, versus for those who did not. It was also done separately for trials attributed to implicit versus explicit decision strategies. This yielded four conditions.We report effect sizes and Bayes factors in addition to NHST p-values, so that the reader can assess both the strength of evidence and conventional significance levels for any effects (Cumming, 2012; Dienes, 2014, 2015). $B_{H[0,.10]}$ refers to a Bayes factor used to test the hypothesis that strategic scores are above chance level of .5, represented as a half-normal with a SD of .10 above chance level, against the H_0, the hypothesis of chance performance.

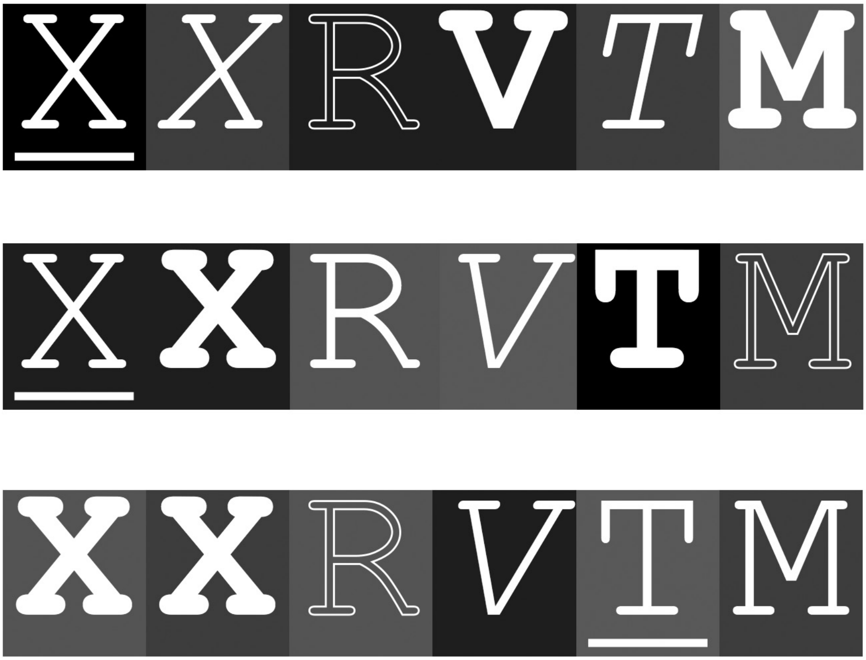

FIGURE 6.2 Three different examples of how the letter string «XXRVTM» (grammar A, Dienes et al., 1995) may appear with random variation in colour and font. The colours of the five elements of each letter string (from left to right) are: black, blue, red, red, blue, purple (top string); red, red, green, purple, black, blue (middle string); green, blue, green, red, purple, blue (bottom string).

The estimated SD of .10 was chosen based on data from a comparable previous study (Norman et al., 2011). A *B* of 3 or above indicates substantial evidence for the alternative above the null hypothesis, a *B* of 1/3 or below indicates substantial evidence for the null above the alternative hypothesis, and a *B* between 1/3 and 3 indicates data insensitivity for distinguishing between the alternative and null hypotheses (Dienes, 2014, 2015). Results are presented in Table 6.2.

Among participants who correctly attributed grammar rules to the letter dimension (i.e., "aware" participants), the Bayes factor was always above 3, and effect sizes were medium, regardless of whether decision strategy was explicit or implicit. This indicates substantial evidence for the alternative hypothesis above the null hypothesis (Dienes, 2014, 2015), in this case that strategic control was present. T-tests comparing performance to a chance level of .5 also showed that strategic scores were significantly above chance both for trials attributed to implicit and explicit decision strategies. Among participants who did not attribute grammar rules to the letter dimension (i.e., "unaware" participants), the Bayes factor was above 3 and the effect size was medium for implicit decision strategies, supporting the presence of strategic control despite a borderline conventional p-value

TABLE 6.2 Means and statistics for implicit versus explicit response strategies, reported separately for aware versus unaware participants.

	*N**	Strategic score *M(SD)* $B_{H[0,.10]}$	t-test compared to chance *(.5)* *t(df)* *p(2-tailed)*	Cohen's effect size *d*
Aware + implicit strategy	36	.59 (.16) 92.40	3.28 (35) .002	.56
Aware + explicit strategy	29	.61 (.28) 273.98	2.17 (28) .04	.39
Unaware + implicit strategy	16	.55 (.10) 3.10	2.08 (15) .055	.50
Unaware + explicit strategy	12	.51 (.16) .49	.14 (11) .89	.06

* *Note*: N differs between cells according to how many participants had responses within response category in question.

in a t-test. For explicit decision strategies, the Bayes factor was between 1/3 and 3 and the effect size was small, which indicates insensitivity for distinguishing between the alternative and the null hypothesis on these trials.[1]

Discussion

The key finding of the current experiment was that participants who did not express conscious structural knowledge of two learned grammars nevertheless showed some ability to strategically control the application of those grammars on a trial-by-trial basis. More specifically, strategic control was found on trials where participants claimed to respond on the basis of intuitive feelings, i.e., "implicit" decision strategies. This was the case even among participants who, after the experiment, expressed no awareness of the general nature of the grammars. Instead, they indicated that rules governing strings were related to irrelevant stimulus dimensions. The data therefore support the hypothesis that strategic control may be possible even when structural knowledge is not fully conscious. Our experiment applied two criteria for identifying cases of unconscious structural knowledge, i.e., that participants were not reporting the use of explicit decision strategies to arrive at their classification decisions, and that they expressed unawareness of the general nature of the acquired rules measured by a global rating after the experiment. Even under the combination of these two criteria, participants chose the target grammar more often than the non-target grammar.

Compared to the studies of Dienes et al., (1995) and Wan et al. (2008), our measure of strategic control was very stringent. Participants had to vary the classification rule between trials, which has been argued to require a higher degree of flexible control than if it is only varied between blocks of trials (Norman et al., 2011) because participants need to monitor both grammars on a moment-by-moment basis. Moreover, our criterion for including participants in the "unaware" subgroup was also very conservative, with only those participants who allocated zero points to the correct stimulus dimension being classified as unaware. Even though this implies that the "aware" subgroup may also contain participants who were less than fully aware that the rule was related to letters alone, it importantly reduces the possibility that the "unaware" subgroup contained participants who believed that the rules were related to letters. It is parsminonious to assume that participants would allocate at least some points to letters if they had even some slight conscious knowledge of the learned rules. Failure to report the stimulus dimension on which the rule was based can therefore be considered a strong indicator that conscious structural knowledge was present.

In sum, our data can be seen to question Jacoby's (1991) general view that strategic control over knowledge requires conscious knowledge, and to also address the more general long-standing debate over whether implicit learning is dependent on conscious awareness of rule fragments (Johnstone and Shanks, 2001; Perruchet and Pacteau, 1990; Redington and Chater, 1996). Even awareness of rule fragments would seem to necessitate awareness of which stimulus dimension mediates the rule. Given that we found grammar knowledge to be expressed without identifying the correct stimulus dimension, it seems implausible that conscious rule fragments can entirely account for AGL.

There is nevertheless a concern that participants who were classified as unaware on the post-experiment measure may have been guided by fleeting awareness of letter rules during some trials of the test phase. It could be argued that the reliability of our self-report measure of rule awareness would be improved if measured on a trial-by-trial basis, and that stronger evidence of strategic control over unconscious structural knowledge would be provided if strategic control were found on individual trials that were both rated as implicit and claimed to be specifically related to irrelevant stimulus properties. The only attempt to date at identifying strategic control over unconscious structural knowledge using such a procedure did not find robust evidence for strategic control on individual trials where participants denied the involvement of the relevant dimension. However, care is needed in comparing across studies. Differences in measured awareness across studies that apply different measurement procedures cannot straightforwardly be interpreted in terms of one measure being more sensitive to changes in conscious awareness than another. Certain measurement procedures could potentially also alter what participants are aware of. As pointed out above, measuring rule awareness on a trial-by-trial basis may for instance increase the likelihood that participants explicitly search for rules and become aware of the general nature of the rule. Future studies in this area will have to develop trial-by-trial measurement procedures that are less likely to interfere with people's hypothesis-testing, and that more adequately distinguish between *attention to* certain stimulus properties and *awareness of* their involvement in the rule, which is another limitation with this procedure.

Our current procedure did use a trial-by-trial measure, namely the structural knowledge attributions. Even with a trial-by-trial measure, noise will produce some misclassification. However, the percentage of trials classified as involving unconscious structural knowledge was 68.80 for participants aware of the relevant stimulus dimension, and 85.93 for participants unaware of the relevant stimulus dimension. It is unlikely that measurement noise, or biased responding by participants, could explain such a large proportion of responses, involving a similar level of strategic knowledge as for responses classified as involving conscious strategic knowledge (both .58).

Although the current data supported the hypothesis that strategic control does not require conscious structural knowledge, more studies are needed to specifically address whether strategic control may require conscious judgement knowledge, i.e., conscious knowledge of *whether or not* a certain letter string is grammatical (Dienes and Scott, 2005). This kind of knowledge in which people are aware that a stimulus belongs to a given category, without having conscious access to the antecedents of the knowledge, has also been referred to as intuitive cognitive feelings (Price and Norman, 2008, 2009) or fringe consciousness (Norman, Price, and Duff, 2006, 2010; Norman et al., 2007). Since, for statistical reasons, the three implicit response categories ("random choice", "intuition", and "familiarity") were combined in the current experiment, this question cannot be addressed from the current data.

As expected, participants who were aware that the rules were related to letters showed strategic control on trials where they claimed to apply "explicit" strategies. This is consistent with previous findings showing that knowledge which is consciously accessible and attributed to the correct source can be strategically controlled (Jacoby, 1991). Strategic control was also found when these participants classified their responses as related to the correct stimulus dimension but as nevertheless based on "implicit" decisions. Strategic control was not found when participants who were generally unaware of the correct stimulus dimension rated their classification decision as "explicit". This is expected if participants based their responses on incorrect, explicit hypotheses related to irrelevant stimulus properties, and therefore supports the validity of the self-report ratings of decision strategy.

Rünger and Frensch (2010) have argued that verbal reports are sensitive measures of consciousness in implicit learning, but only for measures reflecting the content of learning (e.g., our ratings of relevant stimulus dimension) and not metacognitive judgements (e.g., our ratings of explicit versus implicit decision strategy) which they argue are less sensitive and less informative. The current study shows the usefulness of both forms of verbal report measure and exemplifies how the two types of measurements may complement each other in AGL experiments.

Author note

This research was supported by a project grant from the Meltzer foundation. We thank Oskar Blakstad for his contributions in programming the experiment.

Note

1 An ANOVA comparing strategic scores for implicit vs. explicit decision strategies between aware vs. unaware participants showed no significant main effect of awareness [$F(1,39) = 1.91$, $p = .28$, $\eta_p^2 = .03$], no significant main effect of decision strategy [$F(1,39) = .11$, $p = .75$, $\eta_p^2 = .002$], and no significant interaction between decision strategy and awareness [$F(1,39) = 1.31$, $p = .26$, $\eta_p^2 = .03$].

References

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York, NY: Cambridge University Press.

Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.

Destrebecqz, A. and Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, *8*(2), 343–350. doi:10.3758/BF03196171.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00781.

Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In M. Overgaard (Ed.). *Behavioral Methods in Consciousness Research* (pp. 199–220). London, UK: Oxford University Press.

Dienes, Z., Altmann, G. T. M., Kwan, L., and Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1322–1338. doi:10.1037//0278-7393.21.5.1322.

Dienes, Z. and Perner, J. (2007). Executive control without conscious awareness: the cold control theory. In G. Jamieson (Ed.), *Hypnosis and Conscious States: the Cognitive Neuroscience Perspective* (pp 293–314). Oxford University Press.

Dienes, Z. and Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*(5–6), 338–351. doi:10.1007/s00426-004-0208-3.

Franco, A., Cleeremans, A., and Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: how conscious? *Frontiers in Psychology*, *2*, 229. doi: 10.3389/fpsyg.2011.00229.

Fu, Q., Dienes, Z., and Fu, X. (2010). Can unconscious knowledge allow control in sequence learning? *Consciousness and Cognition*, *19*(1), 462–474. doi:10.1016/j.concog. 2009.10.001.

Fu, Q., Fu, X., and Dienes, Z. (2008). Implicit sequence learning and conscious awareness. *Consciousness and Cognition*, *17*, 185–202. doi:10.1016/j.concog.2007.01.007.

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* (pp. 401–444). Thousand Oaks, CA: Sage Publications.

Hepler, J. and Albarracin, D. (2013). Complete unconscious control: using (in) action primes to demonstrate completely unconscious activation of inhibitory control mechanisms. *Cognition*, *128*(3), 271–279. doi: 10.1016/j.cognition.2013.04.012.

Higham, P. A., Vokey, J. R., and Pritchard, J. (2000). Beyond dissociation logic: evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General*, *129*(4), 457–470. doi:10.1037//0096–3445.129.4.457.

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. doi:10.1016/0749-596X(91)90025-F.

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

Lau, H. C. and Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *The Journal of Neuroscience*, *27*(21), 5805–5811. doi 10.1523/JNEUROSCI.4335-06.2007.

Lau, H. and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373.

Mong, H. M., McCabe, D. P., and Clegg, B. A. (2012). Evidence of automatic processing in sequence learning using process-dissociation. *Advances in Cognitive Psychology*, *8*(2), 98–108. doi: 10.5709/acp-0107-z.

Nissen, M. J. and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, *19*, 1–32. doi:10.1016/0010-0285(87)90002-8.

Norman, E. (2015). Measuring strategic control in implicit learning: how and why? *Frontiers in Psychology: Consciousness Research*, *6*:1455. doi: 10.3389/fpsyg.2015.01455.

Norman, E., Price, M. C., and Duff, S. C. (2006). Fringe consciousness in sequence learning: the influence of individual differences. *Consciousness and Cognition*, *15*(4), 723– 760.

Norman, E., Price, M. C., and Duff, S. C. (2010). Fringe consciousness: a useful framework for clarifying the nature of experience-based feelings. In A. Efklides and P. Misailidi (Eds.). *Trends and Prospects in Metacognition Research*, pp. 63–80. New York, NY: Springer.

Norman, E., Price, M. C., Duff, S. C., and Mentzoni, R. A. (2007). Gradations of awareness in a modified sequence learning task. *Consciousness and Cognition*, *16*, 809–837. doi:10.1016/j.concog.2007.02.004.

Norman, E., Price, M. C., and Jones, E. (2011). Measuring strategic control in artificial grammar learning. *Consciousness and Cognition*, *20*, 1920–1929. doi:10.1016/j.concog. 2011.07.008.

Norman, E., Scott, R. B., Price, M. C., and Dienes, Z. (2016). The relationship between strategic control and conscious structural knowledge in artificial grammar learning. *Consciousness and Cognition*, *42*, 229–236. doi: 10.1016/j.concog.2016.03.014.

Perruchet, P. and Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*(3), 264–275.

Price, M. C. and Norman, E. (2008). Intuitive decisions on the fringes of consciousness: are they conscious and does it matter? *Judgment and Decision Making*, *3*, 28–41.

Price, M. C. and Norman, E. (2009). Cognitive feelings. In T. Bayne, A. Cleeremans, and P. Wilken, (Eds.), *Oxford Companion to Consciousness*, pp.141–144. Oxford: Oxford University Press.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *77*, 317–327. doi:10.1016/S0022-5371(67)80149-X.

Redington, M. and Chater, N. (1996). Transfer in artificial grammar learning: a reevaluation. *Journal of Experimental Psychology: General*, *125*(2), 123–138.

Rosenthal, D. M. (2005). *Consciousness and Mind*. Oxford, UK: Clarendon Press.

Rünger, D., and Frensch, P. A. (2010). Defining consciousness in the context of incidental sequence learning: theoretical considerations and empirical implications. *Psychological Research*, *74*, 121–137.

Schmidt, J. R., Crump, M. J. C., Cheesman, J., and Besner, D. (2007). Contingency learning without awareness: evidence for implicit control. *Consciousness and Cognition*, *16*(2), 421–435. doi: 10.1016/j.concog.2006.06.010.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002a). *E-Prime Reference Guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002b). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Scott, R. B. and Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1264–1288. doi: 10.1037/a0012943.

Scott, R. B. and Dienes, Z. (2010). The metacognitive role of familiarity in artificial grammar learning: transitions from unconscious to conscious knowledge. In A. Efklides and P. Misailidi (Eds.). *Trends and Prospects in Metacognition Research* (pp. 37–61). New York, NY: Springer.

Van Gaal, S., Ridderinkhof, K. R., Scholte, H. S., and Lamme, V. A. F. (2010). Unconscious activation of the prefrontal no-go network. *The Journal of Neuroscience*, *30*(11), 4143–4150. doi:10.1523/JNEUROSCI.2992-09.2010.

Wan, L., Dienes, Z., and Fu, X. (2008). Intentional control based on familiarity in artificial grammar learning. *Consciousness and Cognition*, *17*(4), 1209–1218. doi:10.1016/j.concog.2008.*06*.007.

Wilkinson, L., and Shanks, D. R. (2004). Intentional control and implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 354– 369. doi:10.1037/0278-7393.30.2.354.

7

ABSTRACTION IN SEQUENCE LEARNING

Ferenc Kemény and Ágnes Lukács

Introduction

"Not everything is as seems" says Mr. Miyagi to Daniel in the film *The Karate Kid* when Daniel articulates his complaints on Mr. Miyagi treating him as a slave, making him wash and wax his cars, paint his house, etc. Soon it turns out that the clockwise movements required for applying the wax on the car and the counter-clockwise movements for removing it can also be used for blocking others' offensive actions. It is also revealed that the core point was not only to make Daniel wash the cars and paint the house, but to make him practice movement sequences that can be utilized during martial arts. Despite the great success depicted in the film, it is not obvious whether human skill learning behaves in such a way, that is, whether skills are really so independent from the source stimuli. The current chapter reviews the role of abstraction in an important aspect of skill learning: sequence learning.

Sequence learning is a fundamental ability of individuals to adapt to the environment by identifying repeating patterns and using the acquired representations to predict and prepare for future events. The environmental requirements in this case can involve a specific required movement by a sportsperson or a yet unknown combination of notes a cellist has to play. It has been subject to debate whether sequence learning is in fact based on surface cues or abstract information. That is, whether participants learn changes of light intensity reflecting from the paper (surface physical features of letter strings), or a sequence of superordinate categories from the presentation of individual items (abstract features, see discussion by Redington and Chater, 1996). The current review focuses on four aspects of the distinction between abstract versus feature-specific sequence learning: 1) transfer, 2) abstract sequence learning in the absence of transfer, 3) the effect of task complexity in sequence learning and 4) modality-dependence of sequence learning.

Intramodal and crossmodal transfer in sequence learning

When transfer is observed, it is taken as evidence of abstraction. That is, if learning with one set of stimuli affects learning with another set, it indicates learning beyond the level of specific items. In sequence learning, the order of the stimuli, as well as the stimuli themselves, provide information. If the identity of the stimuli is altered, but the underlying transitions are preserved, we expect a positive transfer. Transfer effects of this kind suggest that the underlying structure is learned independent of the specific stimuli. First, transfer results from Artificial Grammar Learning experiments are introduced, then we focus on studies of the Serial Reaction Time Task.

The longstanding debate on the abstractness of sequence learning dates back to the first publication on implicit sequence learning. Reber (1967) used the prototype of Artificial Grammar Learning (AGL) tasks, in which participants viewed a set of letter strings, and were asked to copy them. Unknown to the participants, the strings were generated by a finite state grammar: a mini grammar defining the possible order of the elements. Participants only learned about the presence of an underlying grammar after the "copy-phase", which served as a learning period. In the second phase, participants were exposed to previously unseen sequences that either followed the rules of the same grammar, or violated them. Reber also tested whether the learned grammar representation is truly independent of the source stimuli, and used an entirely different set of letters in the test phase. Results showed that participants could discriminate grammatical and agrammatical sequences even for strings with the new set of letters. Thus, he concluded that solving this task requires and relies on the abstraction of the underlying grammar.

The same assumption inspired a further study that tested not only intradimensional (from one set of letters to another), but also interdimensional and intermodal transfer. Altmann, Dienes and Goode (1995) tested crossmodal transfer in AGL. Four experiments were conducted, in which training and test stimuli differed. The following stimulus sets were used throughout the four experiments: auditory verbal, auditory non-verbal and visual nonsense stimuli. Results showed that performance was best if stimulus sets were the same in the training and test phases. Test performance was, however, still above chance after training in another modality or dimension. That is, in spite of differences in the efficiency of transfer across conditions, results showed a significant positive transfer between different sets of stimuli using the same grammar.

While the above data might appear convincing, Tunney and Altmann (1999) argued that these results are not due to transfer of structural knowledge, but to the recognition of a specific cue. This specific cue is the illegal starting element of ungrammatical sequences. Tunney and Altmann replicated the experimental design of Altmann, Dienes and Goode (1995) with identical results. In a second experiment illegal (low frequency) starters were excluded from the ungrammatical sequences, so that the detection of ungrammaticality could only be based on recognizing illicit sequences within letter strings, i.e. on implicit knowledge of the grammar. This manipulation resulted in the lack of positive transfer in Experiment 2.

The result argued against abstract grammatical information as a basis of transfer from one set of stimuli to the other in Altmann, Dienes and Goode's experiment. Another study by Gómez (1997) also highlighted the importance of sequence-initial clusters in transfer.

The possibility of reliance on sequence-initial clusters is not the only methodological shortcoming of AGL transfer studies. Most transfer experiments also neglect the important issue of how participants can link pre- and post-transfer stimuli. Considering the previously discussed positive transfer in Altmann, Dienes and Goode (1995), it is not evident why transfer is expected to take place at all. In their Experiment 1, Altmann, Dienes and Goode (1995) trained participants with sequences composed of 5 letters, and tested them on sequences of 5 tones. The underlying grammar was identical. It was up to the participants how they mapped the pre- and post-transfer stimuli on each other. A set of 5 letters can be mapped onto a set of 5 tones in 5! = 120 ways; however, only one of these combinations is relevant from the perspective of the task. It is very unlikely, though, that participants can find the correct mapping in a test session of 50 items, especially if there are sequences with no repetitions (e.g. "hes sog pel jix" is grammatical, while the "hes kav jix pes" sequence is agrammatical in Experiment 3 of Altmann, Dienes and Goode, 1995, p. 912). Overall, it is not clear whether transfer takes place at all in the AGL task. The lack of transfer in such studies can be caused by the failure to develop the abstract regularities of the artificial grammar, or by the methodological problems with the applied techniques for testing transfer. We argue for the latter, and provide possible solutions below.

There are at least two possible solutions to link pre- and post-transfer stimuli. One is to use meaningful stimuli with the same semantic content: e.g. words and pictures of the same item (a sequence of the words "*cat dog goat duck*" can be linked to the similar sequence of images of the same animals through meaning, i.e. via an abstract conceptual level). The other option is to explicitly create associations between meaningless pre- and post-transfer stimuli. A previous study (Kemény and Lukács, 2011) used the former method. Participants were trained with an AGL task. In the training phase, sequences of auditory verbal category names (e.g. "*furniture, fruit, mammal, tool*") were presented; the test phase, however introduced category *members* of the same categories (e.g. "table, apple, giraffe, hammer"). These sequences of category members or tokens were either auditorily presented words, or visually presented colour pictures. That is, all participants were expected to transfer structural knowledge (serial order of categories) from a set of categorical stimuli to a set of tokens from the same categories (serial order of members of the same category). Half of the participants had to transfer their knowledge within modality, and the other half between modalities. Participants showed evidence of both intramodal and crossmodal transfer: grammaticality judgement accuracy of participants with preceding auditory training was significantly higher both for the visual and the auditory tests than the performance of participants who only faced the test phase without training. The experiment yielded further interesting results: although performance in both transfer conditions was significantly above-chance level, the crossmodal transfer condition showed significantly higher performance than the intramodal transfer condition. A possible explanation for lower performance in the intramodal condition is within-modality interference. Superordinate categories, like "fruit" or "furniture" have verbal labels, but do not have visual representation. There is no such thing as the image of a general fruit, we only have images of an orange or an apple. Hence category-level training might interfere with token-level testing if both are presented in the auditory modality. Most importantly, however, results show that participants were able to learn sequences of auditorily presented categories, and transferred their structural knowledge onto visual sequences of tokens. While this study was not designed to differentiate between structural and non-structural cues, it provided evidence in favour of abstract learning.

As described above, the other possibility to overcome the mapping problem is to couple pre- and post-transfer stimuli using a pairwise association method. An experiment by Lukics and Kemény (2016) exposed participants to pairwise associations of visual and auditory stimuli. In a categorization task, participants were trained to associate meaningless visual stimuli to meaningless auditory stimuli. This way the mapping between the stimuli of the later training and test phases were already set. Then participants were asked to repeat visual sequences by pressing corresponding response keys in a limited amount of time (a method based on Conway, Bauernschmidt, Huang and Pisoni, 2010). Unknown to the participants, the visual sequences were structured. In the test phase, participants faced structured and random auditory sequences. Despite previous mapping of visual and auditory elements, no difference was found between response rates for structured and random sequences. This pattern of results might be due to the fact that the expression of procedural learning is deficient if participants have to transcode a sequence of auditory elements into visual items. However, no evidence was obtained that elements were abstracted.

The method of this latter study resembles another well-known sequence learning paradigm, the Serial Reaction Time Task (SRTT, Nissen and Bullemer, 1987). In this task, a target stimulus appears at one of four possible locations. The aim of the participants is to press the response key that corresponds to the stimulus location. Participants are asked to be as quick and as accurate with their responses as possible. Upon pressing the correct response key, the location of the target stimulus changes. Unknown to the participants, the location follows a predetermined sequence. Reaction times (RTs) decrease with practice. After some time, the repeating sequence is replaced by a random pattern, and RTs immediately increase (Destrebecqz and Cleeremans, 2003).

The central aim of testing transfer in the SRTT is different from AGL studies, as these studies focus more on the basic constituents of sequence learning. There is a widespread debate on the basis of sequence learning. The perceptual learning hypothesis suggests that learning is based on the prediction of the next stimulus (Remillard, 2003), hence the learned sequence is composed of associations between *stimuli*. Response theory suggests that participants learn to predict the next *response* location (Willingham, Wells, Farrell, and Stemwedel, 2000). In this case, learning is to some extent goal oriented, as a special representation of response goals is required (for related discussion of the AGL task, see Chapter 9 by Popławska-Boruc, Sterczyński and Roczniewska, current volume). The last possibility is pure effector-based motor learning (Deroost, Zeeuws and Soetens, 2006). This suggests that specific "organs", like fingers or muscles, learn the given motor sequence.

Previous transfer studies of the SRTT were interested in identifying what the crucial component of sequence learning is. If the central aspect of sequence learning is the acquisition of the response pattern, then performance is expected to be higher if the response pattern is maintained and all other aspects of the task are modified. Similarly, the stimulus (Remillard, 2003) or effector-based sequences (Deroost, Zeeuws and Soetens, 2006) can be selectively maintained from pre- to post-transfer phases. These studies showed a positive transfer from one hand to the other (Japikse, Negash, Howard and Howard, 2003), as well as from finger responses to arm responses, or between manual and verbal responses (Keele, Jennings, Jones, Caulton and Cohen, 1995).

As in the AGL task, transfer studies of the SRTT also focus on an overlap between pre- and post-transfer phases. While stimuli and structure are easily dissociated in the AGL task, SRTT has several different layers, like response-based, stimulus-based or effector-based information. In contrast to the AGL task, if only the stimuli are altered, the response and effector-based information is still maintained. The maintained information (e.g. the response sequence) helps mapping the pre-transfer stimuli and the post-transfer stimuli. The only possible mapping is clear from any single post-transfer sequence chunk. This is not the case with the AGL task. Mapping between pre- and post-transfer stimuli is only supported by the abstract structure. This facilitation is generally difficult to realize based on a single sequence, and results are not conclusive as to whether a test session of approximately 50 randomly presented grammatical and agrammatical sequences is enough to obtain such a mapping.

In sum, transfer studies of AGL and SRTT traditionally focus on different phenomena. Studies of AGL are more concerned with the transfer of abstract, stimulus-independent grammar between two sets of stimuli, and AGL studies so far provided mixed evidence regarding the existence of transfer at that level. The lack of transfer can reflect the lack of such an abstract grammar representation, as well as a methodological problem, as pointed out above. SRTT experiments on the other hand focus on how different stimulus-, response- or effector-based types of information contribute to learning performance. Thus, some item-based information is always maintained between pre- and post-transfer stimuli. As a result, SRTT transfer studies do not require real abstraction or stimulus-independent transfer. These methodological differences and shortcomings need to be addressed in order to obtain conclusive results on transfer and abstraction in sequence learning.

Abstract sequence learning without transfer

The previous section focused on abstract sequence learning as evidenced by transfer effects. Several studies, however, tested the learning of abstract sequences without transfer. These studies were either designed to prevent the possibility of simply focusing on surface, stimulus-based features, or required the interpretation of the stimuli to solve the task. In a novel design, Goschke and Bolte (2007) used a special task in which participants saw line drawings of objects, and had to name the objects. The pictures were from four different categories: body parts, animals, clothing and furniture. Participants were instructed to pay attention to the categories themselves. Unknown to the participants, the order of categories varied according to a six-element sequence. Like in previous SRT studies, naming latencies increased when the sequential organization was replaced by random presentation of items. In a second experiment, (Experiment 2, Goschke and Bolte, 2007) participants were not instructed to pay attention to the categories, but the first element was identical in all sequence presentations making the beginning of the abstract sequence more salient. This design also resulted in sequence learning. When this salient cue was removed in Experiment 3, sequence learning was still observed even in the absence of instructions to pay attention to categories. These results argue for abstract learning of sequential regularities, even in the absence of response-based information: participants showed a sequence learning effect even without a sequential pattern in the stimuli or responses at the item level.

Using a neural network model based on the primate fronto-striatal system, Dominey and colleagues (Dominey, Lelekov, Ventre-Dominey and Jeannerod, 1998) modelled learning on the SRTT. The central question was whether similar learning takes place in the presence and absence of correlated predictive surface information. The abstract structure was the fixed repetition of given elements, that is, in the case of a 6–1–3–1–6–3 sequence, Stimulus 1 is repeated two items later, Stimulus 6 is repeated 4 items after its first appearance, while every third item is Stimulus 3. Note that in this case, the first three items are unpredictable. In the case of correlated surface features, the items are identical (i.e. the same sequence of the same 3 numbers are repeated over and over again), while in the case of abstract structure learning, each block used a different, but structurally identical sequence (with the first stimulus being repeated 4 items later, the second 2 items later and the third 3 items later). Results showed that the models could learn the abstract sequence as well as the specific sequence. Dominey and colleagues also report two human experiments with the same learning design, with both implicit and explicit conditions (in the explicit condition, participants were visually shown the abstract sequential structure of the stimuli to be presented, and were told that finding this structure would help them solve the task). The correlated surface and abstract sequence could be learned by both implicit and explicit learners, while only explicit learners learned the abstract sequence alone. These results show that abstract sequential learning can take place, but it requires awareness over the structure to be learned.

As we previously described, the basis of learning in the SRTT is debated. It is not obvious which type of information contributes sequence learning: stimulus-, response- or effector-based information, and if more than one type is in action, what the relative contributions of each are. Unfortunately, the structure of the

SRTT makes it difficult to selectively manipulate the different types of sequences. Selective manipulation of the different streams of information often results in a serious loss of statistical power, or in the lack of sequence learning. Based on such results, the Correlated Sequences Approach suggests that sequence learning only appears in the case of at least two correlated streams of information: single response, stimulus or effector sequences are not learned (Meier and Cock, 2010; Weiermann, Cock, and Meier, 2010). Testing this hypothesis required a novel design, the Task Sequence Learning paradigm (Heuer, Schmidtke, and Kleinsorge, 2001; Koch, 2001). In this complex paradigm, participants are informed that they will face different dichotic decision tasks. In the Animals Task, participants have to decide whether the target stimulus is a mammal or a bird; in the Implements Task, they have to decide between kitchen utensils and musical instruments; while in the Plants Task they have to differentiate between trees and flowers (cited from Weiermann and Meier, 2012). All tasks use the same response keys; hence the order of the tasks as well as the order of the responses can be selectively manipulated. Several previous studies have shown that single response sequences or task sequences are not acquired, but correlated sequences lead to sequence learning (Cock and Meier, 2013; Kemény and Meier, 2016; Meier and Cock, 2010; Meier, Weiermann, and Cock, 2012). It is important to note that the task sequence (e.g. Animals–Instruments–Plants–Animals–Plants–Instruments) is at the level of abstract categories, as it is the result of inference: participants see instances of categories, and make a categorical decision based on the task. As a result, learning correlated task and response sequences means learning something not stimulus-bound. It would be important, however, to test whether sequence learning emerges from instruction or is based on the fact that some categories, like trees and flowers, are hierarchically organized into higher-level categories, hence the concepts of the higher-level categories emerge from the activation of the member categories. That is, would sequence learning also appear in the absence of task-based instructions? Or would it be possible to learn a sequence of tasks when the decisions are related to ad hoc categories, i.e. instead of two different categories of animals, participants have to decide between two unrelated categories? These questions of abstract task sequence learning await future investigation (for further related discussion see Meier and Cock, 2012).

The above section described the learning of abstract sequences. All the above-mentioned tasks require the activation of information that is not directly observable from the physical features of stimuli, but is available through the interpretation of the stimuli. These results argue that abstraction might be a consequence of task requirements: if solving the specific experimental task requires the interpretation of the stimuli, abstract learning can take place (see Kemény and Lukács, 2016 for further discussion).

The differentiation between interpreted and non-interpreted stimuli in sequence learning is not new and their processing might rely on different neural pathways. Keele, Ivry, Mayr, Hazeltine and Heuer (2003) argue that there are two separate neural architectures underlying sequence learning: a dorsal and a ventral pathway.

The dorsal stream works with uninterpreted stimuli. As learning is bound to the input modality, stimuli cannot be linked to each other across dimensions or modalities. That is, the dorsal stream only processes sequences with elements from the same dimension. Since the input signal is uninterpreted, it is difficult to verbalize the sequence, which makes it impossible to directly recall it (Keele et al., 2003). Hence, sequence representations of the dorsal stream are implicit. The ventral stream, in contrast, processes categorized input. The categorized nature of input makes it possible to integrate interdimensional stimuli into a single sequence, and to verbalize and/or recall that sequence. Conscious access, however, is not necessary: ventral representations may be either implicit or explicit. In sum, according to Keele and colleagues (2003) the interpreted versus uninterpreted nature of the stimuli is a central aspect in the underlying neural pathways behind learning.

Task complexity

Abstractness can also be a matter of degree associated with task complexity. This is perhaps best exemplified by the wide variety of tasks used in the area of statistical learning and the resulting differences in the abstractness of the acquired representations. The following section will briefly introduce how the original AGL task was modified to test hypotheses related to language acquisition. The different tasks show a wide range of variability in task complexity: the second part of the section focuses on this issue.

The current literature on learning sets of artificial sequences uses a wide array of tasks in the domain of statistical learning. A seminal paper by Saffran and colleagues (1996) provided an empirical demonstration of 8-month-old infants' sensitivity to differences in transitional probabilities (TPs) in sequences of nonsense syllables. In the experiment, infants were exposed to a 2-minute-long monotonous stream of CV syllables. CV syllables formed trisyllabic pseudowords. Syllables within words always followed each other (TPs are 1), while syllables on the pseudoword boundaries could be followed by three possible syllables (TPs are 0.333). In the test phase, participants were exposed to trisyllabic pseudowords (where both TPs are 1) and word boundary triplets (one TP is 1, the other is 0.333). Infants showed different looking times to word boundary triplets and pseudoword triplets, suggesting a sensitivity to statistical information. This paper by Saffran and colleagues introduced the notion of statistical learning, which has later been demonstrated to be an effective learning mechanism outside the verbal domain as well (auditory non-verbal information: Saffran, Johnson, Aslin and Newport, 1999; visual information: Fiser and Aslin, 2002; Kirkham, Slemmer and Johnson, 2002; visuomotor movements: Hunt and Aslin, 2001). Since then, statistical learning has also been extended from TPs over adjacent specific items to learning of TPs over non-adjacent items (e.g. Gómez and Gerken, 1999) and TPs over categories (e.g. Saffran, 2002; Lany and Gómez, 2008). In these studies, a simple artificial grammar (a finite state grammar) is used to generate sentences from a vocabulary of nonsense syllables (or other items). Infants are trained on these grammatical sentences for a short period, and in the test phase they face novel grammatical and ungrammatical sequences. Experiments have demonstrated that infants can learn regularities even when TPs are defined over categories instead of specific items (Lany and Gómez, 2008; Saffran, 2002), and even when a new vocabulary is introduced in the test phase (Gómez and Gerken, 1999; Marcus, Vijayan, Rao and Vishton, 1999; Marcus, Fernandes and Johnson, 2007; Peña, Bonatti, Nespor and Mehler, 2002; Saffran, Pollak, Seibel and Shkolnik, 2007). This is taken as evidence that infants in these cases extracted some abstract structure from the original sentences. It is an open question though whether only categories are abstract (and the sequences are still defined by first-order statistical dependencies between them, and it is the hierarchical build-up of such first-order dependencies that yields a seemingly complex sentence structure) or the acquired structure is itself the result of abstraction.

Shortly after the Saffran et al. (1996) study, the problem of abstraction also generated heated debates in infant learning literature. Marcus and colleagues (1999) argued that statistical learning is not sufficient to account for all types of sequence learning, and a different mechanism, which they call algebraic rule learning, is also necessary. These algebraic rules are defined as "open-ended abstract relationships for which we can substitute arbitrary items" (Marcus et al., 1999, p. 77). The rules in their experiments were very simple, and instead of defining TPs over syllables, they defined abstract relations across categories of syllables: generating ABA or ABB sequences of CV syllables presented in the training phase. In the test phase, infants were more surprised to hear sequences that did not follow the rules of their training grammar than to hear sequences in line with it (both grammatical and ungrammatical sequences were composed of new syllables in the test phase). Marcus and colleagues argue that this is evidence of infants' ability to extract abstract algebraic rules like "the first item is the same as the third item". Marcus and colleagues proposed that there are two different learning mechanisms active and available to infants (at least in language acquisition): one is sensitive to statistical distributions and TPs while the other operates on abstract variables. The idea of two mechanisms was also taken up by Pena et al., (2002), with different arguments.

Abstractness was at the centre of the debate generated by these studies (Altmann and Dienes, 1999; Christiansen and Curtin, 1999; Eimas, 1999; Perruchet, Tyler, Galland and Peereman, 2004; Seidenberg and Elman, 1999; Seidenberg, MacDonald and Saffran, 2002). There is agreement that they at the very least show that infants can form abstract categories, and they might even demonstrate this knowledge, but they do not necessarily rely on algebraic rules. Although TPs are primarily used to describe sequences of specific items (e.g. words), the only difference in the experiments with algebraic rules is that rules are not probabilistic (TPs are always 1), and are defined over abstract categories (A and B). In fact, Pena et al. (2002) argue that in principle, the AXB structures (that is, Stimulus A predicting Stimulus B, with an X intervening random stimulus) in their experiments can be defined by TPs over non-adjacent categories. In Marcus et al.'s experiment, developing abstract representations was based on establishing the perceptual categories of "same" and "different" which might be remarkable, but is not the same as the ability to learn algebraic rules. It is difficult to draw a clear distinction between an algebraic rule and a hierarchical structural description built on TPs.

Gómez and Gerken (2000) give a classification of studies on statistical learning putting abstraction in the centre. All of these studies train participants on a small artificial language, but the presumed level of abstraction differs. Along with the original AGL task (Reber, 1967), studies on segmentation of word-like elements (see above, e.g. Saffran et al., 1996; or Aslin, Saffran, and Newport, 1999; Saffran et al., 1996) are on the first level requiring the least amount of abstraction. In these tasks, participants need to show sensitivity to differences in TPs between specific items. The second level in abstraction considers AXB-type non-adjacent dependencies (first element predicting the third with an intervening random element, Gómez, 2002), long distance dependencies (Friederici, Bahlmann, Heim, Schubotz, and Anwander, 2006) and context-free grammars (de Vries, Monaghan, Knecht and Zwitserlood, 2008). To solve these tasks, participants first need to identify the structured and unstructured elements, and only afterwards can they learn the structure. The third level of abstraction involves tasks that require participants to extract rules and generalize those to new stimuli. Learning ABA-like rules (see above, Marcus et al., 1999) as well as AGL transfer studies (e.g. Altmann, Dienes and Goode, 1995, discussed above) are examples of this level.

The fourth level of abstraction taps into the use of abstract syntactic categories (Gómez and Gerken, 2000). Studies using this approach usually employ a more complex grammar, and a complex vocabulary in which several tokens build up the grammatical categories. Friederici, Steinhauer and Pfeifer (2002) trained participants on a board game. While playing, participants were required to verbalize their moves using an artificial language developed specifically for the task. This artificial language was composed of syntactic rules, as well as the semantic coding of objects. Results showed that after up to six hours of training, syntactic anomalies elicited similar event-related potentials for the newly learned artificial language as in their mother tongue. Another study (Saffran, 2002) used four simple rules and five categories of word-like stimuli employing the AGL procedure, and found above-chance performance even if tokens of the categories were interchangeable throughout both the training and testing phases.

While it is not surprising that task complexity affects learning performance, studies directly addressing task complexity effects are rare. A recent study systematically compared the effect of grammar complexity on learning, together with the effect of semantic anchoring (Van den Bos and Poletiek, 2015). Results showed a quantitative relationship between complexity and learning: participants learned more on simple than on complex grammars, and semantic anchoring was beneficial only in the case of simple grammars.

Constraints on abstract learning

While most of the studies discussed above focused on the abstractness of the acquired information and assumed that the learning mechanism behind the acquisition is abstract and modality- and domain-independent, there are also studies presuming stimulus-dependent, or at least modality-constrained learning mechanisms. Conway and Christiansen (2005) used the same AGL task design with visual, auditory and haptic stimuli looking for differences in learning in different modalities. Results showed both qualitative and quantitative differences: along with a quantitative advantage of the auditory modality, participants of the auditory conditions tended to focus more on sequence-final chunks, while no such preference was observed in the other two modalities.

In a later study, Conway and Christiansen (2006) asked whether participants are able to learn two simultaneously trained grammars. The relationship between the two stimulus sets was manipulated: they were from different modalities, or from the same modality, but different dimensions (e.g. colours versus shapes), or from the same dimensions (different sets of shapes). Results showed that participants were only able to show above-chance performance if the two stimulus sets differed at least in perceptual dimension. Conway and Christiansen (2006) argue that participants could not learn two parallel grammars from the same dimension, since the grammars got mixed up due to dimension overlap. This did not happen with stimuli from different dimensions or from different modalities. Conway and Christiansen conclude that a single, domain- and modality-general learning mechanism cannot explain why only intradimensional stimuli get mixed up, and argue for several parallel learning mechanisms (an argument further developed in Frost, Armstrong, Siegelman and Christiansen, 2015).

Conway and Christiansen (2006), however, do not discuss the possibility of transfer. In their design, participants were trained with two grammars, grammar A and B, with according stimulus sets, stimuli A and B. In the test phase, participants were exposed to grammar A sequences with stimuli A and grammar B sequences also with stimuli A. Only grammar A sequences with stimuli A were considered grammatical, while transfer sequences were considered incorrect (grammar B, stimuli A). Results showed that participants were able to differentiate between grammar A–stimuli A and grammar B–stimuli A sequences only if stimuli A and stimuli B differed in modality or dimension. In this design, however, chance-level performance cannot distinguish between the lack of learning in this dual task and a possible positive transfer. That is, if participants accept grammar A sentences with stimuli A as well as grammar B sequences with stimuli A (transfer), their classification performance would also be at chance level.

A later study aimed at testing stimulus-dependence in the SRTT (Kemény and Németh, 2017). In this study, participants were trained on a visual SRTT without any correlated responses. Participants were instructed that they would see red and grey dots appearing at four different locations, and they were asked to count the grey dots (15% of all trials). Unknown to the participants, a 12-element sequence determined the appearance of the dots. After some training, an auditory task was introduced. Participants heard one of four possible CV syllables, and were asked to press the response key corresponding to the target stimulus. Unknown to the participants, the spatial organization of the response sequence was either the same as or the reverse of the spatial organization of the visual stimuli in the training phase. Two further control conditions were used: one had no structure in the dot-counting task, while the other had no dot-counting task at all. Results showed that the two experimental groups performed significantly worse on learning the auditory–motor sequence of the test phase than the two control groups. That is, not only was no positive transfer observed between the training and test phases in the case of identical sequences, but a previously learned sequence caused a deficit in later sequence learning. According to the authors the results argue for stimulus-dependence in sequence learning: that is, two sequences are only identical if both the structure and the stimuli are identical. This stimulus-dependence is responsible for the lack of positive transfer. The negative transfer, on the other hand, is explained by suggesting a *general sequence processor* mechanism, which is responsible for processing different sequences, but has a limited capacity. Thus, learning one sequence reduces later sequence learning. Importantly, these results argue for both stimulus-independence and stimulus dependence on different levels: stimulus-dependence in the sequential representations, but stimulus-, domain- and modality-independence on the level of the learning mechanism.

Conclusion

The current review focused on implicit sequence learning and revolved around topics related to abstractness of the acquired representations. We have discussed studies that focused on transfer, abstract sequence learning without transfer, task complexity, and modality constraints on sequence learning. Several inconsistencies were identified in connection with abstract sequence learning: the difficulty of mapping in transfer studies, instruction- vs representation-based abstraction, or the requirement of stimulus interpretation to solve the task. We propose that some of these inconsistencies may be resolved by a theoretical distinction on whether a specific task requires abstraction or not. Abstraction in this sense would distinguish between tasks in which the input stimuli are to be interpreted versus tasks where focus on surface features is enough for learning. The classic SRTT (Meulemans, Van der Linden and Perruchet, 1998; Nissen and Bullemer, 1987) for example, does not necessarily require the interpretation of stimuli. There are four target stimuli, and focusing on any feature that distinguishes the target stimuli can be used without interpretation. In the Task Sequence learning (Koch, 2001; Meier and Cock, 2010), on the other hand, participants have to make category decisions on the input stimuli, and the sequence is formed by features that can only be accessed through the interpretation of stimuli. As reviewed above, the AGL task has a great number of different versions, some of which do not require abstraction (e.g. extraction of TPs over specific syllables, Saffran et al., 1996), while others rely on both abstraction and generalization (e.g. extracting patterns like ABA vs. ABB, Marcus et al., 1999). Since task requirements have a crucial role in determining abstraction of the acquired representation, future research should focus more on carefully manipulating task requirements and taking this factor into account in the interpretation of results.

References

Altmann, G. T. M. and Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks. *Science*, *284*(5416), 875–875.

Altmann, G. T. M., Dienes, Z. and Goode, A. (1995). Modality-independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 899–912.

Aslin, R. N., Saffran, J. R. and Newport, E. L. (1999). Statistical learning in linguistic and nonlinguistic domains. In B. MacWhinney, *The Emergence of Language* (pp. 359–380. Mahwah, NJ: Lawrence Erlbaum Associates Publishers .

Christiansen, M. H. and Curtin, S. (1999). Transfer of learning: rule acquisition or statistical learning? *Trends in Cognitive Sciences*, *3*(8), 289–290.

Cock, J. and Meier, B. (2013). Correlation and response relevance in sequence learning. *Psychological Research*, 77(4), 449–462. http://doi.org/10.1007/s00426-012-0444-x.

Conway, C. M., Bauernschmidt, A., Huang, S. S. and Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition*, *114*(3), 356–371. http://doi.org/10.1016/j.cognition.2009.10.009.

Conway, C. M. and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24–39.

Conway, C. M. and Christiansen, M. H. (2006). Statistical learning within and between modalities – pitting abstract against stimulus-specific representations. *Psychological Science*, *17*, 905–912.

Deroost, N., Zeeuws, I. and Soetens, E. (2006). Effector-dependent and response location learning of probabilistic sequences in serial reaction time tasks. *Experimental Brain Research*, *171*, 469–480.

Destrebecqz, A. and Cleeremans, A. (2003). Temporal effects in sequence learning. In L. Jiménez (Ed.), *Advances in Consciousness Research, Vol. 48. Attention and Implicit Learning* (pp. 181–213). Amsterdam, Netherlands: John Benjamins Publishing Company.

de Vries, M. H., Monaghan, P., Knecht, S. and Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: the learnability of embedded hierarchical structures. *Cognition*, *107*, 763–774.

Dominey, P. F., Lelekov, T., Ventre-Dominey, J. and Jeannerod, M. (1998). Dissociable processes for learning the surface structure and abstract structure of sensorimotor sequences. *Journal of Cognitive Neuroscience*, *10*(6), 734–751.

Eimas, P. D. (1999). Segmental and syllabic representations in the perception of speech by young infants. *The Journal of the Acoustical Society of America*, *105*(3), 1901–1911. http://doi.org/10.1121/1.426726.

Fiser, J. and Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 15822–15826.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I. and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 2458–2463.

Friederici, A. D., Steinhauer, K. and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 529–534.

Frost, R., Armstrong, B. C., Siegelman, N. and Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. http://doi.org/10.1016/j.tics.2014.12.010.

Gómez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, *33*, 154–207.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476.

Gómez, R. L. and Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–35.

Gómez, R. L. and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178–186.

Goschke, T. and Bolte, A. (2007). Implicit learning of semantic category sequences: response-independent acquisition of abstract sequential regularities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 394–406.

Heuer, H., Schmidtke, V. and Kleinsorge, T. (2001). Implicit learning of sequences of tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 967–983. http://doi.org/10.1037/0278-7393.27.4.967.

Hunt, R. H. and Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658–680. http://doi.org/10.1037/0096-3445.130.4.658.

Japikse, K. C., Negash, S., Howard, J. H., Jr and Howard, D. V. (2003). Intermanual transfer of procedural learning after extended practice of probabilistic sequences. *Experimental Brain Research*, *148*(1), 38–49. http://doi.org/10.1007/s00221-002-1264-9.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E. and Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, *110*(2), 316–339. http://doi.org/10.1037/0033-295X.110.2.316.

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

Kemény, F. and Lukács, Á. (2011). Crossmodal transfer and unimodal interference in Artificial Grammar Learning. *Learning and Perception, 3*(Supplement 1), 25.

Kemény, F. and Lukács, Á. (2016). Sleep-independent off-line enhancement and time of the day effects in three forms of skill learning. *Cognitive Processing*, *17*(2), 163–174.

Kemény, F. and Meier, B. (2016). Multimodal sequence learning. *Acta Psychologica*, *164*, 27–33. http://doi.org/10.1016/j.actpsy.2015.10.009.

Kemény, F. and Németh, K. (2017). Stimulus dependence and cross-modal interference in sequence learning, *The Quarterly Journal of Experimental Psychology*, *70*(12), 2535–2547. https://doi.org/10.1080/17470218.2016.1246579

Kirkham, N. Z., Slemmer, J. A. and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–42.

Koch, I. (2001). Automatic and intentional activation of task sets. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1474–1486. http://doi.org/10.1037/0278-7393.27.6.1474.

Lany, J. and Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247–1252. http://doi.org/10.1111/j.1467-9280.2008.02233.x.

Lukics, K. S. and Kemény, F. (2016). Szabályok kiemelése nyelvi és nem nyelvi ingerekből. *Magyar Pszichológiai Szemle* 71(2), 241–256.

Marcus, G. F., Fernandes, K. J. and Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, *18*(5), 387–391. https://doi.org/10.1111/j.1467-9280.2007.01910.x.

Marcus, G. F., Vijayan, S., Rao, S. B. and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*, 77–80.

Meier, B. and Cock, J. (2010). Are correlated streams of information necessary for implicit sequence learning? *Acta Psychologica*, *133*(1), 17–27. http://doi.org/10.1016/j.actpsy.2009.08.001.

Meier, B. and Cock, J. (2012). The role of cues and stimulus valency in implicit task sequence learning: a task sequence is not enough. In A. L. Magnusson and D. J. Lindberg (Eds.), *Psychology of Performance and Defeat* (pp. 155–166). Hauppauge, NY: Nova Science Publisher.

Meier, B., Weiermann, B. and Cock, J. (2012). Only correlated sequences that are actively processed contribute to implicit sequence learning. *Acta Psychologica*, *141*(1), 86–95. http://doi.org/10.1016/j.actpsy.2012.06.009.

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

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

Peña, M., Bonatti, L. L., Nespor, M. and Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607. http://doi.org/10.1126/science.1072901.

Perruchet, P., Tyler, M. D., Galland, N. and Peereman, R. (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology. General*, *133*(4), 573–583. http://doi.org/10.1037/0096-3445.133.4.573.

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

Redington, M. and Chater, N. (1996). Transfer in artificial grammar learning: a reevaluation. *Journal of Experimental Psychology: General*, *125*, 123–138.

Remillard, G. (2003). Pure perceptual-based sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 581–597.

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *47*, 172–196.

Saffran, J. R., Aslin, R. N. and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N. and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.

Saffran, J. R., Pollak, S. D., Seibel, R. L. and Shkolnik, A. (2007). Dog is a dog is a dog: infant rule learning is not specific to language. *Cognition*, *105*, 669–680.

Seidenberg, M. S. and Elman, J. L. (1999). Do infants learn grammar with algebra or statistics? *Science*, *284*(5413), 433–433. http://doi.org/10.1126/science.284.5413.433f.

Seidenberg, M. S., MacDonald, M. C. and Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*(5593), 553–554. http://doi.org/10.1126/science.1078094.

Tunney, R. J. and Altmann, G. T. M. (1999). The transfer effect in artificial grammar learning: reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1322–1333. http://doi.org/10.1037/0278-7393.25.5.1322.

Van den Bos, E. and Poletiek, F. H. (2015). Learning simple and complex artificial grammars in the presence of a semantic reference field: effects on performance and awareness. *Frontiers in Psychology*, *6*, 158. https://doi.org/10.3389/fpsyg.2015.00158.

Weiermann, B., Cock, J. and Meier, B. (2010). What matters in implicit task sequence learning: perceptual stimulus features, task sets, or correlated streams of information? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1492–1509. http://doi.org/10.1037/a0021038.

Weiermann, B. and Meier, B. (2012). Incidental sequence learning across the lifespan. *Cognition*, *123*(3), 380–391. http://doi.org/10.1016/j.cognition.2012.02.010.

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

8

THE VERBALIZATION EFFECT ON IMPLICIT LEARNING

Nadezhda V. Moroshkina, Ivan I. Ivanchei, Anatolii D. Karpov, and Irina Ovchinnikova

Introduction

This chapter is dedicated to the problem of implicit and explicit knowledge relationship during learning. Implicit learning researchers have debated actively on the roles of consciousness and the cognitive unconscious in the learning process for fifty years. This discussion started much earlier and is similar to a certain swing. Now and then new papers confirming the "power" of cognitive unconscious emerge; researchers present new experimental techniques and exciting empirical effects (Reber, 1967; Lewicki, Hill and Sasaki, 1989; Chun and Jiang, 1998; Dijksterhuis, 2004, etc.). However, in a few years the proponents of the "weak" cognitive unconscious find certain drawbacks and flows in the methods used, allowing the conclusion that the knowledge acquired by subjects is at least partially conscious or that the acquired skill is very limited (e.g. Dulany, Carlson, and Dewey, 1984; Perruchet and Pacteau, 1990; Shanks and St John, 1994; Hendrickx, de Houwer, Baeyens, Eelen, and Avermaet, 1997; Newell and Shanks, 2014; Vadillo, Konstantinidis, Shanks, 2015). Cleeremans and Jiménez caricatured the approach of powerful cognitive unconscious proponents as "zombie" theories, and that of the powerful consciousness camp as "Commander Data" theories of cognition (Cleeremans and Jiménez, 2002). In doing so, they correctly pointed out that both in the first case (when the role of consciousness in cognition is reduced to zero) and in the latter (when all cognition is regarded as available to total conscious control) consciousness turns into an epiphenomenon, as it plays no functional role.

Curiously enough, both positions produce a similar prediction: intentional attempts to explicate and verbalize experience acquired during learning should not affect learning process itself. Thus, for a long time verbalization was considered by most of the researchers just as a methodological technique for the assessment of the knowledge acquired in learning.

In our work, we will try to show that verbalization, in fact, is an important factor of learning. It affects both qualitative features of learning and its efficiency. We will review various empirical results consistent with this approach and will also suggest methodological and theoretical implications for the further investigation of implicit learning.

Implicit learning discovery: the dissociation between behavior and verbal reports

About fifty years ago (in fact much earlier: recall the works of Hull (1920) and Thorndike (1932, 1935)) psychologists began to accumulate empirical evidence that learning occurs not only unintentionally but also unconsciously in some tasks. This phenomenon was called implicit learning. Today, there are many experimental paradigms for studying implicit learning. Since Reber's time, the most popular one has been artificial grammar learning ("AGL", Reber, 1967). The main idea of AGL experiments is as follows. In the first stage of the experiment, subjects are asked to memorize a set of 20 letter strings (e.g. PVPXVPS or PTTTVPS, etc.). After that, they are informed that the strings they have learned were compiled by a certain system of rules and then they are presented with a new set of strings, some of which are consistent with this set of rules and some are not. The task is to identify which of the new strings are grammatical, i.e. composed according to the same system of rules. Reber found that subjects correctly (higher than chance level) classify new strings and at the same time are unable to report the grammar rules which they seem to have learned in the memorization phase (Reber, 1967).

By the early 1990s, many experimental studies had been conducted, and researchers discovered various effects of unconscious processing in learning. Berry and Broadbent (1984) published the study in which subjects learned to control a dynamic system modeled by a computer program (e.g. a virtual sugar factory). After sixty control cycles, subjects learned to maintain the parameters of the system within predetermined limits but appeared to be unable to verbalize the rules that governed the factory operation. Nissen and Bullemer (1987) report the results of implicit sequence learning study. Lewicki and colleagues (Lewicki, Hill, Czyzewska, 1992) describe a series of ingenious experiments that investigate unconscious learning of hidden covariations in different perceptual tasks, in particular in social perception tasks, etc. Somewhat later the works of Chun and colleagues demonstrate implicit learning of contextual cues in visual search tasks (Chun and Jiang, 1998).

Reber's first experiments were concluded with a free verbal report. The experimenter asked the participants to tell all they knew about the regularity according to which the experimental stimuli were compiled. Reber interpreted the absence of meaningful verbal reports along with quite precise stimuli classification as evidence for unconscious knowledge used for stimuli classification.

The problem of free verbal reports is that it is quite difficult for the experimenter to assess whether subjects' reports are relevant to the actual grammar rule. It is obvious that none of the subjects would give a full verbal description of the artificial grammar used by the experimenter. Then how can we assess to what extent successful classification of strings as grammatical or not is related to subjects' verbalized knowledge? Formalization of verbal reports as a criterion for conscious knowledge utilized by a person has been implemented in two versions. In a standard AGL experiment, Mathews et al. (1989) asked participants after every ten experimental trials to give verbal instructions on which stimuli features one should rely upon to classify stimuli correctly. Detailed instructions on stimuli classification were collected from every participant. This instruction obtained from each subject was presented to new participants, who had to classify the same artificial strings based solely on these instructions. The new participants classified strings correctly at the above-chance level, but the percentage of correct responses was lower than that of the subjects who created the instructions. This served as evidence that the subjects had acquired certain knowledge that they could not transfer to other people, and therefore, such knowledge was implicit. Another version of formalized verbal reports was presented by Dienes, Broadbent and Berry (1991). Their subjects also composed instructions that were not passed to new subjects but used for direct simulation of subjects' behavior according to the rules only, which they were able to verbalize. This study showed identical results: simulation produced more than 50% of correct responses but significantly lower than human subjects.

However, conclusions made by Reber and other authors who stated implicitness of AGL based on verbal reports quickly became a target of criticism. Long before that, subthreshold perception studies already demonstrated that post-experiment reports are not sufficiently precise as a measuring procedure since, on the one hand, subjects tend to reconstruct their memories on the basis of experimental context, and on the other hand, they tend to underestimate the importance of certain information for the experimenter and decide not to report it (Merikle and Reingold, 1992). Moreover, verbal reports are usually retrospective, thus there is a possibility that some information that was conscious during the task execution can be forgotten at the post-experiment interview. In other words, all the criticism which was directed to introspection at one time can be applied fully to the evaluation of knowledge awareness by verbal reports.

New approaches to measurements of consciousness: refusal from the post-experimental verbal report in favor of online measures

Shanks and St John (1994) tried to formulate criteria for a good awareness test for knowledge acquired through experience. These criteria were later enhanced (Newell and Shanks, 2014). Two criteria were suggested initially: information and sensitivity. The information criterion is related to the relevance of knowledge assessed by the awareness test, and to the learning demonstrated by a person. The question that the experimenter asks a subject in the awareness test should address the knowledge according to which the subject made decisions in the experiment. The sensitivity criterion refers to the amount of conscious knowledge accessed by the test of awareness. The test should be sensitive to all task-relevant conscious knowledge. At least, it should be as sensitive as the test of learning. In this case, a subject's lack of awareness of the knowledge guiding his or her behavior cannot be attributed to the insensitivity of the awareness test.

Newell and Shanks added two more criteria in their 2014 paper: reliability and immediacy; while referring to the information criterion as the relevance criterion. The reliability criterion requires awareness measurement not to be affected by the factors which do not affect the behavior of the subject. The immediacy criterion requires measurement of awareness to occur simultaneously with the measurement of behavior or as close to it as possible.

Using these criteria Newell, Shanks and St John argued that the evidence for learning implicitness in most of the experiments are quite unreliable since certain criteria of the awareness measures quality are not satisfied (Shanks and St John, 1994; Newell and Shanks, 2014). For example, the relevance criterion is often not satisfied in AGL, in their opinion. The authors of the papers claim that subjects use unconsciously learned abstract grammar rules (Reber, 1989; Marcus et al., 1999), but many studies have shown that subjects can memorize the whole strings or fragments, or formulate micro-rules, correlated with the artificial grammar rules (Dulaney et al., 1984; Perruchet and Pacteau, 1990; Johnstone and Shanks 2001). Immediacy criterion requirements are also not met in the majority of early implicit learning studies, as interview used for control of knowledge awareness was held after the main test phase.

Timmermans and Cleeremans (2015) describe requirements for awareness tests in terms of exhaustiveness (the test should be sensitive to all the relevant content of consciousness) and exclusiveness (the test should only be sensitive to conscious knowledge). The first requirement matches the sensitivity criterion discussed above, and the latter points to a new problem: the awareness test should not be affected by unconscious knowledge.

As we see, criticism of awareness measures used in implicit learning studies originates in the 1980s and is still present. In response to this criticism, researchers have developed new measures, which seek to meet the stringent demands of critics.

Since the works of Chan (1991) and Dienes with colleagues (Dienes et al., 1995; Dienes and Berry, 1997) measures based on the idea of the subjective threshold have been used extensively. The objective threshold is a level of learning at which subjects can perform the task at the above-chance level. However, it does not mean that representations responsible for this behavior entered the subject's consciousness. The subjective threshold is a level of learning at which a person knows that he or she possesses some knowledge which enables him/her to perform the task efficiently. In such a situation, a person should be confident in his or her correct responses. More specifically, he or she should have a high degree of confidence in answers that are correct, and a low degree of confidence in answers that appeared to be incorrect. Two criteria were suggested within this approach: the guessing criterion and the zero-correlation criterion. The guessing criterion is met when responses are correct at the above-chance level in trials with confidence rated at zero level (when subjects respond, as they think, at random). The zero-correlation criterion is met when the correlation between confidence ratings and trial correctness does not differ from zero. This approach resolves the main problem of retrospective verbal reports: awareness is assessed simultaneously with learning, therefore satisfying the immediacy criterion. This awareness measure takes into account another important point – that learning is not process-pure and contains many components including both conscious and unconscious. Satisfying the guessing criterion and not satisfying the zero-correlation criterion indicates effects of both conscious and unconscious knowledge on learning.

As we can see, new awareness measures based on subjective confidence assessment meet the immediacy criterion; however, they are still not very informative. The researcher finds out that a subject does have some conscious grounds for responses, but it remains unknown what such grounds are. Therefore, Dienes and Scott (Dienes and Scott, 2005; Dienes, 2012) suggested a more meaningful awareness measure – the decision strategy attribution test, which, in essence, is a structured interview that allows subjects to report the grounds for their decisions during the experiment. The test has been developed for AGL experiments. After every classification decision, a subject has to answer what bases he or she used to make this decision. Four possible options are usually available:

A. guessing;
B. intuition;
C. knowledge of grammar rules;
D. recollection of some strings or fragments from the learning phase.

It is assumed that the first two options (A and B) indicate a lack of explicit knowledge of the bases of decisions made. However, if performance in these trials is at the above-chance level, it demonstrates implicit, that is, unconscious, knowledge. If the subject chooses C or D, it indicates that he or she had relied on explicit knowledge while classifying the string.

However, this approach brings us to the question of how well the subjects' reflection skills must be developed for the test to be used, and whether any pre-training is required for it. It can obviously impose certain restrictions on the use of the decision strategy attribution test.

So, at first, the researchers examined the verbalization process exclusively as an auxiliary methodological procedure that allows us to prove that the subject has no conscious knowledge of any hidden regularities in the task. At the same time, the question of interaction between implicit and explicit was not discussed at first – on the contrary, fundamental inaccessibility of implicitly learned regularities to consciousness was postulated (Lewicki, Czyzewska, and Hoffman, 1987; Reber, 1989).

The proposed alternatives to after-experiment interview are, in fact, additional tasks incorporated in the experimental procedure and aimed at assessing the degree of knowledge awareness that determines subjects' decisions. However, the problem of how these additional tasks may affect the learning process itself and alter its natural course has been barely discussed. Let us note that even David Shanks, the ardent critic of unconscious learning (Shanks and St John, 1994; Newell and Shanks, 2014), does not introduce the non-reactivity criterion among other requirements that he proposes in his works, although in other areas (e.g. when using the thinking aloud method in problem-solving) this criterion is widely discussed (Ericsson and Simon 1993; 1998; Fox, Ericsson and Best, 2011; Schooler, 2011). The essence of the non-reactivity criterion is that an additional parallel task introduced in the learning process (for example, to assess how confident you are in your answer or to report the basis of decisions made) should not change the learning process itself. Describing the immediacy criterion, Newell and Shanks just casually mention: "assessments should be made concurrently (so long as they do not influence the behavior) or as soon after the behavior as possible to avoid forgetting and interference" (Newell and Shanks, 2014, p. 4). Nevertheless, there is a reason to expect that subjects' attempts to analyze the knowledge they apply in the course of the experiment have a significant impact on the learning process itself. We will now review some data supporting this idea.

How methods, provoking verbalization, affect learning and application of implicit knowledge

In the following section, we will discuss the studies indicating that participants' attempts to actively explicate the grounds of their decisions block the application of implicit knowledge. We will focus on three similar phenomena: the effect of strategy, the effect of concurrent verbalization and the effect of retrospective verbalization.

At first, the idea of a powerful cognitive unconscious excited scientists so much that they began to postulate extraordinary speed and complexity of unconscious information processing, fundamentally inaccessible to consciousness (see e.g. Lewicki, Hill and Czyzewska, 1992). However, some studies questioned this claim, primarily because the mere effect of implicit learning was hard to replicate. For example, Hendrickx and his colleagues tried to replicate the effects found by Lewicki, and they succeeded only in one of the nine conceptual replications and three exact replications (Hendrickx et al., 1997). The main critical view suggested by Hendrickx was that subjects in Lewicki's experiments noticed the hidden covariation between appearance and psychological features of the models and made decisions consciously relying on this regularity. In their own research Hendrickx and colleagues used a more detailed after-experiment interview that allowed excluding subjects that noticed the presence of hidden covariation. The results of the remaining subjects did not suggest any learning.

However, Lewicki came up with counterarguments (Lewicki et al., 1997). He accused Hendrickx of incorrectly instructing his subjects, provoking them to look for regularities in the learning phase, whereas in his own experiments subjects were encouraged to respond relying on intuition. Thus, the question of subjects' strategies in implicit learning came into being. A similar impact of subjects' strategy on the presence of implicit learning was found by Lleras and von Mühlenen (2004) in their study of contextual cueing in visual search. In the first two experiments, Lleras and von Mühlenen tried to replicate the experiment by Chun and Jiang (1998) that initially discovered learning of contextual clues. Subjects were presented with displays where they had to find targets among distractors as fast as possible. There were two conditions: "old display condition" where the position of certain targets and distractors was always repeated and "new display condition" where the positions of distractors and targets kept changing constantly. No search acceleration was found for old displays as compared to new ones. Despite the exact reproduction of the experimental conditions, the effect of implicit learning was not manifested. When Lleras and von Mühlenen analyzed the after-experimental interviews, they found that participants used different search strategies: "active" strategy (i.e., an active effort to find the target) or "passive" strategy (i.e., intuitive search). In experiment 3, they instructed participants beforehand to use active or passive strategies. As a result, a stable implicit learning effect was found only in the passive strategy group.

Similar results were found in Reber's early AGL studies. One of his experiments involved varying the type of instruction which was presented to subjects before the learning phase (Reber, 1976). In the standard "implicit" condition, subjects were informed that they would be presented with a set of letter strings for memorization. In the "explicit" condition, subjects were additionally told that all the strings were constructed according to a set of certain rules and figuring these rules out would help them to perform the memorization task better. After that, both groups of subjects participated in the same test phase without feedback. They had to classify the set of 44 new strings presented twice, that is 88 trials in total. The results showed that subjects in both groups demonstrated above-chance classification performance; however, subjects from the "implicit" group performed significantly better than subjects from the "explicit" group. Reber also analyzed patterns of responses to the same item. Since each string was presented twice, subjects could give correct-incorrect, correct-correct and incorrect-incorrect pairs of responses to the same stimulus. It turned out that repeated errors were significantly more prevalent in the "explicit" group (this effect was mentioned above). In addition, subjects in the explicit group tended to classify strings as non-grammatical more often than as grammatical. We will return to this fact later. How did Reber explain these results within his approach to implicit learning?

According to Reber (1976, 1989, 1993), one of the conditions for implicit learning to be manifested is a sufficient degree of complexity of the task performed. Implicit learning will be observed if the subject has no applicable explicit scheme at hand that can be used to code the incoming information. From Reber's perspective, the AGL task is a good reference for such a condition: the artificial grammar rules used are quite complex, unknown to the subject prior to the experiment and cannot be successfully processed explicitly within such a short period of time. When the experimenter expressively encourages the subject to search for applicable rules in such a task, the subject comes up with hypotheses that have no relation to the real rule and application of such rules does not really help him/her when classifying new strings. For instance, one of the subjects in the described experiment always classified strings as non-grammatical if one of the letters appeared more than four times, though this fact did not contradict to the artificial grammar rules (Reber, 1976).

In general, the studies of Reber and others demonstrated that explicit rule search activation prevents one from applying implicit knowledge successfully. Based on our approach, the difference between intuitive (passive, implicit) and analytical (active, explicit) strategy lies in the fact that in the first case, the subject makes decisions based on implicit knowledge that is acquired gradually due to accumulation of frequencies of the stimuli presented and is more procedural in nature. In the latter case, when making a transition to an analytical strategy the subject tries to verbalize explicit grounds for his/her decisions. However, such a transition not only makes him/her use another system of representations (declarative memory) but also process the newly acquired information differently. The person is now relying on internal speech that is discrete in nature and emphasizes specific details. As a result, no learning will be observed until a successful explicit hypothesis on what features are significant is produced. A lot of teachers know that an attempt to explicitly verbalize the basis of a decision made can harm learning. The learners themselves can often complain that when studying the rules of their native language they initially make more mistakes than before, when they were acting (writing and talking) intuitively. A lot of people, when trying to understand what is the correct way to spell this or that word or phrase, can first write it down and then look at it, using their procedural and not declarative memory.

It is worth noting that these processes take place when learning occurs and both declarative and procedural systems possess some knowledge. The notion of strategy we are trying to elaborate is related to which of these two knowledge stores a participant will rely on. Thus we make a distinction between the acquisition of knowledge and its application. This view is in accordance with some of the modern approaches to implicit learning (e.g. Poznanski and Tzelgov, 2010; Witt, Puspitawati and Vinter, 2013; Hendricks, Conway and Kellogg, 2013). In these studies, subjects with the same learning experience exhibit different learning patterns due to test phase influences only. We suggest that learning phase influences can have the same impact on the application of acquired knowledge as the test phase influences. Thus, for example, rule search instruction in the learning phase may affect subsequent test performance in the same way as the test instruction to rely on the explicitly formulated knowledge only.

The studies described earlier showed fade-out of the implicit learning effect as soon as the subjects transited to conscious rule or regularity search. We assume that this search is mediated by inner speech processes (see Vygotsky, 1986 for details). We will now describe studies that provoked usage of external speech during learning. The most systematical studies of competing verbalization were conducted in the dynamic systems control paradigm. This research was initiated by the work of

Berry and Broadbent, (1984) where the authors tried to find performance predictors for the dynamic systems control task. This paper used two dynamic systems, "sugar production factory" and "person interaction", and both of them had the same hidden mathematical regularity in place. The "sugar factory" subjects had to control the number of workers in order to achieve and maintain the sugar production level at 9,000 to 11,000 tons; initially the subjects had 600 workers and 6,000 tons of production. The "person interaction" subjects were "interacting" with a computer called Clegg using a specific interface. For all trials the subjects observed the results of their decisions plotted on a graph.

This study demonstrated that accumulation of experience (one group of subjects managed the factory for 30 trials, and another one had 2 series of 30 trials) significantly improves one's ability to control complex systems but has no effect on questionnaire results. On the contrary, a verbal instruction on how to control the system given prior to the task significantly improves one's ability to answer questions but has no effect on performance. A verbal instruction combined with online verbalization leads to a significant improvement in dynamic system control. Thinking aloud in each trial with no other factors, though, showed no effect on either performance or post-experimental questionnaire results (Berry and Broadbent, 1984).

In 1989, Stanley and colleagues tried to replicate Berry and Broadbent's study and test their hypothesis of the impact that competing verbalization has on performance in a system control task (Stanley, Mathews, Buss and Kotler-Cope, 1989). The experimental group subjects were asked to verbalize grounds for their decisions at the end of each block and they were also told that these verbal reports would be given to newcomers as instructions on how to control the system. Each block consisted of ten trials and the next block had new conditions. The subjects performed 20 blocks of 200 trials in one session. The control group subjects did the same task but had no verbalization to do. The results of the experiment showed that the verbalization group subjects performed better in the sugar factory control task than the subjects in the control group.

McGeorge and Burton (1989) also tried to replicate Berry and Broadbent's experiment with the sugar factory the same year, and their results were inconsistent with the initial study. Experiment 1 involved 3 sugar factory control blocks of 30 trials each, and subjects in the experimental group were silent for the first 2 blocks but had to comment on every decision in the third one, verbalizing all the heuristics they used to select the number of workers. The control group was silent throughout all the three blocks.

Their results showed that the subjects who had to verbalize all their decisions showed learning (when the second and the third block were compared) as opposed to the control group that failed to show any differences in the first, second or third blocks. Intergroup comparison also demonstrated that the verbalization group was better. McGeorge and Burton (1989, p. 463) conclude that parallel verbalization is likely to improve learning. And the absence of learning in the control group is linked to absence of graphical feedback (let us note that the initial experiment conducted by Berry and Broadbent involved a graph showing performance dynamics linked to trial number).

A later paper by McLennan and colleagues (Dickson, McLennan and Omodei, 2000) studied the effect of concurrent verbalization in a micro-world control task, which included a significantly larger number of variables linked together by a hidden regularity. The subjects had to extinguish fires, and time was a crucial factor as if the subject did nothing, the fire would spread. The subjects were randomly assigned to three groups: silence (the subject is silent and comments nothing), associative verbalization (throughout the whole experiment the subject comments whatever comes to his mind) or procedural verbalization (the subject has to come up with grounds for his/her decision for each trial). The results demonstrated that the highest performance was observed in the silent group while the worst was in the group with procedural verbalization.

Thus, the results of the studies are more likely to indicate a change in the learning process itself resulting from competing verbalization within the course of the dynamic system control task. Verbalization is beneficial in some conditions (for example, when an explicit instruction is given prior to the task) and harmful in others (e.g. time pressure). A rich and more "ecological" visual environment in the last experiment may probably better suit implicit learning as well while a poorer visual environment (no graph in the experiments by Stanley et al. as well as McGeorge and Burton, unlike the original sugar factory) together with no time pressure seem to be more suitable for explicit learning.

Y. A. Ponomarev (1960) was one of the first researchers to describe the effect of verbalization. His studies did not follow the paradigms that the implicit learning researchers are used to, though the phenomena he studied resemble implicit learning quite closely. He studied how the introduction of a hint at different stages of performing a creative task alters the subjects' behavior. One of his experiments involved a "polytypic panel" task: the subject had to put a set of bars on a panel following a set of specific rules. The subjects performed this task quite easily and then they were offered the next task: a labyrinth. The optimal track in the labyrinth was exactly identical to the final pattern of bars in the previous "polytypic panel" task. The subjects usually made several dozen mistakes while trying to solve the labyrinth task – they turned around the corner to find a dead end. But when offered the labyrinth task after the "polytypic panel" task they showed a significantly lower amount of errors. An interesting fact: when the subjects were asked to explain their decisions made when choosing a path in the labyrinth, the number of erroneous moves rapidly increased (Ponomarev, 1960).

A similar effect was found by S. Belova in her study devoted to how people evaluate psychological features of others based on their first impression (Belova, 2004). She showed her subjects a video with a number of pre-school kids. The subjects had to make a judgment about each kid and assess his/her IQ. All the subjects were assigned to two groups: the first one had to first assess the intellect of all children and then verbalize their evaluation criteria, while the second group had to make a pause after evaluating half the kids, then verbalize their evaluation criteria and continue to evaluate the rest of the children. When the subjects' evaluations were compared with the children's grades in an IQ test, a correlation was found. But the group that had to verbalize their evaluation criteria in the middle of the experiment showed a decrease in accuracy both in comparison to the first half of evaluations given and to the control group. Belova concluded that during verbalization her subjects failed to explicate some of their knowledge, and after that they would only rely on the verbalized criteria when assessing the intellect of the new kids, and their implicit knowledge had no impact on the evaluations given.

Studies of witnesses' memory showed a similar effect that was named "verbal overshadowing". In short, the experiments are conducted as follows. The subjects are presented with a brief video where a certain person commits a crime. After that the subjects are assigned to two groups: the first one has to describe the criminal (verbalization condition) and the second one performs a distracting task (silence condition). After the task, all the subjects have to try and recognize the criminal in a set of eight photos that replicates the typical police identification procedure. The initial experiment (Schooler and Engstler-Schooler, 1990) demonstrated that subjects that had to describe the criminal verbally recognized him on the photo worse than the subjects who performed a distracting task. In the follow-up studies authors have acquired some contradictory evidence; sometimes the response accuracy decreased, while in other cases the decision-making criterion was altered (i.e. an increase in target misses was observed as the subjects said that the photo of the criminal was not presented).

A recent paper devoted to multi-lab replication of Schooler's initial experiment (Alogna et al., 2014) presented convincing results demonstrating a verbal overshadowing effect, i.e. a decrease in criminal identification accuracy after verbalizing his/her distinguishing characteristics. But this is a relatively short-term effect. Another, even bolder, important result that manifested during replication was a change in the decision-making criterion: the subjects that had verbalized their experience switched to a stricter decision-making criterion (the number of target-miss errors was significantly higher than that of false alarms).

The researchers conclude that an attempt to verbalize the acquired experience cannot be used as a neutral measuring procedure as it leads to a change in subjects' decisions.

Schooler and Fallshore successfully demonstrated a similar effect when studying implicit learning (Fallshore and Schooler, 1993). Some of the subjects in an AGL study were informed of the presence of rules governing the strings they had just been memorizing during the classical learning block, and they then were asked to try and verbalize these rules. The control group subjects had to talk on a neutral topic. Then both groups had to classify new strings in a test block. The subjects that had attempted to verbalize the grammar rules performed significantly worse when classifying new strings than the control group.

So if the subjects' strategy has such a strong impact on learning and applying implicit knowledge, how can we use a subject's behavior to understand what mode he/she is operating in? Is he/she using intuition or trying to apply explicit rules? The easiest way is to ask a direct question of what strategy is being used, and this method was utilized in Dienes and Scott's (2005) decision strategy attribution procedure that was described above. But the data we cited shows that the mere attempt to analyze and report bases of decisions made provokes application of an analytical strategy. Therefore, some objective behavioral markers are required for the experimenter to rely on without directly asking the subjects. We would suggest using the decision-making criterion as one of such markers or in other words the ratio of target-miss errors to false alarms. As research in the field of verbal overshadowing shows, subjects that use verbalizations show a more conservative decision-making criterion. The researchers of implicit learning have paid almost no attention to this parameter. We will try to fill in this gap in the following section.

Methodological aspect: strategy markers

The AGL experiment results are often presented as a 2 by 2 table (see Table 8.1).

This kind of data presentation is widely used in signal detection theory (SDT, Macmillan and Creelman, 2005). Such presentation of results allows us to immediately assess not only classification accuracy ("sensitivity", in SDT terms) but the decision-making criterion as well ("bias"): how often the subject classifies strings as grammatical. The easiest measure is just the share of answers "this string is grammatical" (YesRate) though other indicators are also used (C, beta, etc., see Stanislaw, and Todorov, 1999). This parameter is widely discussed in the literature on decision-making theory (e.g. Fleming, Dolan and Frith, 2012). The bias measures allow deeper inspection of the process under investigation. For example, Whittlesea, Jacoby, and Girard identified different recognition strategies analyzing response bias in recognition decisions (Whittlesea, Jacoby, and Girard, 1990). Rahnev et al. (2011) found that attention can influence visual perception changing subjective bias. Applying SDT analysis to second-order decisions (e.g. confidence reports) was also fruitful, allowing investigation of the mechanisms of human metacognition (Fleming, and Lau, 2014).

A high decision-making criterion (or low YesRate) is usually linked to careful decision-making strategy: the observer must be very confident that the stimulus relates to the target class to answer "Yes". This strategy is called "conservative". An opposite strategy is called "liberal". The subjects avoid false alarms when using conservative strategy and avoid misses when using liberal strategy. When the degree of

TABLE 8.1 An example of the results of an AGL experiment from Reber's work (1976).

Answer	*String*				*Total*
	Grammatical		*Non-grammatical*		
Grammatical	H	263	FA	131	394
Non-grammatical	M	177	CR	309	486

uncertainty is high, when there is not enough explicit knowledge for decision-making and the subjects only have vague and non-verbalizable feelings, conservative-strategy subjects would prefer not to risk and answer "No, this is not the target stimulus" ("non-grammatical string") while liberal-strategy subjects would try to guess and answer "Yes, this is the target stimulus" more often.

We looked at the response bias in AGL literature and found that many papers showed a lower YesRate when the subjects were instructed to search for rules in the training block – as compared to the condition when they were instructed just to memorize the strings. This data is presented in Table 8.2.

Although some experiments were quite different, the decision-making criterion is not discussed by the authors and not specified as a strategy marker. Nevertheless, one might conclude that bias may tell us what strategy a person is using (analytic vs holistic) and which knowledge is used (explicit or implicit). But this assumption requires further testing.

In a series of AGL experiments, we tried to relate the YesRate to the grounds the subjects used to justify their answers while classifying strings as grammatical and non-grammatical. The decision strategy attribution test described above was used. We found that subjects with a more conservative criterion, i.e. a low YesRate, used explicit knowledge sources more often (recalling learning stimuli and testing hypotheses on grammar rules) while subjects with a more liberal criterion, i.e. a high YesRate, referred to guessing and intuition as bases for their answers more often (Ivanchei, 2016). Also, when comparing the subjects' results on the same AGL experiment with and without the attribution test, we found subjects' behavior to be significantly different (Ivanchei and Moroshkina, 2018). Our experiment demonstrated that subjects who performed the attribution test turned out to be more accurate and used a more conservative criterion, answering more slowly than the control group subjects who did not use any concurrent awareness measures. Slower response times can be related to analytical decision-making because deliberate decisions require time to be implemented (Kotovsky and Simon, 1973; Smith, Langston, and Nisbett, 1992). Reber et al. (1980) also found that with incidental instruction participants exhibit much shorter RTs than with rule-searching instruction. At the same time, analysis of strings selected as grammatical showed

TABLE 8.2 YesRate in rule search and memorization conditions (we analyzed experiments allowing extraction of such data).

Study	*Rule Search*	*Memorization*
Reber, 1976	0.44	0.49
Shanks, Johnstone, Le Staggs, 1997 (experiment 3)	0.39	0.50
Shanks, Johnstone, Le Staggs, 1997 (experiment 4)	0.46	0.62
Johnstone and Shanks, 2001 (experiment 1)	0.48	0.48
Johnstone and Shanks, 2001 (experiment 2)	0.46	0.50
Opitz and Hofman, 2015 (experiment 1)	0.47	0.63

that the only predictor of string classification as grammatical in the group with no attribution test was its objective similarity to the stimulus from the learning block, while this was not the case for the group that performed the attribution test. This result indicates emergence of explicit hypotheses on grammar rules and it is consistent with similar data obtained in the works of Scott and Dienes (2008). We propose that it is the requirement to report bases of decisions taken that made the subjects analyze their behavior and use conscious decision-making sources to a larger extent, shaping presumable grammar rules as early as in the test block.

Thus, our data, first of all, serves as evidence for the fact that introduction of an attribution test in the testing procedure itself alters the subjects' strategy and secondly indicates that the response bias can indeed be used as an objective marker for such strategy.

In another study, we introduced two regularities to the stimulus material: we varied font features along with strings grammaticality (Moroshkina and Ivanchei, 2013). Our learning procedure allowed us to let subjects learn the association between grammaticality and certain font features. In the learning phase, subjects saw both grammatical and ungrammatical strings, but the grammatical ones were marked in a certain way allowing subjects to memorize them and not the non-grammatical items. In the control group, this association between font and grammaticality remained the same in the test phase. In the experimental group, this association was disrupted. The two groups did not differ in accuracy. However, subjects in the experimental group along with other attributes of explicit learning showed a more conservative response bias, indicating increased conscious control.

Conclusions

Let us summarize the results of our work. We were considering the question of how the process of verbalization affects the acquiring and usage of implicit knowledge.

1. Analysis of previous papers shows that fifty years of active research in the field of implicit learning applied to different tasks resulted in two opposing camps. The first side believes that all the important cognitive work is performed by the unconscious and consciousness has no access to the acquired experience. The opposite side presumes that cognition is always consciously controlled and acquired knowledge can be explicated with the help of a correctly worded request. Strangely enough, though, both of the camps make a similar prediction: subjects' attempts to explicate and verbalize knowledge acquired through learning should not affect the learning process itself. Thus the verbalization procedure is seen by the majority of researchers solely as a methodological measure to assess the degree of one's awareness of knowledge acquired during the learning process.
2. The main trend in contemporary research in the field of implicit learning is more methodological and empirical. Retrospective verbal reports have been

widely criticized as not meeting a number of criteria (Newell and Shanks, 2014; Timmermans and Cleeremans, 2015) and new methods have been developed to assess knowledge awareness. Alternatives to post-experimental interview proposed later (confidence ratings, post-decision wagering as well as a decision strategy attribution test) are in fact additional tasks included in the testing procedure aimed at assessing a subject's awareness of the knowledge he/she uses to make this or that decision. However, the issue of how these additional tasks can affect the learning process itself and alter its natural course has almost never been discussed, except in research in the field of dynamic systems control where attempts were made to study competing verbalization from the very beginning. But inconsistent results obtained in a number of studies (Berry and Broadbent, 1984; McGeorge and Burton, 1989; Dickson, McLennan and Omodei, 2000) have not been duly theoretically processed.

3. From the very first emergence of implicit learning studies some papers appeared now and then showing impacts of subjects' strategies on acquiring and usage of implicit knowledge. When the researchers provoked a conscious search for hidden rules or regularities, accidentally or on purpose, the implicit learning effect decreased or even disappeared (Reber, 1976; Hendrickx et al., 1997; Lleras and von Mühlenen, 2004). This made some researchers even doubt implicit learning as a phenomenon (Hendrickx et al., 1997). Apart from a change in sensitivity to hidden rules, subjects' tendency to search for rules consciously led to a shift in the decision-making criterion in some studies (Reber, 1976; Shanks, Johnstone and Le Staggs, 1997; Johnstone and Shanks, 2001), but these facts received almost no attention from the researchers.
4. Targeted conscious search for rules or regularities, in our approach (following Vygotsky's works), is mediated by the processes of inner speech. Subjects are trying to form some explicit hypotheses on which features of stimuli allow them to be classified as a target category or rejected. The experimenter's instructions play a significant role here as it is the experimenter who can later serve as means of testing these explicit hypotheses (the subjects almost always ask the experimenter what the correct answer was after the experiment ends). Observed effects (a change in sensitivity to hidden rules and an increase in decision-making criterion accuracy) are similar to verbal overshadowing (Schooler and Engstler-Schooler, 1990; Alogna et al., 2014) and verbalization effects (Ponomarev, 1960; Belova, 2004), observed in a number of other tasks.
5. A change in sensitivity to hidden rules and an increase in decision-making criterion accuracy are related to how speech (internal or external) modulates the subject's attention within the course of task execution. As speech is discrete in nature, a subject's attention is driven to specific features of stimuli presented to the detriment of their integral continual features (analytical vs holistic strategy). If a shift from holistic to analytical strategy is provoked as early as in the test block, there will be no transfer of implicit knowledge accumulated in the test block (this is the explanation of the verbalization overshadowing effect used by Schooler in his latest works (Schooler, 2002); this issue was addressed by

Whittlesea and colleagues in their works on implicit learning and memory (Whittlesea and Dorken, 1993; Whittlesea and Price, 2001). Also, when coming up with verbal hypotheses the subject will rely on his/her memories that are available for report (recollection) and use a priori data and heuristics. As the scope of explicit knowledge does not allow subjects to confidently express all the hidden rules present in the testing material, they are choosing a stricter decision-making criterion.

6. The first methodological consequence from the above-mentioned statements is that a strict decision-making criterion (e.g. YesRate < 50%) seen in implicit learning tasks can be used as an objective marker for analytical strategy. Another consequence is that introduction of additional tasks requiring the subject not only to make decisions in the main tasks but also analyze the bases of such decisions, should provoke analytical strategy. This conclusion was proved in our experiments: introduction of an answer attribution procedure in the test phase resulted in subjects' switch to an analytical strategy and usage of explicit and not implicit knowledge (Ivanchei, 2016; Ivanchei and Moroshkina, 2017).

Acknowledgements

The study was supported by the Russian Foundation for Humanities, project 11355.

References

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahnik, S., Birch, S., Birt, A. R., . . . and Zwaan, R. A. (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, 9, 556–578.

Belova, S. S. (2004). Sub"yektivnaya otsenka intellekta drugogo cheloveka: effekt verbalizatsiy. Sotsial'nyy intellekt: teoriya, izmereniye, issledovaniya [Subjective assessment of intelligence of another person: the verbalization effect]. *Social Intelligence: Theory, Measurement, Research*, 39–62.

Berry, D. C. and Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36(2), 209–231.

Chan, C. (1991). Implicit cognitive processes: theoretical issues and applications in computer systems design. (Doctoral dissertation, University of Oxford.)

Chun, M. M., and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive psychology*, 36(1), 28–71.

Cleeremans, A. and 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.

Dickson, J., McLennan, J. and Omodei, M. M. (2000). Effects of concurrent verbalization on a time-critical, dynamic decision-making task. *The Journal of General Psychology*, 127(2), 217–228.

Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat and J. Williams (Eds.), *Statistical Learning and Language Acquisition* (pp. 337–364). Boston: Mouton de Gruyter Publishers.

Dienes, Z., Altmann, G. T. M., Kwan, L. and Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21(5), 1322–1338.

Dienes, Z. and Berry, D. (1997). Implicit learning: below the subjective threshold. *Psychonomic Bulletin & Review*, 4(1), 3–23.

Dienes, Z., Broadbent, D. and Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 875.

Dienes, Z. and Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, 69(5–6), 338–351.

Dijksterhuis, A. (2004). Think different: the merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology*, 87, 586–598.

Dulany, D. E., Carlson, R. A. and Dewey, G. I. (1984). A case of syntactical learning and judgment: how conscious and how abstract? *Journal of Experimental Psychology: General*, 113(4), 541–555.

Ericsson, K. A. and Simon, H. A. (1993). Protocol Analysis: Verbal Reports as Data (2nd ed.). Cambridge, MA: MIT Press.

Ericsson, K. A. and Simon, H. A. (1998). How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5, 178–186.

Fallshore, M. and Schooler, J. W. (1993). Post-encoding verbalization impairs transfer on artificial grammar tasks. *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Fleming, S. M., Dolan, R. J. and Frith, C. D. (2012). Metacognition: computation, biology and function. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1594).

Fleming, S. M. and Lau, H. C. (2014). How to measure metacognition: frontiers in human neuroscience, 8(443). https://doi.org/10.3389/fnhum.2014.00443.

Fox, M. C., Ericsson, K. A. and Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137, 316–344.

Hendricks, M. A., Conway, C. M. and Kellogg, R. T. (2013). Using dual-task methodology to dissociate automatic from nonautomatic processes involved in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1491–1500. https://doi.org/10.1037/a0032974.

Hendrickx, H., de Houwer, J., Baeyens, F., Eelen, P. and Avermaet, E. (1997). Hidden covariation detection might be very hidden indeed. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 201–220.

Hull, C. L. (1920). Quantitative aspects of evolution of concepts: an experimental study. *Psychological Monographs*, 28(1), i.

Ivanchei, I. (2016). Osoznavaemye i neosoznavaemye processy obrabotki informatsii pri usvoenii iskusstvennoi grammatiki. [Conscious and unconscious processing in artificial grammar learning.] Unpublished doctoral dissertation, Saint-Petersburg University.

Ivanchei I. I. and Moroshkina N. V. (2018). The effect of subjective awareness measures on performance in artificial grammar learning task. *Consciousness and Cognition*. 57, 116–133.

Johnstone, T. and Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology*, 42(1), 61–112.
Kotovsky, K., and Simon, H. A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. Cognitive Psychology, 4, 399–424.
Lewicki, P., Czyzewska, M., and Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 523–530.
Lewicki, P., Hill, T. and Czyzewska, M. (1992). Nonconscious acquisition of information. *American psychologist*, 47(6), 796.
Lewicki P., Hill T. and Czyzewska M. (1997). Hidden covariation detection: a fundamental and ubiquitous phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(1), 221–228.
Lewicki, P., Hill, T. and Sasaki, I. (1989). Self-perpetuating development of encoding biases. *Journal of Experimental Psychology: General*, 118, 323–337.
Lleras, A. and von Mühlenen, A. (2004). Spatial context and top-down strategies in visual search. *Spatial Vision*, 17(4), 465–482.
Macmillan, N. A. and Creelman, C. D. (2005). Detection Theory: a *user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Marcus, G. F., Vijayan, S., Rao, S. B., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77–80.
Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R. and Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: a synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1083.
McGeorge, P. and Burton, A. M. (1989). The effects of concurrent verbalization on performance in a dynamic systems task. British Journal of Psychology, 80(4), 455–465.
Merikle, Philip M. and Reingold, Eyal. M. (1992). Measuring unconscious perceptual processes. In R. F. Bornstein and T. S. Pittman (eds.), *Perception without Awareness* (pp. 55–80). New York: Guilford Press.
Moroshkina, N. V. and Ivanchei, I. I. (2013). Resolving the conflict between two implicitly learned regularities. *Procedia-Social and Behavioral Sciences*, 86, 211–216.
Newell, B. R. and Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *The Behavioral and Brain Sciences*, 37(1), 1–19.
Nissen, M. J. and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, 19(1), 1–32.
Opitz, Bertram and Hofmann, Juliane (2015). Concurrence of rule- and similarity-based mechanisms in artificial grammar learning. *Cognitive psychology*, 77, 77–99.
Perruchet, P. and Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119(3), 264–275
Ponomarev, Y. A. (1960). *Psihologija tvorcheskogo myshleniia* [*Psychology of Creative Thinking*]. Moscow: Prosveshcheniye.
Poznanski, Y. and Tzelgov, J. (2010). Modes of knowledge acquisition and retrieval in artificial grammar learning. *Quarterly Journal of Experimental Psychology* (2006), 63(8), 1495–1515. http://doi.org/10.1080/17470210903398121.
Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P. and Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, 14(12), 1513–1515. https://doi.org/10.1038/nn.2948.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855–863.
Reber, A. S. (1976). Implicit learning of synthetic language. *Journal of Experimental Psychology: Human Learning & Memory*, 2, 88–94.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219.

Reber, A. S. (1993). Implicit Learning and Tacit Knowledge: an Essay on the Cognitive Unconscious (1st ed.). New York: Oxford University Press.

Reber, A. S., Kassin, S. M., Lewis, S. and Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning & Memory*, 6(5), 492–502. doi:10.1037//027393.6.5.492.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology*, 16, 989–997.

Schooler, J. W. (2011). Introspecting in the spirit of William James: comment on Fox, Ericsson, and Best. *Psychological Bulletin*, 137(2), 345–350.

Schooler, J. W. and Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology*, 22(1), 36–71.

Scott, R. B. and Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 34(5), 1264–1288.

Shanks, D. R., Johnstone, T. and Le Staggs, A. (1997). Abstraction processes in artificial grammar learning. *The Quarterly Journal of Experimental Psychology: Section A*, 50(1), 216–252.

Shanks, D. R. and St John, M. (1994). Characteristics of dissociable human learning-systems. *Behavioral and Brain Sciences*, 17(3), 367–395.

Smith, E. E., Langston, C. and Nisbett, R. (1992) The case for rules in reasoning. *Cognitive Science* 16, 1–40. doi:10.1207/s15516709cog1601_1.

Stanislaw, H. and Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, 31(1), 137–149.

Stanley, W. B., Mathews, R. C., Buss, R. R. and Kotler-Cope, S. (1989). Insight without awareness: on the interaction of verbalization, instruction and practice in a simulated process control task. *The Quarterly Journal of Experimental Psychology*, 41(3), 553–577.

Thorndike, E. L. (1932). *The Fundamentals of Learning*. New York, NY, US: Teachers' College Bureau of Publications.

Thorndike, E. L. (1935). *The Psychology of Wants, Interests and Attitudes*. New York: Century.

Timmermans, B. and Cleeremans, A. (2015). How can we measure awareness ? An overview of current methods. In M. Overgaard (Ed.), *Behavioral Methods in Consciousness Research* (pp. 21–46). Oxford: Oxford University Press.

Vadillo, M. A., Konstantinidis, E. and Shanks, D. R. (2015). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*, pp 1–16.

Vygotsky, L. S. (1986). *Thought and Language*. (Trans. E. Hanfmann, G. Vakar and A. Kozulin) Cambridge, MA: MIT Press.

Whittlesea, B. W. and Dorken, M. D. (1993). Incidentally, things in general are particularly determined: an episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, 122(2), 227.

Whittlesea, B. W. A., Jacoby, L. L. and Girard, K. (1990). Illusions of immediate memory: evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, 29(6), 716–732.

Whittlesea, B. W. and Price, J. R. (2001). Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory & Cognition*, 29(2), 234–246.

Witt, A., Puspitawati, I. and Vinter, A. (2013). How explicit and implicit test instructions in an implicit learning task affect performance. *PLOS One*, 8(1), e53296.

9

FOCUSING ON GOAL RELEVANCE – IS IT CRUCIAL TO ARTIFICIAL GRAMMAR LEARNING?

Agnieszka Popławska-Boruc, Radosław Sterczyński, and Marta Roczniewska

Introduction

The aim of the present chapter is to summarize a research project which focused on the importance of goal relevance in the implicit learning process. Contrary to Eitam, Schul and Hassin (2009), we hypothesize that implicit learning occurs without goal activisation. Moreover, if no aim is set for participants, in a dual artificial grammar learning (AGL) paradigm individuals learn letters more efficiently than colors, because the former are more relevant due to e.g., the habit of reading. While Reber and Allen (1978) demonstrated that observational (goal-independent) learning does occur, no previous studies investigated what people learn in fact when they are exposed to two or more grammars simultaneously. To address this literature gap, we begin by discussing theoretical underpinnings of the implicit learning phenomenon, paying particular attention to the AGL paradigm. We then describe in detail Eitam and colleagues' (2009; 2013) studies, including a review of research gaps that follow their design. A description of our three systematically modified auto-replications is then followed by careful consideration of the studies' contribution and limitations. We close the chapter with a discussion of future directions of research.

The implicit learning phenomenon

Implicit learning is a process which allows us to acquire knowledge about regularities which exist in our environment without conscious intention to do so and despite the fact that we have no conscious access to the rules that underlie these regularities (Hendricks, Conway, and Kellogg, 2013; Mealor, Dienes, and Scott, 2014; Reber, 1967). The process of implicit learning is contrasted with explicit learning which is an active, conscious process which requires effort. Implicit learning is a passive process which does not require cognitive investment because it takes place even when we are merely exposed to appropriate stimuli (Hendricks et al., 2013). This process is also automatic, thereby, we are largely unaware of it happening and therefore do not control it, making it fast, effortless and unconscious (Bargh, 1994; Moors and De Houwer, 2006).

Implicit learning has been studied with the use of various paradigms (Pothos and Bailey, 2000; Seger, 1994). It is worth mentioning that approaches to this subject matter are very diverse with respect to the level of rule complexity and the effort which participants must put into completing the task. The most popular paradigms are as follows:

- the sequence learning task,
- the dynamic system control task,
- the AGL task.

AGL is one of the most popular paradigms used to study implicit learning. In the classic version of this task, we start out by generating a finite state of rules governing the sequence of letters in a letter string. This set of rules will allow us to create shorter and longer letter strings and determine the tabbing order of all contained letters (Higham, Vokey, and Pritchard, 2000; Pothos, 2007; Reber, 1967). In the classic procedure, participants in the experiment are first asked to memorize a number of rule-abiding, thereby 'grammatical' ('regular') letter strings, and then to classify letter strings from a different set (albeit controlled by the same grammar) as either 'grammatical' or 'non-grammatical' ('irregular'). In contrast, participants in the control group are either asked to memorize non-grammatical letter strings in the learning phase or do not take part in the learning phase at all, and are immediately asked to classify letter strings as either grammatical or non-grammatical.

Multiple studies have shown that participants in the experimental group are able to correctly classify letter strings above chance level (Chang and Knowlton, 2004; Knowlton, Squire, Champagne, Kronenberg, Willoughby, and Zouzounis, 1996; Zizak and Reber, 2004) or more effectively than the control group which did not take part in the learning phase (Popławska and Wierzchoń, 2008; Pothos, 2007; Vokey and Higham, 2005). This suggests that they have acquired some sort of tacit knowledge during the learning phase. There is an ongoing discussion about the nature of this knowledge, whether it is abstract (Meulemans and Van der Linden, 1997; Perruchet and Pacteau, 1990) and to what extent it is conscious (Higham et al., 2000).

Attention and motivation in the process of implicit learning

Information processing requires cognitive resources. Hirst and Kalmar (1987) define cognitive resources as the fuel, structures, processes, and skills which are necessary to carry out a task, thereby underlining the fact that too often cognitive resources are perceived merely as fuel, a source of mental energy. In this paper, we define cognitive resources as that which limits human performance of a task and can be allocated in varying degrees to different tasks. Cognitive resources can be defined in three areas:

- as mental energy, dependent on biochemical processes (e.g. blood glucose level); these resources are often linked to willpower and perseverance (Muraven, Tice, and Baumeister, 1998);
- as information which is necessary to solve a task (Norman and Bobrow, 1975);
- as data processing structures which are either physical (i.e. parts of the brain; Damasio, 1994) or psychological (i.e. the phonological loop or the visuospatial sketchpad; Baddeley, 1986).

Cognitive resources also include attentional resources, which can be allocated in varying degrees to fit the task at hand. It is therefore important to state what we perceive attention to be – a mental energy resource or rather selection mechanism, which determines what kind of data will be processed in any given moment (Von Hecker, Dutke, and Sedek, 2000). If we assume that implicit learning is an automatic process it follows from there that it does not use up cognitive resources. However, this conclusion is difficult to accept when we consider the complexity of the knowledge acquired in paradigms such as the AGL task or the dynamic systems control task. It has to be said that the results of research on the use of cognitive resources in the process of implicit learning, as well as on the role of attention, have not been conclusive. Investigating the role of attentional resources in implicit learning is usually done by means of the dual-task paradigm. Hayes (1987, in: Dienes, Broadbent, and Berry, 1991) obtained results which suggest that acquisition of tacit knowledge takes place regardless of how much we are invested in the process in terms of the use of our cognitive resources. In his research, he used an extra task, namely the procedure of generating random figures. However, Dienes et al., 1991 were not able to replicate the results of the aforementioned studies. Research by Jiménez and Méndez (1999) showed that study participants are capable of acquiring tacit knowledge during a Serial Reaction Time (SRT) task while simultaneously counting objects that appear on the screen. What is more, participants learn about the links between the primary and secondary task when they are actively engaged in the latter. For example, they can link a shape they are presented with to an area on the screen where the target stimulus will appear. We may say then that the process of implicit learning in the SRT task is independent of the number of cognitive resources which are engaged in it, but is highly dependent on mechanisms of attentional selection which allow participants to grasp the complex structure of the dual task, wherein the primary task is interconnected with the secondary task. Research conducted by Nissen and Bullemer (1987;

in: Jiménez and Méndez, 1999) showed that counting tones which appear during the SRT task impairs the process of implicit learning. On the other hand, research by Cohen, Ivry, and Keele (1990) showed that when the sequence of the strings is simple, implicit learning does take place despite the fact that participants have to attend to two different tasks at the same time. Jiménez and Vázquez (2011) demonstrated that even if we have to attend to two different tasks at any given time, like in the dual-task paradigm, if the two tasks involved are interconnected (e.g. when the secondary tasks allows us to infer where the target stimulus will appear) implicit learning will still take place. What is more, the process is not impaired despite the fact that cognitive resources available to us are reduced simply because we have to perform two tasks at the same time. Some of the few studies conducted on AGL involving a parallel secondary task were those conducted by Hendricks et al. (2013). While memorizing letter sequences, participants were expected to remember the number sequences which had been presented to them on the screen beforehand. As it turned out, the process of implicit learning in this condition was just as effective as it was in the control condition. The only exception pertained to participants who, in the learning phase, had to perform an extra task – in their case the process of implicit learning was impaired. The results of these experiments allowed the researchers to infer that cognitive resources are needed more in the testing phase than in the implicit learning phase of the AGL task. The authors of another experiment showed that cognitive resources are essential in a situation where tacit knowledge has to be transferred. When different letters were used in the learning and testing phase (the grammar stayed the same) participants in the dual-task paradigm were not able to correctly sort letter strings into grammatical and non-grammatical ones.

As we mentioned earlier on in the chapter, cognitive resources are important in the testing phase when participants have to sort letter strings into grammatical and non-grammatical ones. If we ask them during this phase to simultaneously perform an extra secondary task it may force them to change the response strategy they adopted in performing the primary task. This is exactly what Deroost, Vandenbossche, Zeischka, Coomans, and Soetens (2012) observed when they used the Stroop test alongside the implicit learning process. We ourselves obtained similar results (Popławska, Roczniewska, and Sterczyński, 2014) when we presented participants with two sets of artificial grammars, associated with various features of the stimuli as well as sensory modalities. Participants changed their response style to a more conformational one when they were simultaneously presented with two artificial grammars. We may conclude that by simplifying our response strategy individuals make it less resource-intensive.

The presented studies do not offer a conclusive answer to the question regarding the importance of attentional resources in the process of implicit learning. A more precise manipulation of attentional resources is needed, for instance by means of instructing participants to focus on the goal of the task.

Goal relevance in the process of implicit learning

Eitam, Schul, and Hassin (2009) set out to determine what role goal relevance plays in the process of implicit learning. In their studies they used two sets of artificial grammar which were presented to participants simultaneously. One grammar set determined the sequence of letters in the letter strings, whereas the other set governed the colors of the background on which the letter strings were presented. Both grammars were balanced in terms of difficulty and the number of transitions between particular letters and colors, respectively. This allowed the researchers to manipulate the learning process and precisely assess which features of the strings would be memorized – the letters, the colors, or both of these features. The researchers stated in the instruction which feature was meant to be learned and stated whether it was goal relevant or not. Participants were divided into two groups. One group was asked to memorize the order of the letters and to ignore the colors, whereas the other group was asked to do just the opposite – memorize the sequence of background colors and ignore the letters. In the testing phase, the researchers assessed levels of correctness in classifying new strings as either grammatical or non-grammatical among participants for whom a given feature was goal relevant in terms of the instruction they received (because they had memorized either letters or colors) and those for whom it was goal irrelevant in terms of the instruction they had been given. The results allowed the researchers to conclude that participants were better at classifying only that feature of the strings which had been relevant to them in terms of the instruction they received. They completely ignored the other feature. The results of the study were replicated in a follow-up study in which participants were asked to memorize both the order of letters and the order of background colors. An analysis revealed that participants in that study were able to classify new letter strings with above chance level correctness regardless of whether they had to sort them according to the order of colors or letters. Thus, it is clear that they had acquired the complex rules of both sets of grammar. This observation allows us to infer that participants can ignore the feature of the letter strings which is irrelevant to them in terms of their goal. Eitam et al. (2009) very clearly suggest that implicit learning will not take place if we do not inform participants about the goal of the task. What is more, in the second study, implicit learning was impaired, especially in the case of memorizing the order of colors. This may confirm the thesis that implicit learning is to some extent dependent on cognitive resources, in this particular case on attentional resources. In the second experiment, these resources had to be allocated to two different features, and because of this, the effect of implicit learning was weaker than it was in the first experiment. The researchers summed up all the studies with the conclusion that both the effect and the direction of implicit learning are modulated by motivation. In their follow-up studies, Eitam and colleagues (2013) hypothesized that attention is not a prerequisite to implicit learning; what is necessary is setting a goal and focusing participants' attention on that goal. In the AGL task this means instructing participants which feature of the stimulus they are supposed to memorize.

The material which participants are supposed to memorize is located in the attention field, but if a particular feature of the stimulus is irrelevant in terms of the goal then implicit learning will not take place.

Because in previous studies (Eitam et al., 2009) both dimensions subject to implicit learning pertained to different features of the strings – either the order of letters or the sequence of background colors – in yet another replication, Eitam et al. (2013) decided to use equivalent dimensions of the presented stimuli. In order to achieve this they created stimuli material which consisted of colored circles where the order of the outer circle colors was governed by one artificial grammar and the order of the inner circle colors was governed by another artificial grammar. According to the authors, this warranted even allocation of attentional resources to both features (the color of the outer rings and the color of the inner rings) and made sure they were perceived by participants as equally important. Therefore, attention would not be dependent on a particular feature of the stimulus. Neither of the features would be favored, which could have been the case in previous experiments (Eitam et al., 2009). The results of the experiment showed that participants acquired only those rules of artificial grammar which were relevant to them in light of the instruction they had been given. Thus, the group which was supposed to memorize the color sequence of the outer circles classified new strings better if they were based on the grammar which governed this feature of the stimuli, whereas the other group was better at classifying strings which were based on the grammar that determined the color sequence of the inner circles. According to Eitam (Eitam et al., 2013) this is proof that attentional resources alone are not enough to learn the rules of artificial grammar even if the material presented is structured in such a way that both features of the stimuli are located in the participants' field of spatial attention. The results of this experiment were replicated in a follow-up study (Eitam et al., 2013), in which each circle in a circle string was presented to participants separately. In the classic studies within the AGL paradigm, participants can see an entire string (made up of letters, for example) on the screen. In contrast, in this experiment each element of the string was presented separately for about 500 ms and each presentation was followed by a 500 ms break. According to the researchers (Eitam et al., 2013), this was supposed to prevent cognitive overloading in participants and make sure each feature of the stimulus was granted sufficient attentional resources. The experiment was structured in such a way as to exclude the possibility of there being inadequate cognitive engagement for a given stimulus which might occur when participants are presented with entire strings at once. The experiment yielded the same pattern of results as the first experiment. Participants learned only those grammar rules which determined the feature of the strings they were instructed to focus on.

Eitam et al. (2013) summed up their research with the following conclusions:

- implicit learning is related only to those features of the stimulus which are relevant to the goal of the task;

- implicit learning occurs when the features of a stimulus differ between dimensions (letters vs. background colors), and also when the features of a stimulus differ within a dimension (color – in case of circle strings);
- implicit learning occurs with great involvement of cognitive resources when participants have to simultaneously memorize both features of a stimulus or when the cognitive load is not substantial.

It is worth emphasizing that Eitam et al. (2013) discuss a number of studies in which implicit learning took place even though participants were not instructed to focus on any specific goal (Cock, Berry, and Buchner, 2002; Rowland and Shanks, 2006). However, all these studies used a different paradigm than AGL. What is more, every time participants learned grammar rules which governed the stimuli they ignored, it turned out to have been important in terms of the primary task. It helped them predict where the primary, goal-relevant stimulus would appear. Thus, although they were told to ignore certain stimuli, in a wider context these stimuli helped them better predict where the primary, goal-relevant stimulus would appear. These types of experiments were mostly conducted in the SRT paradigm or in the modified version of the Contextual Cuing Paradigm. It is worth emphasizing that the qualitative effect of implicit learning does not depend on the quantity of attentional resources which are engaged in the process (Rausei, Makovski, and Jiang, 2007). Eitam et al. (2009) obtained similar results, when participants were instructed to memorize the order of both kinds of stimuli features – letters and background colors.

What is interesting is that subconscious goal priming can also affect the implicit learning process. Research by Eitam, Hassin, and Schul (2008) demonstrated that subconscious priming of words connected to achievement results in deeper data processing and consequently better results obtained in the implicit learning process. Participants were subliminally primed with words such as ambitious, a race, competiveness etc. Then they were asked to perform a dynamic system control task. Participants who were primed with achievement-related words performed better on this test than people who were subliminally primed with neutral words. This goes to show that priming associated with motivational processes affects implicit learning. In light of these research results, it seems that people can only acquire tacit knowledge when it is relevant to their goal. It can be directly related to it (e.g. an instruction to memorize the order of background colors) or indirectly related to it (e.g. when some feature of the material is related to goal-relevant stimuli and thereby allows us to predict where the primary stimulus will appear next) like in the SRT task, for example.

On the basis of these studies, one cannot conclude that implicit learning is goal-oriented because there was no condition in which the participants were presented with material to memorize without being told to focus on any feature of it (the lack of a control group). Is there a material that could be used in the AGL task which would automatically activate the goal without researchers having to explicitly state it to participants? Eitam and colleagues (2014) demonstrated that human faces may be such stimuli. Although they were completely irrelevant in terms of performing the primary task and participants could therefore ignore them, implicit learning took place and participants memorized the artificial grammar-controlled sequence of faces in the experiment. A similar pattern was not noted when faces were replaced with pictures of buildings. Researchers believe this is because to us humans faces are chronically primed as the target of data processing because they are important in terms of our social functioning. In our previous studies (Popławska et al., 2014) we also demonstrated effective implicit learning of letter sequences in an AGL paradigm without explicitly stating the learning goal to the participants. In two studies conducted participants were instructed to just observe the screen, and still they were able to classify letter strings according to complex grammar rules above chance level.

The above-mentioned studies point to a possibility of implicit learning occurring in the absence of an explicit goal statement (*Hypothesis 1*). However, no previous study investigated whether individuals are able to acquire tacit knowledge when exposed to more than one implicit rule. In order to address this research gap we used the dual-task paradigm and thereby presented participants simultaneously with two artificial grammars without telling them what the goal of the task was. The novelty of the designed study is to show observational learning in the dual-task paradigm, which was never tested before, and to test which feature of the stimuli they are most likely to learn.

Method

Overview of the experiments

In the experiments reported here, we used a *2 AGL* paradigm (see Eitam, Schull, and Hassin, 2009). The typical AGL procedure comprises two phases. During the first phase, the so-called learning phase, participants are presented with a series of objects. The objects, depending on the experimental condition, either comply with or violate specific rules, which are complicated and unfamiliar to participants. Objects presented to participants in the learning phase belong to one of the following categories: regular – rule-abiding – in the experimental condition; or irregular – rule-violating – in the control condition. Participants are instructed either to memorize the objects or just watch the objects. When the learning phase is completed, participants are informed about the existence of rules and asked to classify new objects with respect to these complementary rules. During the second phase, the so-called classification/testing phase, participants are presented with new objects that belong to two categories (regular and irregular) and must determine whether these objects comply with or violate the rules established by the first series. The modal finding is that those participants who are presented with regular objects during the learning phase, correctly classify new material. Typical rates of accuracy in this group significantly exceed chance level and those obtained by controls who are presented with irregular objects during the learning phase. This suggests that participants have learned the grammar or its related regularities (Eitam et al., 2009). Research literature has noted a few AGL paradigm mutations that allow us to observe the process of simultaneous acquisition of two independent sets of rules. Rules provided during the learning phase can be connected with individual properties of the presented objects (e.g. Eitam et al., 2009) or divided into separate objects (e.g. Conway and Christiansen, 2009).

From the participants' point of view, the dual-task AGL paradigm is quite similar to classic AGL. Participants are presented with a series of objects, informed about the existence of a rule and asked to classify new objects. However, the objects are more complicated than in the single version of AGL. The complexity of the task, understood as the number of properties that vary among objects, has to be big enough to differentiate objects with respect to two sets of rules. Properties connected with particular rules can be distinctive, in this case objects are perceived as belonging to two distinct categories (e.g. letters and their background colors; see: Eitam et al., 2009). On the other hand, properties can be closely related – objects are then perceived as more holistic but also more complex (e.g.. colored circles inside bigger colored circles, see Eitam et al., 2013). The property of an object connected with one rule can be more salient than its property connected with another rule. Moreover, participants' attention can be manipulated and directed towards a particular property of the object. However, because both regularities remain implicit, one of them should be interpreted as primary and the other as secondary, which makes this paradigm similar to the typical explicit dual task (Necka, 1997).

There are many possible experimental designs, which can be run in dual AGL, which makes this paradigm useful to explore a wide range of problems, e.g.: implicit processes rivalry (Popławska et al., 2014), modality dependence (Conway and Christiansen, 2009), and the role of attention (Eitam et al., 2009). Complete observation should contain all the possible conditions. Combinations of learning and classification conditions are listed in Table 9.1.

TABLE 9.1 Classification expectations in a 2AGL paradigm.

	Classified material			
	Regular-regular	*Regular-irregular*	*Irregular-regular*	*Irregular-irregular*
Learned material				
Regular-regular	fulfil rules	fulfil/violate rules	fulfil/violate rules	violate rules
Regular-irregular	fulfil rules	fulfil rules	violate rules	violate rules
Irregular-regular	fulfil rules	violate rules	fulfil rules	violate rules
Irregular-irregular	irregular	irregular	irregular	irregular

Study 1

Method

Participants

A total of 121 volunteers; 66 women, 55 men, mean age M = 25.2 years, (SD = 4.28), participated in the experiment. Participants were recruited from employees of two companies (IT and textile) or were students of Economics. Participants took part in the study individually and were randomly assigned to one of the four conditions. The experiment was conducted on a laptop computer in a room assigned by the coordinator. During the experiment, nobody aside from the experimenter and the participant was allowed to be inside the room.

Procedure

Participants first completed the learning phase which was followed up with information about the existence of rules, and then went on to complete the classification phase (Reber, 1967). In the classification phase, the participants were instructed to classify new objects that varied in two dimensions (color sequence order and letter sequence order), as either grammatical or non-grammatical.

Training phase. In each trial during the learning phase objects constructed as sequences of colored letters appeared in the centre of the screen for 5 seconds. The sequences of letters were determined by one grammar set and the sequences of colors by another. Each stimulus appeared only once in the course of the learning phase. Unlike in other other studies (e.g., Eitam et al., 2009; Reber, 1967; Vokey and Brooks, 1992), we did not create separate pools of objects for the learning phase and classification phase. Stimuli were not rigidly assigned to phases, to prevent constant bias towards a particular classification.[1] Objects presented during each phase were randomized trial by trial. Thus, each participant was presented with an individually assigned set of objects in an individually assigned order. Participants were instructed to closely observe the presented objects but we did not instruct them to focus on any particular properties of the presented objects.

Test phase. Following training, participants were informed that the objects they saw adhered to a complex set of rules. They were then asked to categorize 64 new objects as either grammatical or not. A total of 16 test stimuli were grammatical in both dimensions, whereas 16 others violated both grammars, 16 stimuli were grammatical in terms of the color sequence order and ungrammatical in terms of the shape/meaning sequence order, and 16 stimuli ware ungrammatical in terms of the color sequence order and grammatical in terms of the shape/meaning sequence order (see Table 9.2). In both dimensions, the ungrammatical stimuli started and ended with a grammatically correct letter. The test objects were presented in the centre of the screen until participants responded. No feedback was given, in order to minimize explicit learning during the test phase (e.g., Eitam et al., 2009).

Material and equipment

The objects presented to participants during both phases were letter strings in bold Arial font size 60. It was created from two grammars regulating the sequences of letters (Grammar 1) and colors (Grammar 2). The two grammars were comparable in their complexity, the first one contained sixteen potential paths between nodes and three recursions, the second one contained eighteen potential paths between nodes and no recursions (see Figure 9.1).

A total of 172 different sequences of letters were generated from Grammar 1 and a total of 130 different sequences of colors were generated from Grammar 2. For each regular (rule-abiding) sequence in both grammars, its irregular (rule-violating) equivalent was generated. The ungrammatical equivalents were equal in length and started and ended with the same element as their grammatical counterparts. The length of sequences generated from both grammars ranged from 3 to 8 elements. Objects presented to the participants were randomly combined trial by trial, with the restriction of the color and letter sequences being equal in length. The randomization procedure consisted of three steps. At the beginning of the trial, the length of the sequences was selected with respect to

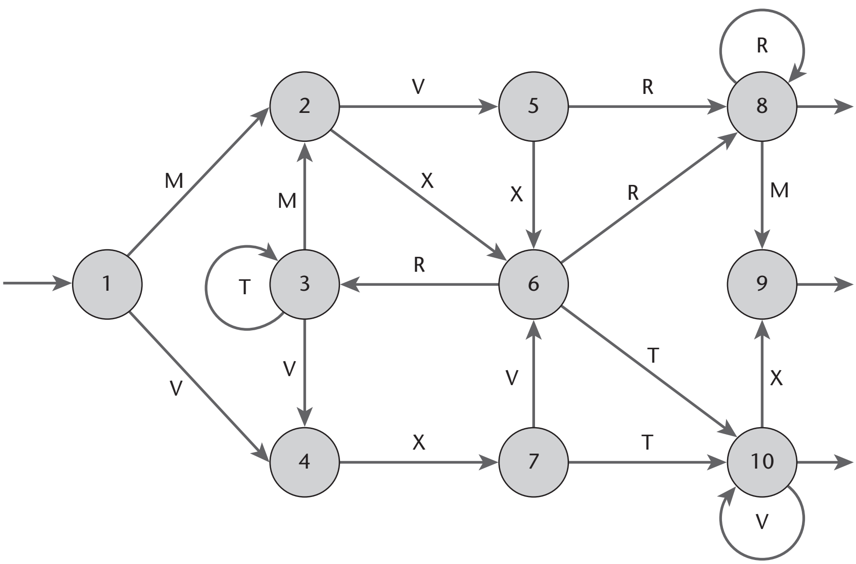

FIGURE 9.1A Grammars governing objects used in all three referred studies. Grammar 1 – letter sequence used in Studies 1–3.

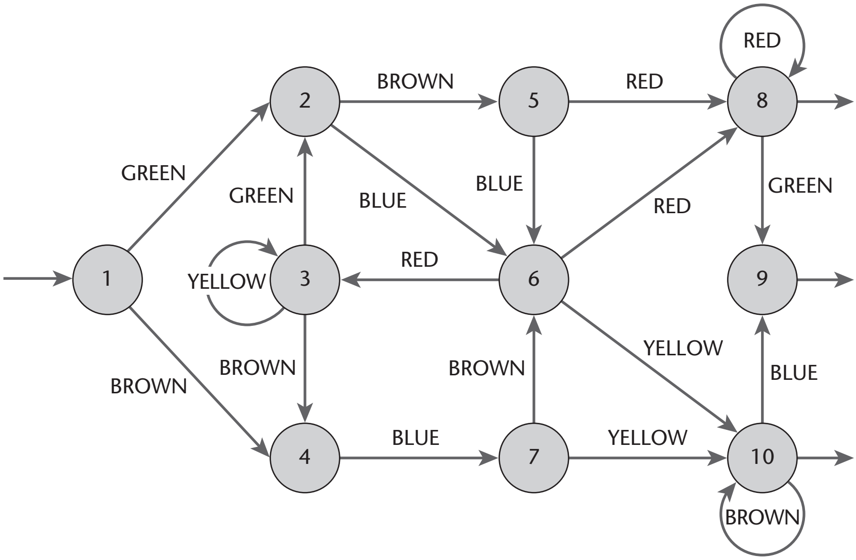

FIGURE 9.1B Grammar 1 – color sequence used in Studies 2–3.

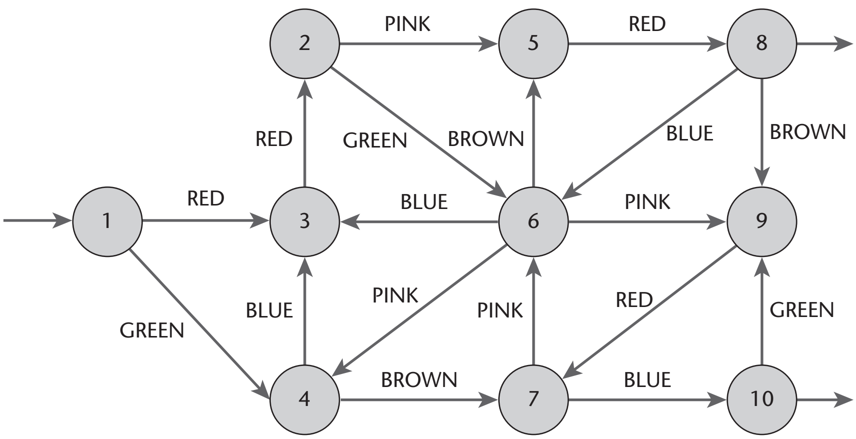

FIGURE 9.1C Grammar 2 – color sequence used in Study 1.

the number of each length sequences in the whole pool. Next, the letter order within the selected length sequences was drawn. Finally, the color order within the same length sequences was drawn. All three randomization steps were run without replacement, so each sequence presented to the participant appeared only once during whole experiment. The experiment was run in Inquisit 3.0.xxx on a laptop PC.

Design

The experiment had a 2 (learned letters: grammatical vs. ungrammatical) × 2 (learned colors: grammatical vs. ungrammatical) × 2 (accuracy criterion: letters vs. colors) mixed design (see Table 9.2). The sequence of letters was governed by Grammar 1; the sequence of colors was governed by Grammar 2. The sequence of letters and colors in the objects presented during the learning phase was manipulated between participants. Measures of knowledge acquired from each grammar were calculated within the same data, which made this variable a within-group factor.

Results and discussion

Classification data was computed into two separate variables. The answers YES (the object follows the rules) and NO (the object violates the rules) were recoded into two separate variables indicating the correctness of classification referring to each rule. The recoding scheme is described in Table 9.2. The percentage of correct answers was used as an indicator of learning particular grammar.

The primary aim of the study was to verify if IL occurs when people are exposed to two rules simultaneously without directing their attention to this fact. To answer this question we computed a three-way mixed design ANOVA: 2 (learned letters: grammatical vs. ungrammatical) x 2 (learned colors: grammatical vs. ungrammatical) x 2 (accuracy criterion: letters vs. colors), with the latter being with-subject. The results revealed a two-way interaction: learned letters x accuracy criterion $F(1, 117) = 3.91$ $p = .05$ [*eta2*] = .03. Participants who observed grammatically sequenced letters during the learning phase, classified objects presented during the test phase with greater accuracy with respect to Grammar 1, $M = 52\%$ ($SD = 6$) than participants who observed ungrammatical letter sequences during the learning phase $M = 49\%$ ($SD = 5$), $t(119) = 2.59$ $p = .01$ $d = .48$ (see Table 9.3). What is more, the percentage of correct answers obtained by participants who observed grammatical letter sequences during the learning phase was higher than chance level $t(59) = 2.38$ $p = .02$ $d = .31$, whereas the percentage

TABLE 9.2 Different types of objects classification recode scheme according to each grammar criteria.

Criterion	*Decision*	*Object*			
		Regular letters		*Irregular letters*	
		Regular colors	*Irregular colors*	*Regular colors*	*Irregular colors*
Grammar 1	fulfils	correct	correct	incorrect	incorrect
(letters)	violates	incorrect	incorrect	correct	correct
Grammar 2	fulfils	correct	incorrect	correct	incorrect
(colors)	violates	incorrect	correct	incorrect	correct

TABLE 9.3 Learned letters × accuracy criterion 2×2 interaction.

Learned letters sequences (Grammar 1)	*Classification correctness*				
	Grammar 1 (letters)		*Grammar 2 (color)*		*n*
	Correct	*SD*	*Correct*	*SD*	
Ungrammatical	49%	5.1%	50%	5.7%	61
Grammatical	52%	5.5%	50%	6.1%	60

Note: Correct = mean percentage of correct classifications.

of correct answers obtained by participants who observed ungrammatical letter sequences during the learning phase did not differ significantly from chance level $t(60) = 1.24$ $p = .22$ $d = .16$. The two groups of participants classified objects presented during the test phase with respect to Grammar 2 – which governed their color sequence – equally well: $t(119) = .32$ $p = .75$ $d = .06$; their level of accuracy did not differ much from chance level and was: $M = 50\%$ ($SD = 6.1$), $t(59) = .43$ $p = .67$ and $M = 50\%$ ($SD = 5.7$), $t(60) = 0$ $p = 1$, respectively .

An analogous interaction between learned colors and accuracy rates was statistically insignificant: $F(1, 117) = 3.91$ $p = .05$ [*eta2*] = .03. Participants who observed objects with grammatically sequenced colors during the learning phase did not classify objects more accurately ($M = 50\%$ ($SD = 5.7$)) than participants who observed ungrammatical objects ($M = 49\%$ ($SD = 6{,}0$)) with respect to Grammar 2: $t(119) = .94$ $p = .35$ $d = .17$. The percentage of correct answers obtained by both groups of participants did not differ significantly from chance level and was, respectively: $t(59) = .46$ $p = .65$ $d = .06$ and $t(60) = .86$ $p = .39$ $d = .11$. Participants who, during the learning phase, observed objects with grammatically sequenced colors did not classify objects more accurately ($M = 50\%$ ($SD = 5.5$)) than participants who observed non-grammatical objects ($M = 49\%$ ($SD = 5.4$)) with respect to Grammar 1 – which governed its letter sequences: $t(119) = .80$ $p = .43$ $d = .15$. The percentage of correct answers obtained in both groups of participants did not differ from chance level and was: $t(59) = .04$ $p = .97$ $d = .005$ and $t(60) = 1.18$ $p = .24$ $d = .15$, respectively.

TABLE 9.4 Learned colors × accuracy criterion 2×2 interaction (insignificant).

Learned color sequences (Grammar 1)	*Classification correctness*				
	Grammar 1 (letters)		*Grammar 2 (color)*		*n*
	Correct	*SD*	*Correct*	*SD*	
Ungrammatical	51%	5.4%	49%	6.0%	61
Grammatical	50%	5.5%	50%	5.7%	60

Note: Correct = mean percentage of correct classifications.

The results provide support for the hypothesis that inclination of AGL processing towards letter sequences is higher than inclination towards color sequences.

The results of Study 1 demonstrate that goal formulation is not a necessary condition for implicit learning to occur. In line with Hypothesis 1, participants were able to learn the hidden regularities, although no explicit instructions as to what to learn (or focus on) were given. Moreover, it appears that out of the two rules that were presented, participants were able to acquire the grammar of letters, but not the grammar of colors. We think it is because reading is more automatic than color-naming[2] – letters attract greater attention as more relevant features of the stimuli than colors. Similarly to faces (Eitam et al., 2014), letters may not need explicit goal pursuit instructions. Therefore, our next hypothesis is as follows:

Hypothesis 2. People acquire grammar related to letters more effectively than grammar related to colors.

Obviously, both relevant and irrelevant information in the perception field is being processed simultaneously. In theory, participants in this study could learn the two grammars. However, the battle over cognitive resources available to the attention process led to their allocation to selected information. Contrary to a study by Eitam and colleagues (2009), in our study we did not explicitly state our goals to participants. Participants were initially asked to '*observe*' the screen; we did not point their attention towards either the letters or the colors. This allowed us to observe whatever became relevant to participants when their attention was not directed at anything in particular.

However, we cannot rule out the possibility that – because letters and colors operated on different grammars – the grammar of letters was easier to acquire and apply than the grammar of colors. Given this assumption, we planned and executed Study 2, wherein both letters and colors were placed according to exactly the same grammar.

Study 2

Method

Participants

A total of 40 participants; 33 women, 7 men, mean age $M = 31.1$ years, ($SD = 7.16$), participated in the experiment. Participants were university students and were given extra credit for their participation. They took part in the experiment one at a time.

Procedure

The procedure was identical to that of Experiment 1 except for one change. Both dimensions of objects' variance were determined by the same grammar. A total of 172 different sequences of letters were generated from Grammar 1 and the same number of different sequences of colors were generated also from Grammar 1 (see Figures 9.1a and 9.1b). The presented objects were a combination of color and letter sequences rendered exactly like in Experiment 1.

Design

Because Grammar 1 worked well in Study 1 (compared to the control condition and chance level) we decided to reduce the experiment to a single group within subject design.[3] All participants observed grammatical objects which abided by Grammar 1 in both its dimensions: color sequence and letter sequence. To verify and compare the effectiveness of learning, within subject t-test and one sample t-test testing differences from chance level, was planned.

Results and discussion

Just like in Experiment 1, data was computed into two separate indicators of correctness with respect to each dimension (see Table 9.2). P-value is presented as one-tailed, according to previous observation.

Correctness referring the letter sequence $M = 54\%$ ($SD = 6.3$) was significantly higher than chance level $t(39) = 3.56$ $p = .001$ $d = .56$. Correctness referring to color sequence $M = 51\%$ ($SD = 6.3$) did not differ from chance level $t(39) = .90$ $p = .37$ $d = .14$. The percentage of correct answers with respect to letter sequence was significantly higher than the percentage of correct answers with respect to color sequence $t(39) = 1.73$ $p = .05$ $d = .27$.

The results presented here are consistent with the main finding of Study 1 – no explicit instructions have to be given to participants for them to acquire the implicit knowledge of rules governing the presented items. This confirms Hypothesis 1. Again, the results demonstrated that – when exposed to two rules simultaneously – people spontaneously focus on one of them. Namely, participants classified grammatical strings as such with greater accuracy when the strings were grammatical with respect to letters; in the case of strings which were grammatical with respect to colors, participants were not able to perform above chance level. The results of this experiment confirm Hypothesis 2, and are consistent with previous findings – people 'chose' the grammar of letters over the grammar of colors, even though both letters and colors were in their field of attention, and we did not tell them to focus on either aspect of the strings.

Because we used exactly the same grammar for both letters and colors, we cannot explain the observed effect with grammar difficulty. However, it is important to underline that colors may have been perceived as a secondary feature of the letters (font color is one of the features of the letter, just like font size or font style). This means that the design of the stimulus presentation in the two studies described so far may have predisposed participants to treat letters as more important than colors in the perception field. Taking this into consideration, we conducted Study 3, wherein colors were 'extracted' from the letters, forming rectangles placed behind black letter strings. We expected that, regardless of how the colors were presented, individuals would learn the grammar of letters more effectively than the grammar of colors (Hypothesis 2).

Study 3

Method

Participants

A total of 31 participants, 22 women and 9 men, mean age $M = 29.46$ years ($SD = 7.12$), participated in the experiment. Participants were university students. They were given extra credit for their participation in the study and took part in the study one at a time.

Procedure

The procedure and experimental design were identical to that of Experiment 2.

The only difference lay in the objects that were used. In both of the previous experiments, objects were colored letters. In this experiment, objects were black letters in Courier font presented on colored backgrounds. Both the sequence of letters and the sequence of background colors were governed by Grammar 1.

Results and discussion

The data was computed identically to both of the previous experiments (see Table 9.2).

Correctness with respect to letter sequence $M = 53\%$ ($SD = 5.8$) was significantly higher than chance level $t(30) = 2.53$ $p = .02$ $d = .45$, and again the correctness with respect to background color sequence M = 52% ($SD = 7.1$) did not differ significantly from chance level $t(30) = 1.35$ $p = .19$ $d = .24$. However, the percentage of correct answers with respect to the letter sequence did not differ significantly from the percentage of correct answers with respect to background color sequence $t(30) = .51$ $p = .61$ $d = .09$.

The results obtained in Study 3 are consistent with our previous findings: in the presence of two implicit regularities, individuals seem to be able to acquire only one set of rules when they are not instructed to pay attention to a particular feature of the presented objects. This might suggest certain limitations of implicit learning. Again, we observed that it was the grammar of letters that was learned and subsequently applied (as compared to chance level performance), not the grammar of colors. Nevertheless, in this experiment the accuracy of classification with respect to letters did not differ significantly from the accuracy of classification with respect to colors. This interesting finding brings to focus the role of dimension separation as a means of making them equally important to the recipient.

General discussion

Studies' contribution

Eitam and colleagues (2009) argue that AGL is selective, and – consequently – learning occurs for goal-relevant dimensions only. In their studies, however, the goals were externally set for participants – they had been informed beforehand what to focus on. To expand on their idea we demonstrated that participants need not be told directly to focus their attention on a particular feature of an object and also that their previous experiences, habits, or motivational standards may influence implicit information processing. This points somewhat to the top-down approach to understanding implicit learning processes.

Possibly because the grammars used in the studies were complex, learning them both equally well may have been too difficult a task for participants' capacity. To some degree, implicit learning is resource-consuming (Dienes et al., 1991). In their line of research Shanks, Rowland, and Ranger (2005) demonstrated that implicit learning is debilitated under conditions of divided attention – having to perform a secondary task impaired participants' sequence learning. In our previous studies, we showed that in the presence of another implicit rule, people's effectiveness with regard to the primary AGL task was not affected; however, their decision-making strategy changed towards a simpler one, indicating a type of trade-off (Popławska et al., 2014). Also, Eitam and colleagues (2009; Experiment 2) demonstrated that participants can simultaneously learn, and then apply, two implicit grammars; however, in this case the effect of learning was in fact weaker. The aforementioned studies thus point to the resource-consuming nature of implicit learning, which calls for attention to be allocated to certain aspects of objects in the perception field. Such selection might be based on many distinct criteria. For one thing, the dimensions may be attended to according to their relevance to individuals' current goals, giving priority to some subset of representations over others (Eitam et al., 2009). It was Bruner (1957) who postulated and empirically demonstrated that goal activation creates perceptual readiness, leading to a faster identification of goal-relevant information. Multiple studies conducted since then have shown that goal-relevant objects become more accessible, and are implicitly assessed as more positive than goal-irrelevant objects when the goal is active (e.g., Brendl, Markman, and Messner, 2003; Ferguson and Bargh, 2004; Roczniewska and Kolańczyk, 2014). Furthermore, self-regulation standards that people hold tend to determine what becomes essential for goal pursuit. For example, ideals may predispose people to focus on objects coherent with current motives, whereas oughts may prompt individuals to adapt a monitoring strategy that involves attending to hindrances (Kolańczyk and Roczniewska, 2015). Finally, previous experiences and habits may predispose people to focus on certain aspects of the stimuli in the visual field, which can further affect their judgment. For instance, the prevalence of vertical over horizontal symmetry perception is present among sighted (not blind) participants, suggesting it is derived from visual experience (Cattaneo, Fantino, Silvanto, Tinti, Pascual-Leone and Vecchi, 2010). Also, the habit of reading in a particular direction can affect the way people scan the perceptual field, regardless of whether it actually contains letters (Chokron and De Agostini, 2000). Namely, left-to-right readers express a preference for stimuli with a rightward directionality, whereas right-to-left readers – a leftward directionality.

The studies described in this chapter demonstrate that implicit learning can occur even in the absence of an explicitly formed goal. Participants in all three studies were instructed to just observe the screen (not memorize the items); yet, they were able to acquire tacit knowledge about the rules governing the presented objects. However, we observed that only one of the two regularities was acquired and applied by participants. This raises a question of attention's selection criteria – what prioritizes letters over colors? The exact mechanism predisposing individuals to focus on letters rather than colors was studied in the experiments described. Possibly, Latin letters, although forming senseless chunks, activate the habit of reading. This, in turn, prioritizes the processing of information related to letters over colors. A context wherein letters appear on a screen may automatically elicit the behavior of reading. Such habits do not require awareness, intention, or control. Although reading does not require controlled attention, it nevertheless uses enough attentional resources (especially since the chunks do not form meaningful words) to reduce the amount of attention accessible for color information processing. This took place not only in all of our three studies, but in other similar studies – Eitam and colleagues (2009) reported that grammar underlying letter strings was generally learned more effectively than that governing color strings, regardless of instructions provided by experimenters.

Limitations and future research

Although the present studies contribute to the literature on goals and implicit learning, some limitations need to be acknowledged. For one thing, the levels of learning in all three experiments have been comparatively lower than those observed in other papers (e.g., Eitam et al., 2009; Scott and Dienes, 2008). This may either be explained by participants' low motivation, or the difficulty of grammars that were to be acquired. Hence, it is possible that these factors are partially responsible for the trade-off effect that we observed. As for the motivation, individuals who partook in Studies 2 and 3 were rewarded for their participation with course credit; however, this incentive cannot be seen as a warranty of participants' high motivation as they were 'remunerated' regardless of their performance. Eitam, Hassin and Schul (2008) demonstrated that – although implicit learning processes are involuntary – achievement motivation can increase performance in the AGL task. Hence, this explanation regarding low motivation cannot be ruled out, and the effect we observed should be tested in future studies with motivation as its moderator. Regarding grammar complexity, each of the grammars we applied consisted of 10 nodes and 16 to 18 possible paths. Indeed, when we compare it to other studies administering the AGL paradigm, we may observe that a great deal of them used less complex structures (5–6 nodes; see: Eitam et al., 2009; Scott and Dienes, 2008). Also, the 16 items in our studies were presented only once during the learning phase; other researchers (e.g., Witt and Vinter, 2012) present more stimuli or repeat them, which allows participants to learn more effectively. The difficulty and complexity of the grammar serves as an explanation for the lower testing phase results. Therefore, it seems possible that in a condition of lower cognitive load (simpler grammars) the trade-off effect would have been smaller, allowing participants to learn both grammars. However, this limitation does not preclude the main finding in this study: people are able to acquire implicit regularities without being instructed to focus their attention on specific dimensions.

An interesting finding that was replicated across all three of our studies is that, in general, participants were able to acquire the grammar of letters, but not the grammar of colors; neither when the font of a letter string was a particular color nor when the colors appeared as rectangles in the background of the letters. This effect cannot be explained by goal relevance (as no goal was formulated in the instruction) or self-regulation standards, but rather by habits; namely the habit of reading, activated by the presence of letters. The above argumentation points to the possibility that if the letters do not form a familiar alphabet (e.g., the Arabic alphabet for an average Western citizen), they may not engage attention as strongly, hence leaving a possibility for colors to become most important, or at least become equally likely to be selected. Future studies should thus compare cross-cultural (based on alphabet) effects of 2AGL learning tasks with respect to colors and letters.

Conclusion

To conclude, we suggest that the relevance of dimensions in implicit learning is not only a function of explicit goals, but also habits that govern human behavior. We demonstrated that reading can predispose individuals to focus on letters over other stimuli in the perception field, making other dimensions less attended to and therefore harder to acquire.

Acknowledgements and funding

This research was supported by the Polish Ministry of Science and Higher Education, Grant no N N106 282239, and SWPS University of Social Sciences and Humanities, Sopot Campus, Grant no WSO/2014/A/09. We thank students Magdalena Stefańska-Lotkowska, Patryk Wójcik, and Barbara Szymańska for collecting data.

Notes

1 In most studies on AGL, objects are rigidly allocated to particular phases of the experiment (the learning phase or the testing phase). However, there is a risk that doing so will distort the results because there is a risk of constructing two pools of objects which might mutually interfere with each other and thereby also with the process of implicit learning

because of object features that researchers do not control in the study. Namely, participants may classify objects in the testing phase as regular or irregular not on the basis of their tacit knowledge of the rules but because of a shortcut they associate with a given letter sequence or because a flag is similar to a color sequence. The risk of tainting participants' freshly acquired tacit knowledge with this mistake is all the more real because most of the reported studies within the AGL task paradigm are generic in their nature.

2 This theory is the most common theory which explains the Stroop effect.

3 In a pilot study we demonstrated the effectiveness of implicit learning of such grammar in the case of a single AGL for both colors and letters.

References

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Bargh, J. A. (1994). The Four Horsemen of Automaticity: awareness, efficiency, intention, and control in social cognition. In R. S. Wyer and T. K. Srull (Eds.), *Handbook of Social Cognition* (Vol. 1, pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brendl, C. M., Markman, A. B., and Messner, C. (2003). The devaluation effect: activating a need devalues unrelated objects. *Journal of Consumer Research*, *29*(4), 463–473. http://doi.org/10.1086/346243.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152. http://doi.org/10.1037/h0043805.

Cattaneo, Z., Fantino, M., Silvanto, J., Tinti, C., Pascual-Leone, A., and Vecchi, T. (2010). Symmetry perception in the blind. *Acta Psychologica*, *134*(3), 398–402. http://doi.org/10.1016/j.actpsy.2010.04.002.

Chang, G. Y. and Knowlton, B. J. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 714–722. http://doi.org/10.1037/0278-7393.30.3.714.

Chokron, S. and De Agostini, M. (2000). Reading habits influence aesthetic preference. *Cognitive Brain Research*, *10*(1–2), 45–49. http://doi.org/10.1016/S0926-6410(00)00021-5.

Cock, J. J., Berry, D. C., and Buchner, A. (2002). Negative priming and sequence learning. *European Journal of Cognitive Psychology*, *14*(1), 27–48. http://doi.org/10.1080/09541440042000151.

Cohen, A., Ivry, R. I., and Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 17–30. http://doi.org/10.1037/0278-7393.16.1.17.

Conway, C. M., and Christiansen, M. H. (2009). Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*(4), 561–580. http://doi.org/10.1080/09541440802097951.

Damasio, A. R. (1994). *Descartes' error: emotion, rationality and the human brain*. New York: Putnam (Grosset Books).

Deroost, N., Vandenbossche, J., Zeischka, P., Coomans, D., and Soetens, E. (2012). Cognitive control: a role for implicit learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1243–1258. http://doi.org/10.1037/a0027633.

Dienes, Z., Broadbent, D. E., and Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 875–887. http://doi.org/doi:10.1037/0278-7393.17.5.875.

Eitam, B., Glass-Hackel, R., Aviezer, H., Dienes, Z., Shoval, R., and Higgins, E. T. (2014). Are task irrelevant faces unintentionally processed? Implicit learning as a test case. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1741–1747. http://doi.org/10.1037/a0037627.

Eitam, B., Glicksohn, A., Shoval, R., Cohen, A., Schul, Y., and Hassin, R. R. (2013). Relevance-based selectivity: the case of implicit learning. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1508–1515. http://doi.org/10.1037/a0033853.

Eitam, B., Hassin, R. R., and Schul, Y. (2008). Nonconscious goal pursuit in novel environments: the case of implicit learning: research article. *Psychological Science*, *19*(3), 261–267. http://doi.org/10.1111/j.1467-9280.2008.02078.x.

Eitam, B., Schul, Y., and Hassin, R. R. (2009). Goal relevance and artificial grammar learning. *Quarterly Journal of Experimental Psychology*, *62*(2), 228–238. http://doi.org/10.1080/17470210802479113.

Ferguson, M. J. and Bargh, J. A. (2004). Liking is for doing: the effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology*, *87*(5), 557–572. http://doi.org/10.1037/0022-3514.87.5.557.

Hayes, N. A. (1987). Systems of Explicit and Implicit Learning. Unpublished doctoral dissertation, University of Oxford, England.

Hendricks, M. A, Conway, C. M., and Kellogg, R. T. (2013). Using dual-task methodology to dissociate automatic from nonautomatic processes involved in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1491–1500. http://doi.org/10.1037/a0032974.

Higham, P. A., Vokey, J. R., and Pritchard, J. L. (2000). Beyond dissociation logic: evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General*, *129*(4), 457–470. http://dx.doi.org/10.1037/0096-3445.129.4.457.

Hirst, W. and Kalmar, D. (1987). Characterizing attentional resources. *Journal of Experimental Psychology: General*, *116*(1), 68–81. http://doi.org/10.1037/0096-3445.116.1.68.

Jiménez, L. and Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 236–259. http://doi.org/10.1037/0278-7393.25.1.236.

Jiménez, L. and Vázquez, G. A. (2011). Implicit sequence learning and contextual cueing do not compete for central cognitive resources. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 222–235. http://doi.org/10.1037/a0020378.

Knowlton, B. J., Squire, L. R., Champagne, N., Kronenberg, B., Willoughby, K., and Zouzounis, J. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 169–181. http://doi.org/10.1037/0278-7393.23.1.220.

Kolańczyk, A. and Roczniewska, M. (2015). The affective self-regulation of covert and overt reasoning in a promotion vs. prevention mind-set. *Polish Psychological Bulletin*, *46*(2), 228–238. http://doi.org/10.1515/ppb-2015-0031.

Mealor, A. D., Dienes, Z., and Scott, R. B. (2014). Unconscious sources of familiarity can be strategically excluded in support of conscious task demands. *Psychology of Consciousness: Theory, Research, and Practice*, *1*(3), 229–242. http://doi.org/10.1037/cns0000027.

Meulemans, T. and Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 1007–1028. http://doi.org/10.1037/0278-7393.23.4.1007.

Moors, A. and De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. http://doi.org/10.1037/0033-2909.132.2.297.

Muraven, M., Tice, D. M., and Baumeister, R. F. (1998). Self-control as limited resource: regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*(3), 774–789. http://doi.org/10.1037/0022-3514.74.3.774.

Necka, E. (1997). Attention, working memory and arousal: concepts apt to account for the 'process of intelligence'. In G. Matthews (Ed.), *Cognitive science perspectives on personality* *and emotion* (Vol. 124, pp. 503–554). Amsterdam: Elsevier. http://doi.org/10.1016/S0166-4115(97)80129-X.

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

Norman, D. A. and Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7(1), 44–64. http://doi.org/10.1016/0010-0285(75)90004-3.

Perruchet, P. and Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*(3), 264–275. http://doi.org/10.1037/0096-3445.119.3.264.

Popławska, A., Roczniewska, M., and Sterczyński, R. (2014). The influence of resources limitation on decision-making processes in artificial grammar learning task. *Studia Psychologiczne*, *52*, 33–46. http://doi.org/10.2478/v10167-010-0076-3.

Popławska, A. and Wierzchoń, M. (2008). The role of perceptual and introspective consciousness in implicit knowledge acquisition: the influence of presentation time. *Psychological Studies*, *46*, 37–48.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*(2), 227–244. http://doi.org/10.1037/0033-2909.133.2.227.

Pothos, E. M. and Bailey, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(4), 847–862. http://doi.org/10.1037//0278-7393.26.4.847.

Rausei, V., Makovski, T., and Jiang, Y. V. (2007). Attention dependency in implicit learning of repeated search context. *Quarterly Journal of Experimental Psychology, 6*. 1321–1328. http://doi.org/10.1080/17470210701515744.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. http://doi.org/10.1016/S0022-5371(67)80149-X.

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

Roczniewska, M. and Kolańczyk, A. (2014). Competence over communion: implicit evaluations of personality traits during goal pursuit. *Polish Psychological Bulletin*, *45*(4), 418–425. http://doi.org/10.2478/ppb-2014-0046

Rowland, L. A. and Shanks, D. R. (2006). Attention modulates the learning of multiple contingencies. *Psychonomic Bulletin and Review*, *13*(4), 643–648. http://doi.org/10.3758/BF03193975.

Scott, R. B. and Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1264–1288. http://doi.org/10.1037/a0012943.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*(2), 163–196. http://doi.org/10.1037/0033-2909.115.2.163.

Shanks, D. R., Rowland, L. A., and Ranger, M. S. (2005). Attentional load and implicit sequence learning. *Psychological Research*, *69*(5–6), 369–382. http://doi.org/10.1007/s00426-004-0211-8.

Vokey, J. R. and Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 328–344. http://doi.org/10.1037//0278-7393.18.2.328.

Vokey, J. R. and Higham, P. A. (2005). Abstract analogies and positive transfer in artificial grammar learning. *Canadian Journal of Experimental Psychology*, *59*(1), 54–61. http://dx.doi.org/10.1037/h0087461.

Von Hecker, U., Dutke, S., and Sedek, G. (2000). *Generative Mental Processes and Cognitive Resources*. Dordrecht: Kluwer Academic Publishers.

Witt, A. and Vinter, A. (2012). Artificial grammar learning in children: abstraction of rules or sensitivity to perceptual features? *Psychological Research*, *76*(1), 97–110. http://doi.org/10.1007/s00426-011-0328-5.

Zizak, D. M. and Reber, A. S. (2004). Implicit preferences: the role(s) of familiarity in the structural mere exposure effect. *Consciousness and Cognition*, *13*(2), 336–362. http://doi.org/10.1016/j.concog.2003.12.003.

10

IMPLICIT LEARNING UNDER ATTENTIONAL LOAD

Michał Wierzchoń and Monika Derda

Introduction

The relation between attention and implicit learning has been intensely debated in the literature from both theoretical and empirical points of view (see e.g. Cleeremans, Destrebecqz and Boyer, 1998; Jiménez, 2003). Theoretically, it has been proposed that implicit learning is an automatic process (see e.g. Schumacher and Schwarb, 2009) from which implicit knowledge may be acquired without paying attention to stimuli and learning protocol (Wolf and Müller, 2012). This statement is important, as it is closely related to the very distinction between explicit and implicit information processing (Revonsuo and Rossetti, 2000). This is because theories postulating an independent character of implicit and explicit systems usually postulate a qualitative difference between these two modes of processing (Willingham and Goedert-Eschmann, 1999) and propose that lower vulnerability to attentional distraction is the main reason why the former was developed (Curran and Keele, 1993). Empirically, multiple studies have investigated whether implicit knowledge can be acquired under attentional load (see e.g. Wierzchoń, Gaillard, Asanowicz and Cleeremans, 2012 for a review). Interestingly, even though the problem has been investigated since the 1980s, a consensus has still not been reached. This may be partially a result of the multiple empirical paradigms applied in order to investigate the problem. Respectively, as one can see in the following paragraphs, discussions of results discrepancies observed under artificial grammar learning and serial reaction time tasks follow different directions. However, it seems that even taking into account these differences, one may find not only empirical evidence that speaks in favour of attentional demands of implicit learning (see e.g. Jiménez and Vázquez, 2005; Shanks, Rowland and Ranger, 2005), but also a large body of studies confirming the alternative hypothesis (e.g. Barker, 2012; Chang and Knowlton, 2004; Rowland and Shanks, 2006). In this chapter, we will discuss in detail the theoretical and empirical evidence supporting the hypothesis that implicit learning is independent from attentional resources. We will focus on the theoretical distinction between implicit and explicit information processing modes and the empirical evidence collected by studies investigating implicit learning under attentional load. We will also present a set of unpublished study results investigating the problem that will further support possible interpretations.

Implicit learning as an automatic process

Early papers defined implicit learning as a process through which intuitive knowledge about the underlying structure of a complex stimulus environment can be developed (Reber 1967; 1989). Typical features used to further characterise the process include (1) unconscious mode of processing, (2) abstract representation of knowledge acquired, (3) incidental character of learning and (4) its robustness over time, lack of attentional resources and psychological disorders resulting in deficits of conscious processing (Cleeremans, Destrebecqz and Boyer, 1998; Jiménez, 2003; Reber, 1989). All of these features have been discussed and sharply criticised in face of the results collected with implicit learning studies (see: Berry and Dienes, 1993; Shanks and Channon, 2002 for a review). This has not only empirical consequences, but also touches on the very problem of the independence of implicit and explicit processing. Many researchers assume that implicit learning is automatic and thus should require minimal cognitive resources (see e.g. Frensch and Rünger, 2003; Schumacher and Schwarb, 2009). This assumption is derived from the broader theoretical context suggesting that we can process information because of two separate systems: the implicit and the explicit. In fact, whenever a new study calls into question one of these features, the very assumption of the independence of implicit and explicit information systems is questioned. This debate adopts multiple forms and is often conducted in the context of perception, memory and consciousness domains (see e.g. Evans, 2008; Graf and Schacter, 1985; Reber and Squire, 1998; Squire, 1987; but see: Henke, 2010). Proponents of the so-called dual channel approach used to suggest that implicit information can influence behaviour independently of that processed explicitly, whereas an alternative view says that those two types of information processing closely depend on each other (see e.g. Frensch and Rünger, 2003; Maniscalco and Lau, 2016).

The idea that attentional requirements differ between the two modes of information processing described above has been developed in the context of automatic processing studies (see: Posner and Snyder, 1975; Shiffrin and Schneider, 1977 for the classical approach). Several critical features have been proposed to differentiate between automatic and controlled processes. Automatic processes are typically defined as being effortless, unconscious, autonomous, loosely controlled and involuntary (Hasher and Zacks, 1979; Logan, 1988). Nevertheless, it has often been proposed that the features do not always occur simultaneously (Neumann, 1984; Bargh, 1994). Leaving aside the problem of its involuntary and unconscious character, we will here focus on the attentional requirements of automatic processing. It is usually assumed that controlled processing consumes attentional resources and automatic processing does not (Strayer and Kramer, 1990). This assumption may be tested in a dual-task paradigm. Therefore, a task for which an automatic mode of processing is questioned (i.e. a primary task) is paired with another task that is assumed to be highly demanding in terms of engaged attentional resources (i.e. a secondary task). It is assumed that if a primary task is indeed automatic, the additional, secondary task should not impair participants' performance. However, when one observes a decrease in primary task performance, one may conclude that the task requires attentional resources and is cognitively demanding (Logan, 1978; Logan, 1979; Posner and Snyder, 1975). Following the dual-task paradigm logic in the context of implicit and explicit learning studies, proponents of implicit and explicit learning independence would expect that implicit learning should not be influenced by the attentional load resulting from performing the secondary task (see: Frensch and Rünger, 2003; Jiménez and Méndez, 1999; Wierzchoń, Gaillard, Asanowicz & Cleeremans, 2012), whereas researchers assuming that both implicit and explicit learning are possible due to the very same information processing system would expect that the attentional load should affect implicit learning effects (see: Shanks, Rowland and Ranger, 2005).

Implicit learning under attentional load

Before looking into the details of if and under which circumstances implicit learning was reported to be impaired under dual-task conditions, let us briefly introduce the two main paradigms through which implicit learning has been documented: artificial grammar learning, developed by Reber (1967), and sequence learning, first introduced by Nissen and Bullemer (1987).

As suggested by the implicit learning definition, the common factor in all implicit learning tasks is that participants acquire some information about the underlying structure of the material without having any intention of doing so (see Reber, 1989). Both paradigms share similar learning conditions, that is, participants are aware of stimuli but know nothing about the existence of hidden rules that organise the learned material. Thus, knowledge of the hidden rule is acquired incidentally. In both paradigms, participant behaviour indicates that they have acquired some knowledge of the rules (i.e. the acquired knowledge is exhibited by the indirect measure of knowledge). It is broadly assumed that implicit learning is observable within both paradigms (see e.g. Cleeremans, Destrebecqz and Boyer, 1998) and both are often used in the context of studies investigating cognitive and neurodegenerative disorders (see e.g. Rüsseler, Gerth and Münte, 2006; Smith, Siegert, McDowall and Abernethy, 2001). However, whether the nature of the knowledge acquired in both tasks is the same is still debatable (Perruchet, 2008; Perruchet, and Pacton, 2006).

In the canonical version of artificial grammar, a learning task consists of two phases. In the first phase, strings of letters constructed around a set of rules (e.g. finite-state grammar) are presented to participants. In the second phase, the participants are asked to classify the set of new strings (half grammatical and half ungrammatical, consecutively randomised) presented serially. Despite the fact that participants reported little verbal knowledge about the rules on which the strings were built, they were able to classify them above chance level (see e.g. Reber, 1989). Thus, it is often argued that they acquired knowledge implicitly and that it is represented unconsciously.

In a typical serial reaction time task experiment (Nissen and Bullemer, 1987), participants are asked to react to a stimulus presented at one of four possible locations. Each location corresponds to one of four possible response keys. The presented elements follow a structured sequence of events. At each trial, after seeing a stimulus, participants are asked to press a corresponding key as quickly and accurately as possible. Unknown to the participants, the sequence of stimuli follows a repetitive pattern. It is usually observed that reaction times tend to decrease progressively during practice (learning effect), but increase dramatically when the hidden sequential structure is modified in any of several ways (the so-called transfer effect: see e.g. Destrebecqz and Cleeremans, 2001; Wierzchoń, Gaillard, Asanowicz & Cleeremans, 2012). This pattern of results is usually interpreted as proof that participants have become sensitive to the sequential regularities contained in the material during the course of training.

Assuming that artificial grammar learning and sequence learning measure the very same process, one would expect similar effects of attentional load induced by a concurrent secondary task regardless of the specific implicit learning paradigm applied. It is important to note here that, regardless of an implicit learning procedure applied, there are at least three possible mechanisms that may diminish implicit learning effects under the dual-task paradigm. Each is related to the influence of the secondary task at a different stage of a learning task. Firstly, one may argue that under attentional load the process of encoding regularities is disrupted in such a way that weaker knowledge representation is formed. Secondly, attentional load may influence the judgement knowledge that is needed to perform a task when the acquired implicit knowledge is tested (see e.g. Dienes and Scott, 2005). Finally, the attentional load may affect motor reactions and thus result in lower performance of an implicit learning test with no effect on implicit learning itself (see also Wierzchoń, Gaillard, Asanowicz & Cleeremans, 2012).

Artificial grammar learning under attentional load

Most artificial grammar learning studies have shown that attentional load induced by a secondary task does not influence implicit learning (as tested with the classification phase of the artificial grammar: Broadbent, 1989; Dienes and Scott, 2005; Hayes, 1989. See Table 10.1 for the review). It was even proposed that attentional load could facilitate the acquisition of implicit knowledge (Perruchet, 2008), as participants are not focusing on rule identification. However, identification is usually ineffective because of the rule difficulty (see also Reber, 1989). Other studies have challenged these conclusions by demonstrating impaired classification performance under attentional load (see e.g. Dienes, Broadbent and Berry, 1991; Chang and Knowlton, 2004). It was usually proposed that this result was an effect of explicit learning's large contribution to classification task performance. This was independently confirmed with multiple studies investigating the availability

TABLE 10.1 Overview of research papers investigating effects of attentional load on implicit learning in an artificial grammar learning paradigm. RNG – random number generation task; EL – explicit learning; IL – implicit learning; AGL – in an artificial grammar learning (AGL) paradigm.

Author	*Secondary task*	*IL under dual-task*	*Results*
Hayes, 1989	**RNG**	observed	intact classification under standard incidental memory instruction (IL measure), but impaired under intentional learning instructions (EL measure)
Dienes, Broadbent and Berry, 1991, Exp. 2	**RNG**	impaired	impaired classification, and other measures of IL (d' and sequential letter dependencies tests) both under intentional and incidental instructions
Chang and Knowlton, 2004, Exp. 2	**articulatory suppression**	observed	knowledge about abstract rules can be acquired, but articulatory suppression reduces later sensitivity to chunk strength
Dienes and Scott, 2005	**RNG**	observed	no effects on classification performance and measures of the conscious or unconscious status of judgement knowledge (i.e. guessing criterion and Chun difference); decreased proportion of attributions to conscious structural knowledge (EL)
Hendricks, Conway and Kellogg, 2013	**Digit span task**	observed/ impaired	Exp 1: three dual-task conditions: DA (secondary task during acquisition phase), DT (testing phase), DAT (both acquisition and testing phase); grammaticality impaired in DT, observed in DT and DAT condition, chunk strength judgements observed in all conditions
Ziori, Pothos and Dienes, 2014	**RNG**	observed/ impaired	AGL in natural context (cities on map); performance under dual-task condition depends on knowledge types: grammaticality observed under dual-task conditions, chunk strength similarity not found in the dual task

of artificial grammar learning judgements in awareness (see e.g. Dienes and Seth, 2010; Wierzchoń, Asanowicz, Paulewicz and Cleeremans, 2012). It seems worth noting that intact classification under dual-task conditions does not imply that implicit learning is entirely independent of attentional resources as the secondary task may not deplete attentional resources enough (i.e. may be too easy to exhibit effects of attentional load on implicit learning). We will come back to this possibility in the following paragraphs.

Serial reaction time task under attentional load

Analogously to artificial grammar learning, serial reaction time task studies exhibit huge differences between the results of experiments investigating the problem. Many studies report successive cases of implicit sequence learning under dual-task conditions (Cohen, Ivry, and Keele, 1990; Frensch, Buchner, and Lin, 1994; Reed and Johnson, 1994; Shanks and Johnstone, 1998; see Table 10.2 for a review). However, a large body of studies has shown that attentional load results in reduced learning effect, longer reaction times and a less pronounced transfer effect (Jiménez and Vázquez, 2005; Shanks and Channon, 2002; Shanks et al., 2005). When trying to explain the discrepancies of the observed results, it seems worth noting that serial reaction time task procedures differ between experiments. One incongruity that influences the effect of a secondary task is the structural complexity of a sequence acquired through the procedure. For instance, it was shown that sequences with unique associations are learned under attentional load, whereas learning ambiguous sequences requires attention (see: Cohen et al., 1990). Similarly, deterministic sequence learning seems to be impaired more by divided attention than by probabilistic learning (Jiménez and Vázquez, 2005). Finally, the interference resulting from attentional load decreases with serial reaction time task training (i.e., when a serial reaction time task is automatised; see Cohen and Poldrack, 2008).

Another factor that seems to mediate the observed effects of attentional load is secondary task and serial reaction time task integration. Studies have demonstrated that the degree of overlap between the processes involved in performing the secondary task and the serial reaction time task (which was manipulated by means of task priority and stimulus onset asynchrony) modulates the way in which dual-tasking interferes with learning (Schumacher and Schwarb, 2009). It was observed that strong integration reduces dual-task interference (Rah, Reber, and Hsiao, 2000; Schmidtke and Heuer, 1997). Also, the temporal characteristics of the serial reaction time task affect the observed effects of cognitive load by disorganising serial reaction time task consistency (e.g., by prolonging the stimulus onset asynchrony and thus disturbing the temporal organisation of the sequence; see: Stadler, 1995).

Regardless of the type of implicit learning procedure applied, different studies applied different secondary tasks (see Tables 10.1 and 10.2, column B) and, as one may expect, this seems to strongly influence observed results. Assuming that implicit learning is attentionally demanding, one may expect that when comparing

TABLE 10.2 Overview of research papers investigating effects of attentional load on implicit learning with a serial reaction time task. RNG – random number generation task; TC – tone-counting task; VSC – visual stimuli counting; MA – mental arithmetic; EL– explicit learning; SL – sequence learning; SOC sequences – Second-Order Conditional sequences; SRT – with a serial reaction time task; ASRT – alternating SRT; WM – working memory.

Nissen and Bullemer, 1987	**TC**	impaired	acquisition of the sequence under TC was minimal
Cohen, Ivry and Keele, 1990	**TC**	observed/impaired	simple structured sequences can be learned with TC but more complex ones require attention
Curran and Keele, 1993	**TC**	observed	no influence of TC regardless of the level of sequence awareness
Reed and Johnson, 1994	**TC**	observed	SOC sequences can be learned under TC (no control conditions; TC used to minimise opportunities for explicit learning)
Frensch, Buchner and Lin, 1994	**TC**	observed	both unique and ambiguous sequences can be learned under TC; both time of secondary task onset and time interval between the response to a stimulus and the presentation of the next stimulus affect SRT performance
Stadler, 1995	**TC, memory-load task**	observed/impaired	no influence of memory load, but impaired SL under TC; TC disrupts learning by preventing consistent organisation of the sequence
Heuer and Schmidtke, 1996	**verbal, visuo-spatial and auditory go/no go tasks**	observed/impaired	only auditory go/no go task (a task similar to TC, but with no requirement to update and memorise the number counted) interferes with the SL; interference seems to be specific to certain secondary tasks
Mayr, 1996	**TC, learning of a second sequence**	observed	learning of spatial and object sequences simultaneously was as efficient as learning of single sequences; the effect occurs even under TC
Schmidtke and Heuer, 1997	**auditory go/no go task**	observed/impaired	impaired performance under dual-task conditions can be caused by a task integration that impairs SRT (reduced SRT under go/no go task with a random sequences of tones; repeated sequences of tones integrated with SRT enhanced learning)

Frensch, Lin and Buchner, 1998	**TC**	impaired	TC primarily affects expression of learning (practice effects in SRT did not differ under TC), but also implicit learning itself (when learning assessment was performed under TC)
Shanks and Johnstone, 1998	**TC**	observed	replication of Reed and Johnson, 1994, SOC sequences can be learned under TC
Schvaneveldt and Gomez, 1998	**TC**	observed	probabilistic sequences (first- and second-order conditional) are learned under TC; transfer effect results under TC suggest limitations in performance, but not in learning
Jiménez and Méndez, 1999	**VSC**	observed	no effect of VSC (target shape-counting performed on stimulus on which SRT was being carried out) on SL of probabilistic sequences generated with finite-state grammar
Rah, Reber and Hsiao, 2000	**TC**	observed/impaired	contingency of tone sequence in TC and SRT influence learning; SOC sequences can be learned under TC contingent with SRT (reverse results in Exp. 4, when TC was not contingent); attention was not manipulated across conditions
Jiménez and Méndez, 2001	**VSC**	observed	replication of Jiménez and Méndez, 1999; SL can be acquired and expressed under VSC even when Ss cannot anticipate the next location in cued generation task (EL test)
Hsiao and Reber, 2001	**TC**	impaired	significant learning of the SOC sequence; the effect was influenced by RSOA of tones and level of TC performance
Shanks and Channon, 2002	**TC**	impaired	SL of SOC affected by TC during training, regardless of the presence of TC in the transfer block
Jiménez and Vázquez, 2005	**TC, TC associated with SRT**	observed/impaired	TC affected expression and acquisition of SL; greater interference was observed with deterministic sequence (EL); no influence of TC on SL, when a task is associated with SRT
Shanks, Rowland and Ranger, 2005	**VSC**	impaired	VSC impairs SL of SOC (regardless the presence of secondary task at transfer); acquired knowledge, as assessed by a generation task, was consciously accessible

(continued)

Table 10.2 *(continued)*

Poldrack et al., 2005	**TC**	observed	fMRI study; behavioural data: no effects of TC after intensive training; fMRI data: before training, SRT with TC elicited activation in a wide network of frontal and striatal regions, as well as parietal lobe; after training – SRT under TC showed less activity in bilateral ventral premotor regions, right middle frontal gyrus, and right caudate body
Rowland and Shanks, 2006	**Ignoring irrelevant distractor**	observed	no evidence of impaired learning in groups trained on the SRT task in the presence of distractor stimuli, SL observed because selection task was easy and not resource consuming
Nejati, Farshi, Ashayeri and Aghdasi, 2008	**TC**	observed/impaired	SL under TC observed in younger adults but impaired in elderly group
Cohen and Poldrack, 2008	**letter counting task**	impaired	letter counting task impaired SRT but dual-task effect decreased with training (SRT lasted 3h)
Schumacher and Schwarb, 2009	**tone-identi-fication task**	observed/impaired	dual task disrupts SRT only when the processing for the two tasks overlap (i.e. parallel response selection for both tasks interfere – short SOA) and with equal priority of tasks (as compared to SRT priority)
Hemond, Brown and Robertson, 2010	**VSC, learning of a second sequence**	observed/impaired	SL can be enhanced by concurrently learning sequence of coloured cues and impaired by VSC (counting the number of red cues)
Németh et al., 2011	**Sentence comprehension; word recognition; MA**	observed/impaired	ASRT task, general skill learning was not affected by the dual-task conditions. Sequence-specific learning observed in the Word and MA conditions, but impaired in the Sentence condition. Secondary task effects results from language-specific processing but not task difficulty
Barker, 2012	**Six types of dual tasks (related to spatial/temporal aspects of WM)**	observed	incidental learning observed in all conditions, but awareness scores (explicit measures) were diminished

Gabay, Schiff and Vakil, 2012	**TC**	impaired	Ss with developmental dyslexia (DD) and control group. In dual-task no transfer effect in both groups, but a secondary task seems to delay the acquisition of a new skill in the DD group only
Wierzchoń, Gaillard, Asanowicz & Paulewicz, 2012	**RNG and TC**	observed/impaired	SRT task, learning observed under dual conditions, but transfer effect was smaller under dual-task conditions; SRT was impaired more under RNG than TC
Halvorson, Wagschal and Hazeltine, 2013	**Tone identification task**	observed/impaired	Additional instruction manipulation: Ss were instructed to complete two separate tasks (visual and auditory) or a single, integrated task. In the separated tasks group SL was observed, while in the integrated group SL was impaired
Coomans, Vandenbossche and Deroost, 2014	**Dot changes counting task**	observed/impaired	Comparison between 8–10-year-old children and 18–22-year-old adults. Young adults exhibit more pronounced SL than children. The secondary task impaired task performance in SL in adults, but not in children
Thompson, Sanchez, Wesley and Reber, 2014	**Ego depletion**	impaired	Serial Interception Sequence Learning (SISL); ego depletion procedure (applied prior to or during learning); impaired expression of sequence-specific knowledge
Vandenbossche, Coomans, Homblé and Deroost, 2014	**TC and VSC**	observed/impaired	Comparison between young (18–25) and older adults (55–75). SL was observed under dual-task conditions for younger group, but reduced for older adults
Borragán, Slama, Destrebecqz and Peigneux, 2016	**Time load dual back task**	observed	SRT, two-day experiment, facilitative effect of cognitive fatigue induced by secondary task; performance improvement was more pronounced for sequential than the motor components of SL

different secondary tasks we should observe different effects of attentional load on implicit learning (see: Heuer and Schmidtke, 1996; Stadler, 1995). Surprisingly, this question has rarely been directly addressed experimentally (but see: Stadler, 1995; Wierzchoń, Gaillard, Asanowicz and Cleeremans, 2012). It is well known that different secondary tasks have different attentional requirements. For example, a tone-counting task is not thought to be demanding, while the opposite is true of random number generation. Interestingly, serial reaction time task studies investigating the effects of attentional load on implicit learning have most often applied a tone-counting task, whereas artificial grammar learning studies have used random number generation more often (none of the artificial grammar studies applied tone-counting).

Summing up, the results of studies investigating the effects of attentional load on implicit learning are not consistent. There are multiple reasons for this which are related to both the details of the implicit learning procedures and the types of secondary task applied. However, two observations seem to be crucial in the context of the theoretical value of those results. First, it seems that a secondary task needs to be properly chosen, as a secondary task that is too easy might be performed simultaneously, even with an attentionally demanding primary task. Thus, it may be that implicit learning is indeed not automatic, but attentional load manipulation was too weak to prove this. Secondly, one should check whether experiments showing no effects of attentional load on implicit learning indeed speak in favour of the null hypothesis, or whether statistical evidence is unclear.

New empirical evidence

Here, we briefly present a set of results of experiments aiming to test the effect of attentional load on implicit learning. Some of the main disadvantages of the studies described in the previous sections are related to the fact that we are testing the null hypothesis when we assume no effect of attentional load on implicit learning. However, to our knowledge all of the previous experiments either provide evidence for the alternative hypothesis (assuming the difference between attentional load and control conditions), or do not provide evidence in favour of any hypothesis. In other words, the null hypothesis was never directly tested. Thus, a method should be applied that allows us to test whether indeed there are no differences between conditions. Recently, the use of Bayesian statistics was proposed in order to investigate unconscious mental states (Dienes, 2015). In the same vein, we reanalysed a set of a simple experiments conducted in recent years in our laboratory, analysing whether they support the hypothesis that assumes that attentional load does not affect implicit learning. In other words, we tested whether our data are more likely assuming no difference between control and dual-task conditions. This can be done with the BF_{01} analysis that allows us to test whether the data are more probable under the null hypothesis. The result of BF_{01} larger than 3 is interpreted as moderate evidence in favour of the null hypothesis (see Dienes, 2015 for more details).

Artificial grammar learning studies

The procedures of all experiments in this group follow the same scenario. We applied the typical procedure of an artificial grammar learning task. We have used a simple grammar and the set of materials described in early papers on implicit learning (Dienes, Broadbent, and Berry, 1991). First, we asked participants to memorise a string of letters based on artificial grammar (acquisition phase). Seven grammatical strings were presented twice. Each string was presented for 5s. Then, we checked whether participants followed the instruction asking them to recall the presented strings. Finally, we informed participants that the strings they had seen were constructed based on a set of rules and presented them with a set of new strings, some of them following the very same rule and some violating it. Participants then classified the strings as grammatical and ungrammatical. Forty-two strings were presented in this phase, half of them being grammatical. Crucially, in the experimental groups we additionally asked participants to perform a secondary task simultaneously with the acquisition phase. These manipulations aimed to influence attentional load, thus allowing us to test whether implicit learning, as measured with classification task accuracy, requires attention. Below, we report the results of two experiments based on this scenario, presenting only the results investigating the impact of attentional load on classification performance.

Experiment 1

66 (22 participants in each group) undergraduate students voluntarily participated in the study. The participants were randomly assigned to three groups: control (C), DIVA-simple (DS), DIVA-hard (DH). In the control group the artificial grammar learning task was used. In addition to the AGL task, in DS and DH groups the secondary divided attention task was performed (DIVA, Szymura and Nęcka, 1998). The participants in the simple DIVA condition (DS) had to hold the position of a horizontal line in the middle of a vertically oriented rectangle by pressing a mouse button. The line was falling continuously and each button press allowed participants to lift the line. When the line crossed either the upper or lower border of the rectangle, participants heard an unpleasant tone. They were instructed to avoid such situations. In the more difficult version of the task (DH) the position of the line changes randomly.

The results indicate that classification accuracy was above the chance level in all groups (C: 59%, DS: 63%, DH: 65%; $t > 2.8$; $p < .01$). However, the secondary task did not affect classification performance regardless of the task difficulty condition (ANOVA with three-level factor group and dependent variable: accuracy; $F[2,63] = 1.04$, ns). In order to check whether attentional load does indeed not influence implicit learning, we additionally computed a Bayesian ANOVA using JASP®, the free open-source statistics package (JASP Team, 2016). The data was more probable under the null hypothesis, as $BF_{01} = 3.64$.

Experiment 2

76 (25 in control and 51 in experimental group) undergraduate students voluntarily participated in the experiment. We randomly assigned them to two experimental groups and a control group. All participants followed the classical artificial grammar learning task procedure. The random interval generation task (Vandierendonck, De Vooght and Van der Goten, 1998) was used as a secondary task in experimental condition. Participants had to click on the left mouse button as randomly as possible during the acquisition phase.

The classification performance was above chance level in all groups (control: 57%, experimental: 61% $t > 2.20$, $p < .05$). We did not observe any effects of attentional load on classification performance (ANOVA with two-level factor group and dependent variable: accuracy: $F[2,73] = 1.24$, ns). Similar to the previous experiment, we checked whether the attentional load did not influence implicit learning using a Bayesian ANOVA. Again, the data was more probable under the null hypothesis, as $BF_{01} = 3.41$.

Serial reaction time task studies

In the following two experiments, we applied the typical serial reaction time task procedure that follows exactly the same procedure as we have used in our previous studies (Wierzchoń, Gaillard, Asanowicz and Cleeremans, 2012). We presented participants with a series of stimuli appearing sequentially at one of four possible locations. Participants had to react as quickly and as accurately as possible by pressing a key corresponding to the location of a stimulus. In the first 13 blocks (96 trials each), the order of stimuli presentation follows second-order contingency sequences (SOC1 or SOC2 sequences – see: Reed and Johnson, 1994). At the 14th block, the sequence was replaced with another second-order contingency sequence, and then at the 15th block the initial sequence was again introduced. Blocks 1–12 served as a learning phase, whereas the last three blocks were used as a sequence learning test phase. Here, we calculate the difference between reaction times in the 14th block, where the sequence was changed, and the averaged reaction times for the 13th and 15th blocks (the so-called transfer effect index). Analogously to the artificial grammar learning experiments described above, in the experimental groups we additionally asked participants to perform a secondary task simultaneously with the first 12 blocks of the serial reaction time task, aiming to test the effects of attentional load on the sequence learning process. We then tested whether implicit learning, as tested with the transfer effect, requires attentional resources. Below, we report the results of two experiments based on this scenario, presenting only the results investigating the impact of attentional load on the transfer effect.

Experiment 3

55 undergraduate students voluntarily participated in the experiment and were randomly assigned to two experimental and a control group. 19 participants were

assigned to the control group (C); 18 participants were tested within each of the experimental groups. All participants followed the classical serial reaction time task procedure as described above. A mental arithmetic task was used as a secondary task in experimental conditions so that in 25% of trials in learning blocks a random digit replaced a stimulus in the serial reaction time task. Participants had to add 'one' (M1 group) or 'three' (M3 group) (depending on the task condition) to a presented digit and articulate a result. It was assumed that adding 'three' should be a more demanding task than adding 'one'.

The efficacy of training was assessed with the transfer effect that was observed in all experimental conditions (C: 73.23ms; M1: 43.75ms; M3: 42.11ms; $Fs > 27.72$, $p < .001$). We also observed an effect of attentional load on the transfer effect (ANOVA with three-level factor group and dependent variable: transfer effect, $F[2.52] = 4.322$, $p < .05$). However, we did not observe any differences between dual-task conditions (i.e. between easy and difficult mental arithmetic conditions: $t = .14$, ns). We further compared those conditions with a Bayesian independent samples t test. The data was more probable under the null hypothesis, as $BF_{01} = 3.08$. Thus, even though the serial reaction time task performance was impaired under attentional load conditions, the difficulty of the secondary task did not further affect the results. Importantly, a significant transfer effect was observed in all experimental conditions, suggesting that participants had acquired sequence knowledge even under dual-task conditions.

Experiment 4

69 undergraduate students voluntarily participated in the experiment and were randomly assigned to two experimental groups and a control group. 19 participants were assigned to the control group (C); 23 participants were tested in the first experimental group (RNG1), and 27 in the second experimental group (RNG20). All participants followed the classic serial reaction time task procedure as described above. A random number generation task was used as a secondary task in experimental conditions so that participants were asked to articulate random digits between '1' and '9' (RNG1) or '2' and '29' (RNG20), depending on the condition. It was assumed that both conditions should be equally demanding for participants, but the latter would (adversely) affect the temporal structure of the serial reaction time task (as the time required to articulate a random number was much longer). Similarly to Experiment 3, the efficacy of training was assessed with the transfer effect that was observed in all experimental conditions (C: 73.22ms; RNG1: 35.03ms, RNG20: 32.69ms; $Fs > 40.39$, $p < .001$). We observed an effect of attentional load on the transfer effect (ANOVA with three-level factor group and dependent variable: transfer effect, $F[2.66] = 11.29$, $p < .001$). However, we again did not observe any differences between dual-task conditions ($t = .27$, ns). We further compared these conditions with a Bayesian independent samples t test. The data was more probable under the null hypothesis, as $BF_{01} = 3.37$. To sum up, we observed an impaired transfer effect under

the dual-task condition. Interestingly, the type of the secondary task applied did not further affect the influence of the secondary task on sequence learning efficiency. Again, it is worth noting that a significant transfer effect was observed in all experimental conditions, thus participants had acquired sequence knowledge even under dual-task conditions.

Final conclusions

The question of whether attentional load influences implicit learning is crucially related to the problem of the relation between implicit and explicit systems (Revonsuo and Rossetti, 2000), Thus, it seems worth investigating evidence that speaks in favour of the hypothesis that assumes that attentional load does not affect implicit learning. Such a pattern of results confirms the automatic character of implicit learning that may be interpreted in line with models assuming at least relative independence of the implicit and explicit learning systems (Frensch and Rünger, 2003; Jiménez, and Méndez, 1999; Wierzchoń, Gaillard, Asanowicz and Cleeremans, 2012, Willingham and Goedert-Eschmann, 1999). The literature review presented in this chapter clearly shows that the current results on this subject are not consistent. A large body of experiments suggests that attentional load does not affect implicit learning; however, multiple studies suggest the opposite (see Tables 10.1 and 10.2). We hypothesised that some of these inconsistencies may result from the differences between the implicit learning task applied and attentional load induction methods.

We proposed a series of studies using various types of the secondary tasks in the context of two established implicit learning procedures. Particularly, we presented two studies conducted in an artificial grammar learning paradigm showing that a secondary task applied over an acquisition phase did not affect classification accuracy (regardless of the actual type of secondary task applied). We confirmed this conclusion with Bayesian analysis, thereby allowing us to present evidence in favour of the null hypothesis. The results observed in the context of the serial reaction time task were less clear. We observed an impaired transfer effect under secondary task conditions. However, the further manipulation of the secondary task difficulty did not affect learning efficacy. Importantly, sequence learning was still observed under attentional load (as we observed a significant transfer effect in all the experimental conditions). This suggests that the less pronounced transfer effect could result from the fact that participants had to perform an additional task during the training phase, i.e. attentional load may affect motor reactions or other processes related to the task execution and thus result in lower performance of an implicit learning test with no effect on implicit learning itself (see also Wierzchoń, Gaillard, Asanowicz and Cleeremans, 2012). Serial reaction time task performance may be also impaired due to the changed time characteristics of the task (e.g. it simply takes longer to pronounce 23 than 3, as required by the RNG1 and RNG20 protocols described in Experiment 4 – see: Stadler, 1995 for a similar view).

In the introduction we questioned at which stage of learning the secondary task may disturb implicit learning. Given the fact that we have observed learning effects in all the experiments (even though they were slightly diminished in the SRT task), the most plausible explanation for now is that attentional load results in lower performance of an implicit learning test with no effect on implicit learning itself. Depending on the task, this could be an effect of motor (as in DIVA) or other types of interference related to the tasks execution (as in both SL studies). It is important to note, that the secondary tasks we have applied aimed to disturb different aspects of attention, thus they should not be directly compared. However, the congruency of the effects seems to suggest that implicit learning is observed regardless of the detailed nature of the attentional load task.

The effects of attentional load on sequence learning should be further investigated. The current empirical evidence is still not clear. For example, one may argue that the difficulty of the secondary task was not properly controlled (we have assumed the differences in the difficulty based on the literature and controlled the accuracy of the performance over the course of the experiments, but we have not analysed differences in difficulty within a given task statistically). We added the difficulty manipulation to rule out the possibility that attentional load manipulation was too weak. However, one may try to develop even more difficult tasks and see whether it will affect the performance.

We believe that our data adds to evidence supporting the thesis that attentional load does not affect implicit learning. It also seems worth applying the idea of comparing not only results collected in studies applying different types of secondary tasks, but also manipulating condition difficulty within one type of secondary task to vary the assumed attentional demands. However, in future studies this manipulation should be more carefully controlled. Importantly, our results confirm that data of implicit learning studies should be analysed with Bayesian models in order to investigate the evidence supporting not only the alternative, but also the null hypothesis.

Acknowledgements

Preparation of this paper was supported with a SONATA BIS grant, funded by the National Science Centre, Poland (Grant 2012/07/E/HS6/01037).

References

Bargh, J. A. (1994). The four horsemen of automaticity: intention, awareness, efficiency, and control as separate issues. In R. S. Wyer, and T. K. Srull (Eds.), *Handbook of Social Cognition: Vol. 1, Basic Processes* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Barker, L. (2012). Defining the parameters of incidental learning on a serial reaction time (SRT) task: do conscious rules apply? *Brain Sciences*, *2*(4), 769–789. DOI: 10.3390/brainsci2040769.

Berry, D. C. and Dienes, Z. (1993). *Implicit Learning: Theoretical and Empirical Issues*. Hove, England: Erlbaum.

Borragán, G., Slama, H., Destrebecqz, A. and Peigneux, P. (2016). Cognitive fatigue facilitates procedural sequence learning. *Frontiers in Human Neuroscience, 10*. DOI: 10.3389/fnhum.2016.00086.

Broadbent, D. E. (1989). Lasting representations and temporary processes. In H. L. Roediger III and F. I. M. Craik (Eds.), *Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving* (pp. 211–227). Hillsdale, NJ: Erlbaum.

Chang, G. Y. and Knowlton, B. J. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 714–722. DOI: 10.1037/0278-7393.30.3.714.

Cleeremans, A., Destrebecqz, A. and Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416. DOI: 10.1016/S1364- 6613(98) 01232-7.

Cohen, A., Ivry, R. I. and Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 17–30. DOI: 10.1037/0278-7393.16.1.17.

Cohen, J. R. and Poldrack, R. A. (2008). Automaticity in motor sequence learning does not impair response inhibition. *Psychonomic Bulletin and Review*, *15*(1), 108–115. DOI: 10.3758/PBR.15.1.108.

Coomans, D., Vandenbossche, J. and Deroost, N. (2014). The effect of attentional load on implicit sequence learning in children and young adults. *Frontiers in Psychology, 5*. DOI: 10.3389/fpsyg.2014.00465.

Curran, T. and Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 189–202. DOI: 10.1037/0278-7393.19.1.189.

Destrebecqz, A. and Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, *8*, 343–350. DOI: 10.3758/BF03196171.

Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In M. Overgaard (Ed.), *Behavioural Methods in Consciousness Research* (pp.199–220). Oxford: Oxford University Press.

Dienes, Z., Broadbent, D. and Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 875–887. DOI:10.1037/0278-7393.17.5.875.

Dienes, Z. and Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*(5–6), 338–351. DOI: 10.1007/s00426-004-0208-3.

Dienes, Z. and Seth, A. (2010). Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Consciousness and Cognition*, *19*(2), 674–681.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. A*nnual Review of Psychology*, *59*, 255–278. DOI:10.1146/annurev.psych.59.1030 06.093629.

Frensch, P. A., Buchner, A. and 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*(3), 567–584. DOI: 10.1037/0278-7393.20.3.567.

Frensch, P. A., Lin, J. and 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*(2), 83–98. DOI: 10.1007/s004260050015.

Frensch, P. A. and Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, *12*, 13–18. DOI: 10.1111/1467-8721.01213.

Gabay, Y., Schiff, R. and Vakil, E. (2012). Attentional requirements during acquisition and consolidation of a skill in normal readers and developmental dyslexics. *Neuropsychology*, *26*(6), 744–757. DOI: 10.1037/a0030235.

Graf, P. and Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 501–518. DOI: 10.1037/0278-7393.11.3.501.

Halvorson, K. M., Wagschal, T. T. and Hazeltine, E. (2013). Conceptualization of task boundaries preserves implicit sequence learning under dual-task conditions. *Psychonomic Bulletin and Review*, *20*(5), 1005–1010. DOI: 10.3758/s13423-013-0409-0.

Hasher, L. and Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *08*(3), 356–388. DOI: 10.1037/0096- 3445.108.*3*.356.

Hayes, N. A. (1989). *Systems of Explicit and Implicit Learning*. Unpublished doctoral dissertation, University of Oxford, England.

Hendricks, M. A., Conway, C. M. and Kellogg, R. T. (2013). Using dual-task methodology to dissociate automatic from nonautomatic processes involved in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1491 1500. DOI: 10.1037/a0032974.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*(7), 523–532. DOI: 10.1038/nrn2850.

Heuer, H. and Schmidtke, V. (1996). Secondary-task effects on sequence learning. *Psychological Research*, *59*(2), 119–133. DOI: 10.1007/BF01792433.

Hsiao, A. T. and Reber, A. S. (2001). The dual-task SRT procedure: fine-tuning the timing. *Psychonomic Bulletin and Review*, *8*(2), 336–342. DOI: 10.3758/BF03196170.

Hemond, C., Brown, R. M. and Robertson, E. M. (2010). A distraction can impair or enhance motor performance. *The Journal of Neuroscience*, *30*(2), 650–654.

JASP Team (2016). JASP (Version 0.7.5.5) [Computer software].

Jiménez, L. (Ed.). (2003). *Advances in Consciousness Research, Vol. 48: Attention and Implicit Learning*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Jiménez, L. and Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 236–259. DOI: 10.1037/0278-7393.25.1.236.

Jiménez, L. and Méndez, C. (2001). Implicit sequence learning with competing explicit cues. *The Quarterly Journal of Experimental Psychology: Section A*, *54*(2), 345–369. DOI: 10.1080/713755964.

Jiménez, L. and Vázquez, G. A. (2005). Sequence learning under dual-task conditions: alternatives to a resource-based account. *Psychological Research*, *69*(5–6), 352–368. DOI: 10.1007/s00426-004-0210-9.

Logan, G. D. (1978). Attention in character-classification tasks: evidence for the automaticity of component stages. *Journal of Experimental Psychology: General*, *107*(1), 32–63. DOI: 10.1037/0096-3445.107.1.32.

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(2), 189–207. DOI: 10.1037/0096-1523.5.2.189.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. DOI: 10.1037/0033-295X.95.4.492.

Maniscalco, B. and Lau, H. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neuroscience of Consciousness*, 1–17. DOI: 10.1093/nc/niw00.

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*(2), 350–364. DOI: 10.1037/0278-7393.22.2.350.

Nejati, V., Farshi, M. T., Ashayeri, H. and Aghdasi, M. T. (2008). Dual task interference in implicit sequence learning by young and old adults. *International Journal of Geriatric Psychiatry*, *23*(8), 801–804. DOI: 10.1002/gps.1976.

Németh, D., Janacsek, K., Csifcsak, G., Szvoboda, G., Howard, J. H. and Howard, D. V. (2011). Interference between sentence processing and probabilistic implicit sequence learning. *PLoS ONE*, *6*(3), e17577. DOI: 10.1371/journal.pone.0017577.

Neumann, O. (1984). Automatic processing: a review of recent findings and a plea for an old theory. In *Cognition and Motor Processes* (pp. 255–293). Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-69382-3_17.

Nissen, M. J. and Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. DOI: 10.1016/0010-0285(87)90002-8.

Perruchet, P. (2008). Implicit learning. In J. H. Byrne and H. L. Roediger III (Eds.), *Learning and Memory: a Comprehensive Reference: Vol. 2. Cognitive Psychology of Memory* (pp. 597–621). Oxford, England: Elsevier.

Perruchet, P. and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in cognitive sciences*, *10*(5), 233–238. DOI: 10.1016/j.tics.2006.03.006

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y. and Knowlton, B. J. (2005). The neural correlates of motor skill automaticity. *The Journal of Neuroscience*, *25*(22), 5356–5364.

Posner, M. I. and Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Eds.), *Information Processing and Cognition: the Loyola Symposium* (pp. 55–85), Hillsdale, NJ: Erlbaum.

Rah, S. K. Y., Reber, A. S. and Hsiao, A. T. (2000). Another wrinkle on the dual-task SRT experiment: it's probably not dual task. *Psychonomic Bulletin and Review*, 7(2), 309–313. DOI: 10.3758/BF03212986.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. DOI: 10.1016/S0022-5371(67)80149-X.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. DOI: 10.1037/0096-3445.118.3.219.

Reber, P. J. and Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, *10*(2), 248–263. DOI: 10.1162/089892998562681.

Reed, J. and 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*(3), 585–594. DOI: 10.1037/0278-7393.20.3.585.

Revonsuo, A. and Rossetti, Y. (2000). Dissociation and interaction: windows to the hidden mechanisms of consciousness. In Y. Rossetti and A. Revonsuo (Eds.), *Beyond Dissociation: Interaction between Dissociated Implicit and Explicit Processing* (pp. 351–366). Amsterdam: Benjamins.

Rowland, L. A. and Shanks, D. R. (2006). Sequence learning and selection difficulty. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 287–299. DOI: 10.1037/0096-1523.32.2.287.

Rüsseler, J., Gerth, I. and Münte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology*, *28*(5), 808–827. DOI: 10.1080/13803390591001007.

Schmidtke, V. and Heuer, H. (1997). Task integration as a factor in secondary-task effects on sequence learning. *Psychological Research*, *60*(1–2), 53–71. DOI: 10.1007/BF00419680.

Schumacher, E. H. and Schwarb, H. (2009). Parallel response selection disrupts sequence learning under dual-task conditions. *Journal of Experimental Psychology: General*, *138*(2), 270–290. DOI: 10.1037/a0015378.

Schvaneveldt, R. W. and Gomez, R. L. (1998). Attention and probabilistic sequence learning. *Psychological Research*, *61*(3), 175–190. DOI: 10.1007/s004260050023.

Shanks, D. R., and Channon, S. (2002). Effects of a secondary task on 'implicit' sequence learning: learning or performance? *Psychological Research*, *66*(2), 99–109. DOI: 10.1007/s00426-001-0081-2.

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

Shanks, D. R., Rowland, L. A. and Ranger, M. S. (2005). Attentional load and implicit sequence learning. *Psychological Research*, *69*(5–6), 369–382. DOI: 10.1007/s00426-004-0211-8.

Shiffrin, R. M. and Schneider, W. (1977). Controlled and automatic human information processing: II: perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. DOI: 10.1037/0033-295X.84.2.127.

Smith, J., Siegert, R. J., McDowall, J. and Abernethy, D. (2001). Preserved implicit learning on both the serial reaction time task and artificial grammar in patients with Parkinson's disease. *Brain and Cognition*, *45*(3), 378–391. DOI: 10.1006/brcg.2001.1286.

Squire, L. R. (1987). *Memory and Brain*. New York: Oxford University Press.

Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, memory, and Cognition*, *21*(3), 674–685. DOI: 10.1037/0278-7393.21.3.674.

Strayer, D. L. and Kramer, A. F. (1990). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 67–82. DOI: 10.1037/0278-7393.16.1.67.

Szymura, B. and Nęcka, E. (1998). Visual selective attention and personality: an experimental verification of three models of extra version. *Personality and Individual Differences*, *24*(5), 713–729. DOI: 10.1016/S0191-8869(97)00198-0.

Thompson, K. R., Sanchez, D. J., Wesley, A. H. and Reber, P. J. (2014). Ego depletion impairs implicit learning. *PloS One*, *9*(10), e109370. DOI: 10.1371/journal.pone.0109370.

Vandenbossche, J., Coomans, D., Homblé, K. and Deroost, N. (2014). The effect of cognitive aging on implicit sequence learning and dual tasking. *Frontiers in Psychology, 5*. DOI: 10.3389/fpsyg.2014.00154.

Vandierendonck, A., De Vooght, G. and Van der Goten, K. (1998). Does random time interval generation interfere with working memory executive functions? *European Journal of Cognitive Psychology*, *10*(4), 413–442. DOI: 10.1080/713752284.

Wierzchoń, M., Asanowicz, D., Paulewicz, B. and Cleeremans, A. (2012). Subjective measures of consciousness in artificial grammar learning task. *Consciousness and Cognition*, *21*(3), 1141–53. DOI:10.1016/j.concog.2012.05.01.

Wierzchoń, M., Gaillard, V., Asanowicz, D. and Cleeremans, A. (2012). Manipulating attentional load in sequence learning through random number generation. *Advances in Cognitive Psychology*, *8*(2), 179–195. DOI: 10.5709/acp-0114-0.

Willingham, D. B. and Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: evidence for parallel development. *Psychological Science*, *10*(6), 531–534. DOI: 10.1111/1467–9280.00201.

Wolf, K. and Müller, N. G. (2012). Attention and implicit learning. In Seel N. M. (Ed.) *Encyclopedia of the Sciences of Learning*, (pp. 350–353). Boston, MA: Springer. DOI: 10.1007/978-1-4419-1428-6_717.

Ziori, E., Pothos, E. M. and Dienes, Z. (2014). Role of prior knowledge in implicit and explicit learning of artificial grammars. *Consciousness and Cognition*, *28*, 1–16. DOI: 10.1016/j.concog.2014.06.003.

INDEX