The transition from implicit to explicit knowledge representations in implicit sequence learning

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Clarissa Lustig

aus

Bad Neuenahr

November / 2021

1. Berichterstatter: Prof. Dr. Hilde Haider (Köln)
2. Berichterstatter: Prof. Dr. Robert Gaschler (Hagen)
Tag der mündlichen Prüfung: 29.04.2022

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät zu Köln im April 2022 angenommen.

Zusammenfassung

Implizites Lernen kann als unbewusster Lernprozess betrachtet werden, der ohne Intention stattfindet und es dem Lernenden ermöglicht, sich an Regularien in der Umwelt flexibel anzupassen. Implizite Lernprozesse wurden in der Forschung bereits intensiv untersucht, unter anderem unter der Verwendung der seriellen Wahlreaktionsaufgabe (SRTT; Nissen & Bullemer, 1987). Mit der SRTT können Versuchsteilnehmer eine sequenzielle Struktur lernen, über die sie vorher jedoch nicht informiert wurden. Anhand von Reaktionszeit-Daten lässt sich erkennen, dass die Versuchsteilnehmer die Sequenz tatsächlich oft implizit lernen, ohne sie jedoch hinterher explizit benennen zu können.

Die Betrachtung der Bedingungen und Prozesse, unter denen Wissen in impliziten Lernsituationen bewusst wird, ist ein Bereich, zu dem nach wie vor intensiv geforscht wird und in dem es bislang keine einheitliche Theorie gibt. Vielmehr gibt es eine Reihe unterschiedlicher Sichtweisen über die Entstehung expliziten Wissens in einer impliziten Lernsituation. Außerdem wurden klassische Bewusstseinstheorien bislang wenig zur Erklärung bewusster Gewahrwerdung beim impliziten Lernen herangezogen (Cleeremans & Jimenéz, 2002)

Das Ziel dieser Arbeit lag darin, genauer zu untersuchen, unter welchen Bedingungen Wissen in einer impliziten Lernsituation bewusst wird, sowie die zugrunde liegenden Prozesse genauer zu verstehen. Zur Beantwortung der Fragestellung wurden die relevanten Theorien diskutiert sowie zwei Studien durchgeführt. In der Überblicksarbeit wurden zwei theoretische Sichtweisen im impliziten Lernen, die Unexpected Event Hypothesis (UEH; Frensch et al., 2003) sowie die Stärkeannahme (Cleeremans & Jimenéz, 2002), in Zusammenhang mit zwei klassischen Bewusstseinstheorien, der Global Workspace Theory (GWT; Baars, 1997) sowie den Higher Order Thought Theorien (Lau & Rosenthal, 2011) betrachtet. Im Ergebnis scheint die UEH in Kombination mit beiden Bewusstseinstheorien vielversprechend zu sein, eine Vielzahl der Befunde im impliziten Lernen, die sich auf die Entstehung expliziten Wissens beziehen, zu erklären.

In Studie 1 wurde die Rolle unerwarteter Ereignisse für die Entstehung expliziten Wissens untersucht. Ein Befund dieser Studie ist, dass sich das Erleben eines unterwarteten Ereignisses vor allem auf die Entstehung expliziten Wissens auswirkt. Implizites Wissen scheint von dem Erleben unerwarteter Ereignisse hingegen nicht beeinflusst zu werden.

In der zweiten Studie wurde das Verhalten der Versuchsteilnehmer bei der Entwicklung expliziten Wissens untersucht. Konkret wurde der Grund für den starken Abfall in den Reaktionszeiten (RT-drop; Haider & Frensch, 2005, 2009), der oft mit dem Entstehen

expliziten Wissens auftritt, genauer betrachtet. Es hat sich gezeigt, dass der RT-drop eine Verhaltenskonsequenz und nicht ein Vorbote für die Entwicklung expliziten Wissens ist. Sobald das Wissen bewusst ist, kann die Bearbeitungsstrategie in der Aufgabe von einer stimulusbasierten zu einer wissensbasierten werden. Sind die Bedingungen jedoch ungünstig für eine wissensbasierte Strategie, wird die stimulusbasierte Strategie beibehalten.

Insgesamt liefern die Überblicksarbeit und die zwei Studien eine gute Basis für weitergehende Forschungsfragen durch die Untersuchung einzelner Aspekte der UEH. Sowohl die theoretische Verknüpfung mit den Bewusstseinstheorien als auch die Erkenntnisse zu den Prozessen unerwarteter Ereignisse könnten dabei helfen, Forschungsfragen präziser zu formulieren und experimentelle Manipulationen gezielter zu gestalten.

Abstract

Implicit learning can be seen as unconscious learning process that takes place without the intention to learn and that allows to adapt to regularities in the environment. Implicit learning processes have been broadly investigated, amongst others by using the Serial Reaction Time task (SRTT; Nissen & Bullemer, 1987). With the SRTT, participants can learn a sequential structure without being informed about this structure. By means of the behavioral data as for instance reaction times one can see that the participants can learn the sequence implicitly but often do not show any explicit knowledge.

The conditions and processes, under which explicit knowledge occurs in implicit learning, belong still to the contents of current research and there still does not exist any theoretical consensus among the researchers. Rather, different viewpoints about the development of explicit knowledge had been established. Furthermore, classical theories of consciousness had rarely been used to explain awareness in implicit learning (Cleeremans & Jimenéz, 2002).

The goal of the current work was to investigate under which conditions explicit knowledge arises in implicit learning and to better understand the underlying processes.

To answer these questions, the relevant theories had been reviewed. Furthermore, two studies had been conducted. In the review, two theoretical points of view in implicit learning, the Unexpected Event Hypothesis (UEH; Frensch et al., 2003) and the strengthening account (Cleeremans & Jimenéz, 2002), had been discussed in the context of two classical theories of consciousness, the Global Workspace Theory (GWT; Baars, 1997) and the Higher Order Thought Theories (Lau & Rosenthal, 2011). As a result, the UEH in combination with both theories of consciousness seems to be promising in explaining a multitude of findings regarding to the development of explicit knowledge in implicit learning.

In Study 1, the role of unexpected events for the development of explicit knowledge was investigated. We found that the experience of an unexpected event influences the development of explicit knowledge. In contrast, implicit knowledge remains rather unaffected from unexpected events.

In the second study, the behavioral components of awareness in implicit learning had been investigated. Here, the role of the strong decrease of reaction times (RT-drop; Haider & Frensch, 2005, 2009) that often accompanies the occurrence of explicit knowledge was examined. It has been shown that the RT-drop might be a behavioral consequence of rather than a precursor for the development of explicit knowledge. As soon as explicit knowledge arises, the strategy might shift from a stimulus-based to a knowledge-based task performance. If the conditions are inappropriate for a strategy shift, the old strategy may be maintained.

In sum, the review and the two studies provide a good base for further research by investigating single aspects of the UEH. Both, the theoretical connection of the UEH with theories of consciousness, as well as the findings about the processes of unexpected events can help to formulate precise research questions and to systematically design future experiments.

Content

1. Introduction

Implicit learning is an unconscious learning mechanism that can take place without the learners' intention to learn. Learning the mother tongue or social behavior in children are frequently used examples to illustrate unintentional learning as it occurs incidentally and without any cognitive effort. A further example from everyday life could be cooking. While some persons learn cooking explicitly by following the instructions in a recipe, others may learn it incidentally by observing or helping other people. More globally, implicit learning fulfills an important function as it enables the learner to adapt to the regularities in the environment.

To investigate implicit learning scientifically, paradigms as, for instance, the Serial Reaction Time Task (SRTT; Nissen & Bullemer, 1987) are used. In the SRTT, a sequential structure is built into the task that the participants are not informed about. Nevertheless, the behavioral data show that the participants can learn this sequential structure. However, the participants often do not even notice that they have learned something and thus, do not show any explicit sequence knowledge. Therefore, it is concluded that learning in this situation is implicit.

Despite more than thirty years of research history, there seems to be little consensus regarding some of the core issues in implicit sequence learning. While researchers agree that implicit sequence learning takes place without a conscious intention to learn (Abrahamse et al., 2010), there is still disagreement on how knowledge becomes conscious aware in an implicit learning situation. There exist various viewpoints about how explicit knowledge arises in implicit learning, each of them with its strengths and weaknesses in explaining the underlying processes as well as the cognitive and behavioral phenomena that accompany awareness. Furthermore, although implicit learning can be seen as unconscious learning process, classical theories of consciousness are rarely used to explain implicit learning and awareness (Cleeremans & Jimenéz, 2002).

The aim of the following thesis is to contribute to this debate about how and under which conditions knowledge can become conscious aware in implicit learning situations. In a review and two studies, we aimed to clarify how classical theories of consciousness fit with current viewpoints in implicit learning and thus, how they can contribute to explain awareness in implicit learning. Furthermore, single aspects about the reasons for the occurrence of explicit knowledge as well as the behavioral components of awareness were investigated in more detail. These investigations might help to better understand the underlying processes and to provide a good experimental basis for future research.

The structure of the thesis pursues the main goal to give an answer to these questions. First, two common theories about consciousness (Chapter 2) and two viewpoints about the development of explicit knowledge in implicit sequence learning (Chapter 3) are introduced. Then, a review as well as two studies are presented, each of them examined respectively one of the core aspects in more detail. In the review (Chapter 5) two different viewpoints of how explicit knowledge in implicit learning situations arises were discussed in the light of two prominent theories of consciousness. Furthermore, the review describes, why one of the two viewpoints, the Unexpected Event Hypothesis (UEH; Frensch et al., 2003) provides a good explanation for the development of explicit knowledge in implicit learning in combination with aspects of both theories of consciousness. Based on the assumptions of the UEH, the first study (Chapter 6) further investigated the role of unexpected events for the transition from implicit to explicit sequence knowledge while the strength of knowledge representations was kept constant. Here, response-effects as perceivable changes in the environment were used to create those unexpected events. The second study (Chapter 7) focused on specific aspects of the transition from implicit to explicit sequence knowledge and investigated the suddenness of insight in a given rule in implicit learning as well as the behavioral components of these insight processes. Here, the sudden decrease of reaction times (RT-drop; Haider & Frensch, 2005, 2009) that often accompanies the development of explicit knowledge was investigated regarding to its functional role for awareness (precursor vs. consequence). Finally, the thesis closes with a discussion of the main findings of the review and the two studies particularly with regard to their respective contributions to answer the research question of this thesis (Chapter 8). Furthermore, the results might provide some implications for future directions of the implicit learning research.

2. Theories of consciousness

The term consciousness is widely discussed and until now, after decades of research, there still exists no consistent definition of what consciousness is and how it arises (Cleeremans et al., 2020). For instance, consciousness may be the subjective experience or a mental state that one is aware of and can report. Mental states in this context might be sensory states, such as sensations or perceptions, or intentional states, as for example believes or preferences (Baars, 2003; Baars et al., 2013; Lau & Rosenthal, 2011; Rosenthal, 1985).

In this chapter, two prominent theories of consciousness, the global workspace theory (GTW; Baars, 1997), as well as the higher-order-thought theories (HOTTs; Rosenthal, 1985), are introduced. Both accounts use verbal report as precondition for conscious mental states. Thus, for the purpose of this thesis, no further definition of consciousness seems to be needed. The theories are then used in the following chapters to be applied and challenged in the light of explaining the development of explicit knowledge in implicit sequence learning.

Global Workspace Theory

The Global Workspace Theory (GWT) was developed first by Baars (1997). The author compared consciousness metaphorically with a theater that contains a stage of fleeting working memory (Baddeley & Hitch, 1974) capacity. This stage receives sensory and abstract information from widespread specialized networks in the brain (Baars & Franklin, 2003). But only the information that enters in the spotlight of attention will become completely conscious. The "stage" or Global Workspace (GWS) is expected to be a neuronal system with long-distance and widespread connections in the brain (Dehaene & Nacchache, 2001). Contents can become conscious aware only when they gain access to the GWS and are broadcasted to other cognitive systems. These contents are usually processed in encapsulated modules that are domain specific and specialized. The modules work unconsciously and are only locally connected (Dehaene & Nacchache, 2001). The specialization and encapsulation of modules allow that different contents can be processed in parallel (Baars, 1997). The contents compete with each other for the access to the GWS. Contents that win the competition and are thus accessed by the GWS are no longer processed in specializes systems but broadcasted to multiple unconscious subsystems (Baars, 1997).

Selective attention might be a process that helps a content to win this competition in a winner-takes-it-all manner. According to the GWT, every experience is shaped by unconscious contexts (or frames; Baars et al., 2013) as for instance motives, emotions, or executive functions (Baars & Franklin, 2003) that might guide selective attention (Baars, 1997). Because of the winner-takes-it-all principle, the flow of conscious moments (James, 1890) is always changing depending on the competing contents (Baars et al., 2013). Due to these dynamics, it is, for instance, needed that a stimulus is presented a certain time to be processed consciously (Dehaene & Naccache, 2001).

The GWT has been implemented in different computational and neuronal network models (e.g., Franklin, 2001; IDA, Baars & Franklin, 2003) and received support from various neuronal brain imaging studies. Neuronal areas that are related to consciousness seem to be the frontal cortex as well as the parietal cortex (Baars, 2003). Furthermore, neuronal amplifications of the GWT focus not only on the role of the cortex but also on the role of the thalamus for consciousness (e.g., neuronal GWT, Dehaene & Nacchache, 2001). However, there does not exist only one single area in the brain where one can find the GWS. Rather, the GWS seems to be based on multiple cortical processors that are hierarchically organized (Dehaene & Naccache, 2001; Dehaene et al., 2011). According to Dehaene and colleagues (2011), a stimulus can become conscious in two phases. First, the stimulus enters the hierarchy of processors bottom-up and unconsciously. Once selected for the actual goals, it is processed in a top-down manner by neurons of the GWS. Only one content can become conscious to the same time, whereas all other contents are inhibited. This is assumed to occur in an all-or-none manner and expressed by ignition, a sudden firing of the neurons in the different cortical and subcortical areas (Dehaene et al., 2003).

The function of consciousness sees Baars (1997, 2003) in access that then enables adaption, error detection and control of behavior. Furthermore, consciousness allows optimizing the balance between organization and flexibility.

Higher-order thought theories

Higher-order thought theories (HOTTs) sum up a group of theories instead of one single theory (Lau & Rosenthal, 2011). In contrast to first-order theories (FOTTs), which assume that for a mental state to be conscious the (neuronal) representation of this mental state is sufficient (e.g., Block, 2007) or has to be strengthened (GWT; Brown et al., 2019), the core assumption of all HOTTs is that besides this first-order (FO) representation about a mental state (e.g., “the green grass”), a higher-order (HO) or meta-representation about oneself to be in this mental state (e.g., “it’s me that sees the green grass”) is needed (Lau & Rosenthal, 2011). According to Rosenthal (1985, 2002), this higher-order thought (or higher-order awareness; Rosenthal, 2012) itself does not need to be conscious. So that the higher-order thought or second-order thought becomes conscious, a third-order thought about this higher-order thought, and thus, a thought about the thought would be needed. However, this is supposed to be seldom as one usually does not think about one's own thoughts (Lau & Rosenthal, 2011; Rosenthal, 1985).The various HOTTs differ in their concrete definition of HO mechanisms (Brown et al., 2019). But all these theories have the assumption in common that even if a FO representation remains constant, a change in HO representations is sufficient for a change in subjective awareness (Lau & Rosenthal, 2011).

As the GWT, HOTTs are supported by evidence of neurophysiological studies. While FO representations depend on neuronal activity in early sensory areas, HO representations are accompanied by activation in the prefrontal and parietal cortex (Lau & Rosenthal, 2011). As described above, in the GWT it is also assumed that consciousness is based on prefrontal and parietal activation (Baars, 2003). However, Lau and Rosenthal (2011; Rosenthal, 2012) suppose that the GWT and HOTT can be distinguished by studies that show consistent task performance with different levels of awareness and thus, a respective activation in the different neuronal areas (e.g., Lau & Passingham, 2006). The GWT supposes that the prefrontal and parietal activity represents both consciousness and task performance. According to the HOTTs, task performance depends on FO instead of HO representation and therefore, on the activation of early sensory areas. Thus, a change in awareness by constant task performance and the activation of the respective brain areas might support these differences in FO and HO representations and thus speak in favor for the HOTTs.

Regarding to the functional role of consciousness, the HOTTs also differ from the GWT that supposes the function of consciousness in behavioral components as action control and optimizing of task performance. In contrast, no special function or utility of consciousness is assumed according to the HOTTs (Lau & Rosenthal, 2011).

Finally, the HOTTs have been criticized because a creature has to be cognitively able to build higher-order thoughts. However, animals and children whose cerebral cortex is not yet fully developed are also conscious aware of their mental states (e.g., Seager, 2004). The different HOTTs give different answers to this point (Brown et al., 2019). Rosenthal (2002; Lau & Rosenthal, 2011) argued, for instance, that the conscious awareness of human adults might be finer graded than that of children or animals. Furthermore, contemporary theories of consciousness aim to integrate classical theories as the GWT and HOTTs to use the advantages of each theory and to encounter possible problems and critics in a combined concept (e.g., Cleeremans et al., 2020; Shea & Frith, 2019).

3. Implicit learning and the development of explicit knowledge

As described before in the introduction section, implicit learning refers to unintentional learning. Furthermore, the content of what has been learnt often remains unconscious. There exist various paradigms to investigate different forms of implicit learning as well as their cognitive mechanisms. For instance, in Artificial Grammar Learning (AGL; Reber, 1967), letter strings that follow a certain grammar are presented to the participants. Only after the training phase, the participants are informed about the existence of the grammar. In a following test phase, the participants are asked to classify letter strings as grammatical or not grammatical. The common finding in the AGL is that participants classify the letter strings correctly as grammatical or not grammatical without being able to name the underlying grammar. Thus, participants seem to have learned implicitly something about the statistical structure of the grammar and about the likelihood that a given string belongs to this structure.

A somewhat different paradigm for the investigation of implicit learning is the Serial Reaction Time Task (SRTT; Nissen & Bullemer, 1987). Within this task, participants usually learn a sequential structure, rather than a statistical structure. In a typical SRTT, squares are presented horizontally on the computer screen. These squares are mapped to keys on a keyboard. In each trial, an asterisk appears as target in one of the squares and the participants' task is to press the corresponding key. Unbeknownst to the participants, the order of the target positions and thus, of the to-be-pressed keys, follows a repeated sequence. A general finding of the SRTT is that most participants show learning during training but are often not able to verbalize their knowledge or do not even notice the existence of a repeated sequence. Thus, learning in this task is assumed to be implicit. Furthermore, amnestic patients often show similar implicit learning effects but fail to report explicit sequence knowledge. This finding usually serves as evidence that the performance of direct (e.g. verbal report) and indirect tasks (e.g. task performance in a SRTT) reflects different memory systems, namely the declarative (explicit) and non-declarative (implicit) memory system (Nissen & Bullemer, 1987; Reber & Squire, 1994, 1998; Keane et al., 1995).

Using different variants of the SRTT as the implicit learning paradigm, the investigated topics are manifold and range from the question of whether attention is needed for implicit learning (Jiménez & Mendez, 1999) to the content of the acquired sequence representations as, for instance, stimulus-stimulus or response-response associations (Abrahamse et al., 2010). Relating to the observation that some participants develop explicit knowledge in implicit sequence learning while others do not, some of the important questions in implicit learning research refer to the conditions for the development of explicit knowledge as well as the underlying cognitive mechanisms. Based on the findings of this research, there exist now various theories about the development of implicit and explicit knowledge in implicit learning that differ particularly in their assumptions about the involved learning systems and the characteristics of implicit and explicit knowledge representations. These theories can be broadly subsumed in single and multiple system views that will be discussed in the following sections (for an overview, see, Frensch et al., 2003).

Single system views

Within the single system account, the existence of only one single learning system is supposed by the different theories. For instance, Shanks and colleagues (Shanks & Johnstone, 1999; Shanks & St. John, 1994) deny that the evidences in the implicit learning research demonstrate the existence of unconscious knowledge in implicit learning situations and attribute the often found lack of verbal report to methodological problems. According to the authors, differences in task performance and awareness, which are often reported as evidence for unconscious knowledge as a result of implicit learning, are grounded in the different sensitivities of the respective measures. Thus, while task performance measures primarily skill acquisition, awareness is assessed by tests like for instance generation tasks (Dienes & Seth, 2010; Haider et al., 2011; Shanks & Johnston, 1999) or verbal report (Rünger & Frensch, 2010). These differences in the sensitivity of the measures might also explain the finding that amnestic patients perform well in implicit memory tests but fail in explicit memory tests. Shanks and St. John (1994) argue that the implicit memory tests might be less sensitive for differences between amnestic patients and healthy control participants than explicit memory tests. If this were the case, amnestic patients might have learned overall less than healthy control participants but only the explicit memory tests are be sensitive enough to register these group differences. Support for this argumentation stem, for example, from simulation studies with the simple recurrent network model (SRN; Kinder & Shanks, 2001, 2003). Kinder and Shanks (2003) investigated whether the SRN, which is based on a single system account, can explain the differences between amnestic patients and healthy control participants. They successfully replicated a study from Keane and colleagues (1995) who showed that amnestic patients are impaired in recollection or recognition tasks but show normal performance in indirect tasks such as priming tasks. Based on the simulations with the SRN, Kinder and Shanks (2003) concluded that direct and indirect memory tasks can be based on a single mechanism and that amnestic patients might suffer from general memory impairment.

Besides the varying sensitivity of different measures in implicit learning, a second methodological problem might be the lacking process purity of awareness tests. Process purity in this case means that explicit awareness tests should not be influenced by implicit knowledge (Curran, 2001; Wilkinson & Shanks, 2004). Destrebecqz and Cleeremans (2001) tried to solve this problem by using the Process Dissociation Procedure (PDP; Jacobi, 1991), for the implicit-explicit differentiation measure in a SRTT. The rationale behind the PDP is that only knowledge that is explicit should be under intentional control and thus suppressible if asked for. Thus, subsequent to the SRTT training, participants were informed about the existence of a sequence in the training phase and asked to generate strings of SRTT elements under either an inclusion or exclusion condition. In the inclusion condition, these strings should contain the training sequence. In contrast, in the exclusion condition, participants should avoid using parts of the trained sequence to generate their strings.

A general finding is that participants, who reported explicit sequence knowledge, were able to suppress their knowledge to generate own strings without parts of the trained sequence in the exclusion condition. In contrast, participants with rather implicit knowledge generated more strings with parts of the trained sequence in the exclusion condition. These participants had little control about their implicitly acquired knowledge and failed to suppress this knowledge in generating the strings. In the inclusion condition, all participants generated a comparable amount of own strings that contained parts of the trained sequence. These findings support the assumption that the PDP is process pure and can distinguish between implicit an explicit knowledge in implicit sequence learning. However, Wilkinson and Shanks (2004) replicated the study of Destrebecqz and Cleeremans (2001) and found knowledge control in form of the suppression of implicit acquired knowledge to generate test strings under the exclusion condition in all experimental conditions. They claimed that the difference to the results of Destrebecqz and Cleeremans (2001) might be explained by a lack of power in the original study and concluded that knowledge that is acquired in implicit learning situations is accessible for control and thus explicit.

Similar to Shanks and colleagues, Perruchet and Vinter (2002) assume that a concept of unconscious knowledge representations is not needed. Instead, conscious mental live is sufficient to explain behavior. The authors propose a mentalistic framework as alternative for computational information processing models that are often based on unconscious processes and that consider consciousness as access to the output of cognitive processes. In contrast, the mentalistic framework denies the existence of unconscious representations. Only the neural processes that form representations are unconscious. The representations themselves are embedded in the phenomenal experience and as such, are conscious. Perruchet and Vinter (2002) tested their assumptions with a computational model (SOC; self-organizing consciousness) and conducted simulations that support the idea that unconscious representations are not needed and that complex conscious representations cause behavior.

Implications of the SOC model for implicit learning are based on the formations of chunks or units (Perruchet & Vinter, 2002). In AGL, the general finding is that participants perform grammatically without being able to name the rules. This finding is usually interpreted in the sense that participants have learnt the rule implicitly and perform well based on their acquired implicit knowledge. However, they are not able to explicitly name the grammar (Reber, 1967). The reinterpretation of this finding in sense of the SOC is that the training modifies how the data are consciously coded and perceived (Perruchet & Vinter, 1998). Conscious units or chunks are formed during training and when these chunks are experienced as being familiar and thus processed fluently, they are interpreted as grammatically. A similar process is assumed for implicit sequence learning. In the SRTT training, also chunks or sequence units might be formed, here in form of, for instance, movement patterns (Perruchet et al., 1990).

A slightly different view stem from Cleeremans and Jiménez (2002; Cleeremans, 2006) who do not question the existence of implicit knowledge representations per se. Rather, they deny in their early work that implicit and explicit learning are based on different learning systems. The authors assume that the cognitive system comprises information processing modules that are interconnected and hierarchically organized. Learning, as a consequence of information processing, changes the pattern of connectivity and of connection weights between these processing modules. These connectivity patterns characterize knowledge representations whose quality is improved during learning. The quality of such a knowledge representation is characterized by the dimensions strength, stability in time and distinctiveness. Strength defines how many processing units are involved to generate a representation and how strong these modules are activated, stability in time depends on how long the representations can be hold active and distinctiveness refers to the overlap of representations. During implicit learning, knowledge representations are acquired. At the beginning of the implicit learning process, the quality of the acquired knowledge representations is low. Nevertheless, these representations can yet affect behaviour, even in the absence of awareness. During learning, the quality of the knowledge representations increases gradually. This increase of quality is necessary for explicit representations. The change from implicit to explicit knowledge representations does not occur in an all-or-none principle, but rather graded. The function of explicit knowledge representations is then the flexible and adaptive control over behaviour (Cleeremans & Jiménez, 2002; Cleeremans, 2006).

The findings of Destrebecqz and Cleeremans (2001) support the assumption of the view that increasing the quality of knowledge representations leads to the development of explicit knowledge in implicit sequence learning. The authors found that participants who received a response stimulus interval (RSI) of 250 ms developed more explicit sequence knowledge than participants who did not receive any RSI. The amount of implicit knowledge was equal for the two conditions. The authors concluded that the knowledge representations can affect behaviour independently of their quality. Furthermore, explicit knowledge is based on high-quality representations. Third, the development of these high-quality representations needs time. Thus, participants who received an RSI of 250 ms had more time to increase the quality of the representations and thus, to develop explicit knowledge.

In sum, representatives of the single system view either deny the existence or the need of unconscious knowledge as a result of implicit learning processes (Shanks & St. John, 1994; Perruchet & Vinter, 2002) or suggest a gradual change from implicit to explicit knowledge representations (Cleeremans & Jimenez, 2002). Furthermore, the single system view is supported by empirical evidence from behavioral experiments as well as simulation studies. However, there are some findings that cannot be explained by single system views. These aspects will be discussed in the following section where the multiple system view is introduced.

Multiple system views

In contrast to the single system view, the multiple system view supposes the existence of at least two independent learning systems that work in parallel. These learning systems differ for example in their processing modes as well as in their representational formats (e.g., Willingham, 1998).

For instance, Willingham (1998) claimed in his control-based learning theory (COBALT) two learning systems that operate in parallel and independently from each other. Of these two learning systems, the ventral system needs attention and is responsible for explicit learning. In contrast, the dorsal system is responsible for implicit learning. However, despite this independence, the two learning systems interact, and this interaction is unilateral. That means that the explicit system can control the implicit system but not the other way around. The implication of COBALT for implicit sequence learning is that sequence learning without awareness is possible.

Keele and colleagues (2003) used the same systems (ventral/dorsal) in their dual-system model. However, as an important difference, the dual-system model supposes that the dorsal system is responsible for implicit learning whereas the ventral system is responsible for implicit learning from which explicit knowledge can arise. Thus, in contrast to COBALT (Willingham, 1998), the dual-system model supports both, the multiple-system view and the single-system view, as implicit and explicit knowledge might occur within the same system (the ventral system).

Supporting empirical evidence for the assumption that the implicit and explicit learning system operate independently from each other stem, for instance, from Reber and Squire (1998) who conducted studies with amnestic patients. They responded on the critic of Shanks and St. John (1994), namely that implicit learning in amnestic patients might also be impaired and that only explicit knowledge tests are sensitive enough to identify these differences, by using a crossover method. They conducted a SRTT where amnestic patients were able to develop more implicit knowledge by practicing the task during training and healthy control participants were able to generate more explicit knowledge by observing and reminding the sequence during training. After training, for amnestic patients and healthy control participants the same tests were used that measured either explicit (recognition tests) or implicit (task performance) knowledge. As a result of this crossover comparison, the amnestic patients had more implicit and less explicit knowledge than the control participants. Reber and Squire (1998) concluded from this crossover results that the performance of the participants in the explicit or implicit tests cannot stem from the same source in the brain. Thus, at least two learning systems might be involved in implicit learning.

Assumptions about the nature of implicit and explicit knowledge representations are formulated in the representational theory of knowledge (Dienes and Perner, 1999). Within this theory, knowledge means to have a representation and can vary in what is represented explicitly and which components are implicit. Three components for knowledge are defined in the representational theory of knowledge, namely the content, the attitude about the content and the person itself that has the knowledge and the attitude. The content is explicitly represented but can have parts that remain implicit. A person has an attitude about the verity of the content. This allows the person to make judgements about it. The content, the attitude and the fact that it is oneself who has the knowledge and the attitude about the content are related. Dienes and Perner (1999) applied their representational theory of knowledge to various research areas, including the AGL task (AGL; Reber, 1967) as implicit learning paradigm. They differentiate between voluntary control about knowledge and the attitude that a person has about the knowledge. For voluntary control that allows the application of the acquired knowledge, the rules of the artificial grammar have to be represented explicitly. However, participants can have indirect control without any explicit knowledge representation. In the case of AGL, this indirect control allows participants to give correct grammatical judgments without knowing the grammatical rule. The attitude that a person has about the knowledge refers to the question how explicit this person can reflect about this knowledge. The explicitness of the attitude in AGL can be measured with confidence judgements about the grammatical judgements as, for example, knowing versus guessing. If the confidence judgements correlate with the amount of correct grammatical judgements, the state of the knowledge is represented explicitly. If, however, the confidence judgements do not correlated with the amount of correct responses, the state of the knowledge is represented implicitly.

In their theory, Dienes and Perner (1999) defined the nature of representations as well as explicit and implicit components. However, they did not make any assumptions, about how exactly explicit knowledge is developed in an implicit learning situation.

To explain the transition from implicit to explicit knowledge representations, Cleeremans (2008, 2011, 2014, 2020) amplified the account of Cleeremans and Jiménez (2002; Cleeremans, 2006) by separating the unconscious and conscious systems in the radical plasticity thesis. The radical plasticity thesis picks up the idea that representations differ in their quality and thus, on the dimensions strength, stability in time and distinctness. Furthermore, during the learning processes, the quality of knowledge representations increases in a continuous, graded manner. In addition, Cleeremans (2008, 2011, 2014, 2020) assumes that such a FO-representation itself can never become conscious. For the development of conscious knowledge, a higher-order representation (Rosenthal, 1985) or meta-representation is needed. These meta-representations build the basis for verbal report. They are formed through the interaction with the environment, other individuals and oneself. During this interaction, internal representations about the environment, other individuals and oneself are re-described. In addition, the quality of meta-representations improves in the same way as that of FO-representations, that is, gradually on the dimensions strength, stability in time and distinctness (Cleeremans, 2011). Evidence for this assumption is provided, for example, by connectionist networks. Cleeremans and colleagues (2007) computed, for instance, a wagering network that is based on the findings of the Post-Decision Wagering Task (PWDT; Persaud et al., 2007). In the PDWT, the participants are asked to place a high or low wager on their decisions within a task. As a result, only participants with explicit knowledge representations and thus, with conscious knowledge about the reasons for their decisions, place high wagers for correct responses. Participants who developed only implicit knowledge representations place rather low wagers on their (even correct) decisions. The wagering network of Cleeremans and colleagues (2007) contains a FO-network, whose task it is to identify digits from zero to nine, and a HO-network, whose task it is to place high or low wagers dependent on the decision about the correctness of the answers of the FO-network. Simulations studies showed that the FO-network gradually learned to improve performance. In contrast, the HO-network performs initially well as it learns that the FO-network is incorrect in classifying the digits. Then, the performance of the HO-network decreased with the increasing performance of the FO-network. At this point in time, the performance of the whole system seems to be based on unconscious knowledge. Finally, the HO-network increased the performance parallel with the performance of the FO-network. In this phase, the performance is supposed to shift from unconscious to conscious processing (Cleeremans et al., 2007).

In contrast to Cleeremans (2011), who assumes the graded development of explicit knowledge representations, Frensch and colleagues (2003) suppose with the Unexpected Event Hypothesis (UEH) that explicit knowledge arises suddenly in an all-or-none manner. An unexpected change in the own behaviour can trigger searching processes to find an explanation for this unexpected event. Such unexpected events in the SRTT are for example premature responses (Haider & Frensch, 2005, 2009; Haider et al., 2011). The result of the search processes is what becomes conscious aware. However, the search processes do not mandatorily cause the detection of a sequence or other regularity that is embedded in the task. If another explanation for the unexpected change in the own behaviour seems to be more plausible, as for example the instruction that premature responses might occur due to a lack of attention, the regularity remains undetected (Haider & Frensch, 2005). As in other theories within the multiple system view, it is not the implicit knowledge that becomes conscious as supposed by single system theorists. Rather, a second explicit representation is developed.

Evidences for the existence of two independent learning systems, and the sudden development of explicit knowledge representations due to the experience of an unexpected event, stem from the research of Haider and colleagues (e.g., Haider & Frensch, 2005, 2009). In various experiments, the graded strengthening of knowledge representations (Cleeremans & Jiménez, 2002; Cleeremans, 2006, 2008, 2011, 2014, 2020) was tested against the UEH (Frensch et al., 2003). In sum, two general findings of these experiments provide support for the UEH. First, more explicit knowledge was found in conditions in which the participants experienced an unexpected event compared to conditions in which the participants did not. The strengthening of knowledge representations was held constant in these experiments. Haider and Frensch (2005, 2009) tested for example the role of premature and correct responses (responses that are given before stimulus presentation) as unexpected events. They found that increasing the probability of premature responses led to more explicit sequence knowledge compared to a smaller probability of premature responses. In another study, Esser and Haider (2017b) manipulated the experienced fluency of performing a SRTT by using deviants. These deviants were presented either in a blocked (blocked order condition) or a random order (random order condition). The total amount of deviant trials was identical in both conditions. Thus, while strengthening of knowledge representations was held constant in the two conditions, only the participants in the blocked order condition experienced an unexpected variation in the task fluency that triggered searching processes and thus, the development of explicit sequence knowledge. In contrast, the participants in the random order condition did not experience the unexpected variation in the fluency and therefore, did not start search processes. As a result, the participants in the blocked order condition developed more explicit knowledge than participants in the random order condition.

A second empirical finding supporting the assumptions of the UEH is that participants who develop explicit sequence knowledge often express a sudden decrease of reaction times (Haider & Frensch, 2005, 2009; see also, Willingham, 1998). This sudden decrease of reaction times that accompany awareness, the so-called RT-drop, might be the point in time when participants change their responding strategy from stimulus-driven to top-down performance due to the acquired explicit knowledge representations (Haider & Frensch, 2005; Haider et al., 2005; Haider et al., 2011; Rose, et al., 2010; Wessel et al., 2012).

In sum, there exists widespread evidence for different theories that suppose independent implicit and explicit learning systems. The following sections focus on the assumptions and predictions of the strengthening account from Cleeremans (Cleeremans & Jiménez, 2002; Cleeremans, 2006, 2008, 2011, 2014, 2020) as well as the UEH (Frensch et al., 2003) as two viewpoints that deal particularly with the transition from unconscious to conscious knowledge representations.

4. Overview of the studies

In order to investigate the main question of this thesis, namely how and under which conditions awareness arises in implicit sequence learning, the following review and two experimental studies considered three different core aspects. These aspects reach from a global review of selected theories to a more detailed investigation of single predictions of the UEH as one of the theories in implicit learning. The review aimed to embed the current theories and research in implicit learning in a broader theoretical context of two classical theories of consciousness. Here, the strengthening account (Cleeremans & Jiménez, 2002; Cleeremans, 2006, 2008, 2011, 2014, 2020) as well as the UEH (Frensch et al., 2003) as two prominent views in implicit learning were reviewed with its strengths and weaknesses in the application of the GWT (Baars, 1997, 2003) or the HOTTs (Lau & Rosenthal, 2011; Rosenthal, 1985). As a result of this review, the application of only the GWT or the HOTTs in implicit learning requires additional clarifications. Therefore, we furthermore discussed the advantages of linking the UEH with a combination of the GWT and the HOTTs in order to explain how explicit conscious knowledge arises in implicit sequence learning.

Due to the existing supporting empirical evidence for the UEH (see, chapter 3), the assumptions and predictions of the UEH were pursued as basis for the following two empirical studies. In Study 1, a SRTT was conducted to investigate the role of specific unexpected events, here the role of response-effects, for the development of explicit knowledge in implicit sequence learning when strengthening of knowledge representations was held constant.

Finally, Study 2 aimed to further investigate a single prediction of the UEH, namely the suddenness of conscious insight in an existing rule once an unexpected event had been experienced. Practically, the behavioral components were investigated that accompany these insight processes, i.e. the sudden decrease of RTs. Here, the concrete research question concerned the role of these so called RT-drops and thus, whether an RT-drop is a precursor for the development of explicit knowledge representations or whether it is a behavioral consequence of conscious awareness.

As described in Chapter 3, the detection of unconscious knowledge in implicit learning might also depend on the sensitivity and process purity of the respective measure (Shanks and St. John, 1994) and on the respective definition of consciousness (e.g., Frensch & Rünger, 2003). To take the methodological problems of explicit measures into account, in Study 1 and Study 2 the Post-Decision Wagering Task (PDWT; Persaud et al., 2007) was used in addition to verbal reports to assess explicit sequence knowledge after the SRTT training. The PDWT is similar to the training SRTT task. But in some trials, a question mark is presented instead of the target and the participants have to guess the next response. In addition, they have to place a high or low wager depending on their confidence in the correctness of their response. The rationale behind the task is that participants with implicit knowledge give more correct responses than it is expected by guessing but do not show any correlation between the correctness of their response and their wagers (zero correlation criterion; Dienes et al., 1995). However, participants with explicit knowledge place high wagers for correct and low wagers for incorrect responses. Thus, the PDWT seems so be a measure that is sensitive enough to investigate whether a participant had developed primarily implicit or explicit knowledge (e.g., Dienes & Seth, 2010).

5. Review: What triggers explicit awareness in an implicit sequence learning situation? Implications from theories of consciousness.

Abstract

This article aims to continue the debate how explicit, conscious knowledge can arise in an implicit learning situation. We review hitherto existing theoretical views and evaluate their compatibility with two current, successful scientific concepts of consciousness: The Global Workspace Theory and Higher Order Thought Theories. In this context, we introduce the Unexpected Event Hypothesis (Frensch et al., 2003) in an elaborated form and discuss its advantage in explaining the emergence of conscious knowledge in an implicit learning situation.

Implicit learning research is concerned with situations in which individuals learn sequential deterministic or probabilistic contingencies, but lack the intention to learn about these contingencies. Such sequential contingencies might comprise sequential actions (such as learning to blindly type on a keyboard), perceptual events (e.g. learning the sequence of notes in a melody or of visual regularities when driving the same route repeatedly) or complex regularities (e.g. arithmetic principles, Prather, 2012). Empirically, two different tasks are common to investigate implicit learning: The Serial Reaction Time Task (SRTT; Nissen & Bullemer, 1987) and the Artificial Grammar Learning Task (AGL; Reber, 1967). In the SRTT, participants typically respond to sequential stimuli with sequential responses. In general, learning of the underlying sequence is most commonly shown by comparing the reaction times of sequential trials with trials containing a new, unknown sequence. In the AGL, participants learn by observing probabilistic contingencies between strings of, for example, letters or other visual stimuli. Here, learning can, for example, be inferred by showing that participants are able to discriminate strings that follow either the old or a new grammar. The hidden regularities in the SRTT and the AGL are usually not mentioned to the participants.

Interestingly, on a subjective level, participants in these paradigms often seem to lack conscious insight that they have learned something. In these cases, it is often presumed that participants acquired unconscious or implicit knowledge. In the simplest sense, when a participant is asked to report the learned sequence at a later time and is unable to do so, is said to have “implicit knowledge”. Participants who can report their knowledge are said to possess “explicit knowledge”. Of course, a large debate is centered around the question how to determine whether the acquired knowledge in implicit learning paradigms is in fact unconscious (see e.g., Newell & Shanks, 2014; Peters & Lau, 2015).

Various objective or subjective measures to determine the conscious or unconscious status of the developed knowledge have been suggested over the years. Objective measures aim to show a performance-based difference between conscious and unconscious knowledge. For instance, process dissociation tasks have been suggested as objective measures (Destrebecqz & Cleeremans, 2001; for critical points see Stahl et al., 2015). These tasks aim to show that unconscious (implicit) knowledge leads to correct responses when the participants are asked to generate the learned sequence on their own (inclusion condition), but also fail to inhibit their acquired sequence knowledge when asked to generate new sequences (exclusion condition). Quite different, subjective measures emphasize the lack of metacognitive knowledge or subjective awareness. While participants with implicit knowledge are often able to predict the next correct response or stimulus, they express subjective uncertainty. For example, they put low instead of high wagers on the correctness of their responses. By contrast, participants with conscious (explicit) knowledge express high certainty on correct responses but low certainty on wrong responses (Dienes & Seth, 2010; Haider et al., 2011; Persaud et al., 2007). Thus, there are conceptual differences when assessing the status of knowledge acquired in an implicit learning situation. Furthermore, presupposed that implicit learning tasks can lead to implicit, unconscious knowledge, there usually are participants who develop explicit, consciously accessible knowledge. Based on these findings, we will adapt the term "implicit knowledge" for subjectively and behaviorally unconscious knowledge and "explicit knowledge" for conscious knowledge.

This difference in conscious availability of the acquired knowledge suggests that implicit learning settings could provide a promising opportunity to investigate how implicit or explicit knowledge is related to the representational status of the acquired knowledge and which mechanism mediates between implicit and explicit knowledge. Is explicit knowledge based on the same acquired representation as implicit knowledge? If so, does the same learning mechanism change the representational quality (e.g. its strength is increased over the learning period) which is then accompanied by a gradual change in conscious accessibility? Or does the conscious status depend on a different access to that representation? A different possibility would be that implicit knowledge is based on different representations than explicit knowledge. These latter two options would imply that an "implicit learning" process leads to strictly implicit knowledge, while explicit knowledge requires an additional process that grants conscious accessibility.

This leads to a close connection between implicit learning research and general theories of consciousness which aim to explain how consciousness originates from unconscious processes. This way, implicit sequence learning can enrich theoretical views on consciousness with new experimental approaches and help to test predictions, thereby, improving current weaknesses or identifying unexpected problems of the respective theories. Likewise, the question how explicit knowledge arises from an implicit learning situation can profit from testing predictions from general theories of consciousness.

The central aim of this article is thus to discuss two current models in implicit learning research that aim to explain how consciously accessible, explicit knowledge can develop from implicit learning situations and to assess their correspondence with current theories on consciousness. One of these views proposes that explicit knowledge depends on the strengthening of associative links (i.e., the representational quality). In this case, it is assumed that conscious knowledge will develop gradually, either by strengthening single transitions (Cleeremans, 2008, 2011; Cleeremans & Jiménez, 2002) or by integrating single elements into sequence chunks (Perruchet & Vinter, 2002; Perruchet, 2002). On the other hand, there are theories that propose an indirect link between implicit and explicit knowledge. One of these theories is called the Unexpected Event Hypothesis (Frensch et al., 2003). From this theoretical viewpoint, unconscious (implicit) knowledge and conscious (explicit) knowledge are based on different representations. Implicit knowledge relies on representations acquired during learning that lead to behavioral changes (e.g. faster RTs). Explicit knowledge develops by detecting mismatches between expected and experienced behavior. If such a mismatch is detected a second (and subjectively explicit) inferential process is triggered that can lead to the discovery of sequential contingencies inherent in the task. In this case, a new representation of the sequential contingencies is acquired, which is independent from the implicitly learned representation. This new representation is explicitly accessible. Thus, explicit knowledge is not assumed to develop gradually from implicit knowledge, as is presumed by strengthening theories, but instead in an abrupt change from an unconscious into a conscious knowledge state.

We will review current evidence for each of these theoretical classes. Furthermore, these two viewpoints and their empirical evidence will be reconciled with two prominent theories of consciousness: One is the Global Workspace Theory (Baars, 1997, 2005), the other is the Higher Order Thought Theory (Lau, 2008; Rosenthal, 1997, 2012). We will discuss the advantages and disadvantages of each perspective and their implications for the development of explicit knowledge in implicit learning situations.

The Serial Reaction Time Task (SRTT) and its Relevance for Consciousness Research

In this section, we will briefly outline how implicit learning paradigms can provide an important opportunity to investigate the connection between unconscious and conscious processes. As we have briefly stated in the introduction, there are various implicit learning paradigms which all have in common that a sequence of responses or stimuli is learned without the intention to do so. Here, we will mainly focus on the SRTT, as we believe that its simple, yet flexible structure makes it a particular interesting instrument to investigate how implicitly acquired representations can become consciously accessible.

The SRTT has the potential to provide both, online and offline measures of learning. During the training part, in which participants are trained with a sequence of stimuli or responses, error rates and reaction times can be utilized to assess learning. Subsequent tests, such as the previously described Inclusion/ Exclusion Task or the Wagering Task, are usually employed to determine whether or to what extent the acquired knowledge is implicit or explicit. However, the training part of the task can also have the potential to detect whether knowledge is explicit. Moreover, it might also have the potential to confine the time frame in which a qualitative change from an implicit to an explicit representation of the learned sequence occurs. It has been found that participants who develop explicit knowledge in a SRTT often show a sudden decrease of reaction times in the training phase. This so-called RT-drop indicates that participants change their behavior qualitatively (Haider & Frensch, 2005; Haider et al., 2005; Lustig et al., 2021, Rose, et al., 2010; Wessel et al., 2012) and do not mandatorily need to process the stimulus anymore to perform the task. This stimulus-independency of explicit sequence knowledge is further supported by studies that show that participants are no longer affected by incongruent stimulus-response characteristics, such as the Stroop Effect (Haider et al., 2011) and the Simon Effect (Koch, 2007), or by a frequency-induced response bias (Tubau et al., 2007). Thus, analyzing reaction times in a SRTT has the potential to classify participants with explicit sequence knowledge online during training, as well as potentially grasping the point in time when explicit knowledge occurs (Rose et al., 2010).

From this perspective, the SRTT is a useful addition to the investigation of qualitative differences between unconsciously and consciously accessible representations, as well as the mechanisms that mediate between them. Mostly, such questions are approached by priming studies (e.g., Del Cul et al., 2009; Kouider & Faivre, 2017; Kouider et al., 2010; Lau & Rosenthal, 2011; Overgaard, 2003). Priming studies provide a well-controlled opportunity to study the differences between conscious and unconscious processing on a trial-by-trial basis. In priming studies, unconscious processing is operationalized either by weak signal strength or by inattention towards the critical stimuli or stimuli features (Dehaene et al., 2006). The advantage of implicit learning paradigms, however, is that they provide an opportunity to create situations where signal strength can be controlled in a wider range (e.g., by controlling training duration or sequence complexity) while the resulting representations have a high temporal stability (Tamayo & Frensch, 2015). At the same time the role of attention towards the sequence can be investigated, with evidence so far suggesting that selective attention (i.e., having participants respond to the sequential feature vs. responding to a sequence-irrelevant feature) towards the sequence does not lead to more conscious knowledge but might be a prerequisite for implicit learning (Jiménez & Méndez, 1999; Turk-Browne et al., 2005). A further difference between priming and implicit learning paradigms is that the former can provide insight how certain singular simple or complex (most often visual) percepts which already have an existing entry in long-term memory can become conscious. By contrast, implicit learning studies can investigate how whole newly acquired knowledge structures about relations between clearly perceivable events in the environment can become conscious. Implicit learning can hereby help to investigate how the cognitive system can learn about its own internal states; how the system changes from a state of not knowing that the internal knowledge base has changed, to knowing that new knowledge has been developed and that current consciously accessible believes about states in the world need to be adjusted or replaced. Thus, implicit learning research could also be seen as an interesting link to research on insight and creativity in which it is often proposed that creative ideas result from or, at least, are largely supported by consciously inaccessible processes (Dietrich & Haider, 2015; Fedor et al., 2017).

The term creative "insight" implies that there is a very brief moment in time in which the cognitive system changes its state from consciously not knowing to knowing. Nevertheless, it is important to see that research on insight and creativity often stresses that such apparently sudden "insight" moments are preceded by other events, including the moment of noticing that there is a problem that needs to be solved over an incubation period (spanning from years to minutes) until finally coming up with a new idea (Cosmelli & Preiss, 2014; Dietrich & Haider, 2015; Hélie & Sun, 2010). Thus, there might be different, hierarchical representational contents an individual can become aware of (Kouider et al., 2010). In an implicit learning task this can encompass consciously noticing that that there is something about the task that has not yet been recognized before, noticing that certain events (finger movements, stimuli, etc.) appear systematically, to lastly discovering the actual sequence that governs the task. However, regardless of the specific content, these changes in awareness could result from a sudden or a gradual change in the underlying representations. While the SRTT has the potential to explore various aspects of these insight stages, what might be of particular interest in the light of the Unexpected Event Hypothesis is the moment where an individual notices that their consciously held beliefs about the task do not match their perceived experiences, which could trigger an inferential process that finally leads to recognizing the sequence in the task.

How to Conceptualize the Transition from Implicit to Explicit Sequence Knowledge

In the following, we will briefly outline two theoretical viewpoints about the development of explicit knowledge in an implicit learning situation. Afterwards, two prominent theoretical perspectives on consciousness will be introduced. These are the Global Workspace Theory (Baars, 1997; Baars & Franklin, 2003; Dehaene & Changeux, 2011; Dehaene & Naccache, 2001) and the Higher Order Thought Theory (Dienes & Perner, 1999; Lau & Rosenthal, 2011; Rosenthal, 2012; see Deheane et al., 2017, or Shea & Frith, 2019, for an argument why both, the Global Workspace Theory and the Higher-Order Thought Theory, are important for consciousness studies). We will elaborate how both theoretical views have already been applied in the field of implicit learning and how they fit the current theoretical views and empirical data on the transition from implicit to explicit knowledge. We will try to illustrate which problems both perspectives have, when they are used to explain how explicit knowledge arises from an implicit learning situation. Lastly, we aim to discuss in which direction future research could go in order to tackle these issues.

Theoretical Views on the Development of Conscious Knowledge in Implicit Learning Situations

As briefly stated in the introduction, there are two distinct theoretical perspectives on how explicit knowledge can develop in an implicit learning situation. One proposes that there is a continuous transition from an unconscious to a conscious state, while the other assumes that knowledge is either implicit or explicit and that the transition happens in a sudden representational change.

The idea of a gradual change in representational quality goes back to Cleeremans and Jiménez (2002). They proposed three different factors which influence the quality of a representation:

(1) Stability, i.e., the time a certain activation pattern can be maintained, (2) strength, i.e., the number of modules involved and their respective activation strength, and (3) distinctiveness, i.e., the extent of overlap between representations within a functional network (see Kinsbourne, 1996, for a similar position). While implicit learning first leads to very weak representations, with practice these representations gradually gain quality and can result in explicit knowledge. This proposal has later been elaborated further in the Radical Plasticity Hypothesis by Cleeremans (2008, 2011, 2014) by adding a hierarchically higher, second-order learning system. The lower-order or first-order learning system develops implicit knowledge through interaction with the environment. This knowledge is never conscious; it is labeled as knowledge within the system. For consciousness to arise, the first-order information needs to be redescribed as a meta-representation; that is, knowledge for the system (Clark & Karmiloff-Smith, 1993). The first-order representation itself becomes an object of a representation for higher-order systems. This higher-order system receives input from the first-order systems and learns that the state of the first-order system has changed as it becomes more accurate and thereby develops a higher-order attitude towards the first-order knowledge (e.g. "knowing that …", "hoping that …", "believing that …"). This higher-order representation is assumed to be a new representation involving a broad pattern of activation over different processing units which is only indirectly shaped by the changes of the connection-weights within the first-order system. The proposed learning mechanism behind the first- and the higher order learning system is the same; both systems gradually improve the quality of a representation with each learning trial. Thus, the higher-order representation gradually becomes more consciously accessible, as it changes from "not knowing" to "knowing".

The other theoretical account, the Unexpected-Event Hypothesis, agrees that it is not the implicit knowledge that becomes conscious itself. Rather, a second learning mechanism leads to the acquisition of explicitly accessible knowledge. However, this second mechanism does not lead to a gradual development of conscious access. The crucial idea of the Unexpected Event Theory is that explicit sequence knowledge can only develop when an individual unexpectedly notices a change in their own behavior. In an implicit learning situation, the interaction with the task leads to a continuous improvement of the responses to the stimuli; they become more accurate and faster. It can be this improvement or, for example, the feeling that the task becomes more fluent or easy, that there is a certain rhythm in one's own responses, or an external event that can trigger an intentional search for the sequence. Generally speaking, the Unexpected Event Theory involves a monitoring process which constantly compares observable expected and actual behavior. This comprises internal, subjectively experienced, as well as externally observable behavioral deviations from one's expectations.

For example, giving the (correct) response before the stimulus was shown could clearly be a surprising event for the participant. Haider and Frensch (2009) varied the RSI to manipulate the opportunity for participants to emit a response before the next stimulus occurred. They found more explicit sequence knowledge when the task allowed participants to emit premature responses. Likewise, Rünger and Frensch (2008) demonstrated that exchanging the trained sequence with a new sequence led to more explicit knowledge, than the same amount of training with only one sequence. The authors attributed this effect to the participants noticing surprisingly slower responses and attributing them to a change in the structure of the task. Controlling for the number of sequential and non-sequential trials, Esser and Haider (2017) showed that arranging regular, sequential and irregular, random trials in mini-blocks led to more explicit sequence knowledge, compared to a condition where regular and irregular trials were mixed randomly. In this case, the authors attributed this effect to providing the participants the opportunity to experience systematic differences in the fluency of the task.

Under the Unexpected Event Hypothesis, these consciously perceivable differences in the expected and observed behavior are critical for recognizing that the own knowledge is different from the expected knowledge. This does not imply that the individual instantly also gains insight into the exact sequence structure. Instead, the detected conflict triggers an attributional process in order to adjust its predictions and reestablish coherence between the distal environment and one's proximal model of it. Comparable monitoring-models have been established in neurocognitive models of conflict-detection and adaption (Botvinick, 2007; Botvinick et al., 2001), metacognitive control (Koriat, 2000, 2012, 2015), or memory (Whittlesea, 2002; Whittlesea & Williams, 2000). Thus, the triggered attributional processes do not necessarily lead to explicit sequence knowledge, if another explanation seems more likely to account for the unexpected experience (see also Reisenzein et al., 2017). Haider and Frensch (2005) have shown that premature responses in an implicit learning task did not result in more explicit knowledge if another, simpler explanation for these responses was provided (e.g. "attentional lapses"). Importantly, the secondary explicit learning process is not assumed to directly have access to the implicit knowledge. Instead, the explicit learning process will result in a new representation based on the available information. In this context, Schwager et al. (2012) have shown that after perceiving an unexpected event, participants will learn a new sequence just as well as the sequence they were trained with.

A suggestion that comprises aspects of both perspectives, the Radical Plasticity and the Unexpected Event Hypothesis, has been made by Scott and Dienes (2008; 2010). They suggested that implicit learning influences familiarity judgements, which enable individuals to correctly distinguish strings that follow the implicitly learned structure from those that do not. With repeated familiarity judgements, the individual will learn that their judgements are correct and thus explicit judgmental knowledge (knowing that one knows) develops gradually. However, in order to acquire explicit structural knowledge (knowing the exact structure of the implicitly learned information), a second explicit learning mechanism is required that tests different hypothesis about the reason for the correct judgements. When and how this explicit learning mechanism is triggered remains unclear.

In the following sections, we briefly describe the main aspects of the Global Workspace Theory and The Higher-Order Thought Theory as well as their relation to both, the Radical Plasticity and the Unexpected Event Hypothesis.

Global Workspace Theory

The Global Workspace Theory is a prominent functional and neuroscientific theory of consciousness. The basic assumption of the Global Workspace Theory is that the brain contains a multitude of functionally highly specialized subsystems working in parallel. Information in these subsystems is unconscious, there is no phenomenal- (Chalmers, 1995; Block; 2007), micro-consciousness (Lamme, 2006), or anything alike associated with information processing in these networks. Per se, these networks work encapsulated, that means they exchange information only within hard-wired or acquired pathways to fulfill their specialized task. This encapsulation enables the brain to handle a massive amount of input in parallel (Baars, 1997). Nevertheless, coherent interaction with the environment requires serial output and therefore a mechanism is needed that selects the most relevant information. Here, the theory postulates a global workspace mechanism which provides the necessary infrastructure, neurologically mainly realized by thalamo-cortical long-distance neurons of the prefrontal and the anterior cingular cortices (see Baars et al.,2013 for a detailed elaboration of the neuronal architecture). The global workspace is able to select relevant information, to prevent interference, to allow the encapsulated modules to exchange information, and to flexibly establish temporary networks between these modules (Dehaene & Naccache, 2001).

The Global Workspace Theory uses a blackboard metaphor for describing how the global workspace works. When a module gets selected to enter the global workspace, it can broadcast its content to any other network in the brain. Other modules can access the information on the blackboard and process it in their specified function. The information from the broadcasted module is no longer encapsulated. It is now said to be amodal, because it is no longer bound to the specialized processes of the module it originated from. Instead, it is now processed in a broad context of unconscious subsystems. These subsystems include, for example, perception, language, intentions, self-concepts, expectations, memory, and exclusive access to working-memory functions (Baars, 1997, 2005; Baars et al., 2013; Baars & Franklin, 2003; Cowan, 2010; Persuh et al.,2018; Schwager & Hagendorf, 2009). Thus, the Global Workspace Theory avoids the assumption that there is a certain instance where consciousness is “created” or that consciousness is an additional phenomenal quality that accompanies certain processes. Instead, consciousness is functionally defined by the global accessibility of representations and the resulting enabling of behavioral consequences.

Crucial to the Global Workspace Theory as a functionalist theory of consciousness is that conscious processing is equalized with the global accessibility of information and the thereby enabled options of processing this information. Neuroimaging shows that this de-capsulation of information is accompanied by a neurological “ignition”, a sudden, strong activation of a vast variety of cortical and subcortical regions (Dehaene & Changeux, 2011; Dehaene & Naccache, 2001; Del Cul et al., 2009; Rose et al., 2010; Schuck et al., 2015; Wessel et al., 2012). Hence, in the Global Workspace Theory the transition from unconscious to conscious processing is seen as an all-or-none phenomenon; there is no gradual consciousness. What may differ, however, is the level of representation that gains access to the global workspace. The levels of representation can vary from low (e.g. simple shapes or levels of intensity) to high (e.g. whole object, words, meaning), which all can be accessed independently and thus account for varying levels of the quality of conscious representations (Kouider et al., 2010).

To explain how a representation can change from an unconscious to a conscious state, while at the same time avoiding to assume that some instance “knows” (i.e., a homunculus) which information to select, the Global Workspace Theory suggests a stochastic bottom-up variation-selection mechanism for explaining how the most relevant information is selected from the enormous amount of unconscious information (“Neural Darwinism”, Changeux & Dehaene, 1989). Every unconscious module constantly competes for access to the global workspace (variation component), while the global workspace sets a selection function depending on current goal states. Only one module or coalition of modules will show the strongest activation in the context of the current goal-state-dependent content of the global workspace and will therefore win the competition for global broadcasting (“winner-takes-it all”, Shanahan & Baars, 2005). If a bottom-up signal surpasses a certain threshold, it is assumed to receive top-down amplification to remain maintained (it is said to receive "attentional amplification"). Thus, Dehaene and colleagues (2006) propose a 2x2 taxonomy in which an unconscious representation can have (a) strong or weak signal strength and, independent from signal strength, (b) can or cannot be amplified by top-down attention. Only when both criteria are given, signal strength is high and top-down attention towards the unconscious content is provided, the sufficient conditions for consciousness are met.

Global Workspace Theory and the Emergence of Conscious Knowledge in Implicit Learning

What are the implications of the Global Workspace Theory for implicit learning research and the explanation how explicit knowledge develops in an implicit learning situation? A first prediction is that the change from an unconscious to a conscious state happens in a sudden "insight" that a sequence has been learned, instead of a gradual change towards consciously accessible knowledge (Marti & Dehaene, 2017). There is some empirical evidence that in fact the transition from implicit to explicit knowledge is reflected in a sudden representational change. These studies aim to examine the point in time where an individual becomes able to verbalize their acquired knowledge or use it strategically. For example, Haider et al. (2011) provided evidence that most of the participants who showed a sudden drop in their RT during learning were able to verbalize their knowledge by the end of training. The RT-drop seemingly reflects the moment where participants switched from stimulus- to plan-driven control (Tubau et al., 2007). Moreover, neuroimaging data showed that a sudden coupling of gamma-band activity and increases of the BOLD-signal in the ventrolateral prefrontal cortex, the medial and ventrolateral prefrontal cortex and the ventral striatum preceded such an RT-drop, respectively strategy change (Lawson et al., 2017, Rose et al., 2010; Schuck et al., 2015; Wessel et al., 2012). These changes might reflect the sudden "ignition" of cortical activity which, as postulated by the Global Workspace Theory, accompanies the transition from an unconscious to a conscious state (Dehaene & Changeux, 2011; Dehaene & Naccache, 2001).

While such results implicate that conscious insight into implicitly learned representations seem to happen rather sudden instead of gradually, they do not provide information about how and why these transitions occur. A non-linear transition can occur due to an underlying slow gradual learning processes as well as through a spontaneously triggered, inferential explicit learning process. In order to explain how an unconscious, implicit representation can become consciously accessible under the Global Workspace Theory, two essential aspects need to be clarified: First, it needs to be explained how implicit, encapsulated information can reach a representational strength high enough to win the competition for access to the global workspace. Second, it needs to be explained how a goal state arises that provides the necessary top-down amplification for the implicit information.

Certainly, with ongoing practice in an SRTT, the representational strength or quality will gradually increase. The Global Workspace Theory allows the proposal that with enough practice, the representational strength or quality could be high enough by itself to win the competition and enter the global workspace (Cleeremans & Jiménez, 2002). This is what happens, when a signal with very high bottom-up strength is presented (e.g. a loud noise). However, it seems rather unlikely that an implicit learning process results in a signal that is strong enough by itself to win this competition without any additional top-down amplification. Likewise, a gradual higher-order learning process, as it is suggested by Cleeremans (2008, 2011), would, in the light of the Global Workspace Theory, require an additional explanation when and how the higher-order learning process changes from a gradual increase in representational quality to a non-linear increase in activation that corresponds with the sudden ignition proposed by the Global Workspace Theory. Rather, it seems that an explanation is needed how the system gets into a state in which the encapsulated module containing implicit information provides the fitting information to the selection function set by the global workspace and thus will receive additional top-down amplification.

Here, the Unexpected Event Theory provides a simple explanation: A monitoring process can detect a mismatch between expected and experienced performance. Because the mismatch is subjectively perceivable in one's own behavioral output or in internally perceived aspects of the task, this mismatch results in a new state of the global workspace that sets a fitting selective function for the implicitly learned representation.

Whatever the mechanism is that triggers the global workspace to allocate top-down amplification to the implicitly learned representation, it is furthermore important to ask, whether it is the implicit representation itself that will become a conscious representation or whether a new explicit representation will develop.

The Unexpected Event Theory suggests that the latter is the case; a detected mismatch only triggers a conscious attributional process with the purpose of finding any explanation for this mismatch. If an explanation different from an underlying sequence is more likely to the participant, the (sequence) knowledge remains implicit (Haider & Frensch, 2005; Wilbert & Haider, 2012). If, however, it seems likely to the participant that an underlying sequence is the reason for their behavior, a new, explicit learning process will learn the sequence, fully independent of the implicitly learned representation (see Schwager et al., 2012).

To sum up, when the Global Workspace Theory should be applied to explaining a transition from implicit to explicit knowledge two questions need more investigation in the future: First, how does the global workspace get into a state that can provide top-down amplification to the implicitly learned representation? Second, is it the implicit information that becomes conscious itself once it is selected or does the content of the global workspace only mobilize a second, explicit learning process?

Higher-Order Thought Theory

While the Global Workspace Theory is a specific theory of consciousness, Higher-Order Thought Theories are an umbrella term for a wider range of theories, which are concerned with the metacognitive aspects of consciousness (Lau & Rosenthal, 2011). Here, we focus on a Higher-Order Thought Theory that goes back to the work of Rosenthal (1997; Dienes & Perner, 1999). In its core, it differentiates between first-order and second-order (or higher-order) states. First-order states refer to simple input-output rules of any sensory or motor system. This can be understood in analogy to the parallel working modules in the Global Workspace Theory. Encapsulated, respectively implicitly learned information can be seen as a first-order state which is unconscious. Not only the human brain, but any simple or complex machine, which shows discriminatory performance, has first-order states (e.g. perceiving light of a certain wavelength results in the output of detecting red).

Consciousness, according to Higher-Order Thought Theories, crucially depends on developing higher-order knowledge about this first-order knowledge. Put simply, consciousness means knowing that one knows. This comprises the ability for self-reflection, self-reference and a propositional attitude (e.g. “I know/believe/guess that it is red that I see”, “It is I, who sees red”, “it is red that I see”, Dienes & Perner, 1999). What is needed for consciousness is a mechanism that allows the brain to draw inferences about its own internal first-order states and about how these relate to states in the environment. Different theoretical suggestions and models have been put forward to describe the learning process behind the acquisition of higher-order knowledge about first-order states (Fleming & Daw, 2017; Lau, 2008; Lau & Rosenthal, 2011).

Higher-Order Thought Theory and the Emergence of Conscious Knowledge in Implicit Learning

The theoretical view put forward by Cleeremans (2008, 2011, 2014) clearly applies Higher-Order Thought Theories to the question how explicit knowledge develops in an implicit learning situation: Through interaction with the environment, a first-order representation is developed and is gradually improving in quality. The higher-order system receives input from the first-order systems and learns that the state in the first-order system has changed and thereby develops a higher-order attitude towards the first-order knowledge (e.g. "knowing that …", "hoping that …", "seeing that …"). This higher-order knowledge is not per se conscious, but can become conscious, once its representational quality is strong enough. A very valuable aspect of this theory is that it connects well established connectionist learning theories with the development of consciousness.

Pasquali et al. (2010) have investigated the relation between first-order sensitivity and higher-order awareness in simulations of neural networks within different paradigms (i.e. Blindsight, Iowa Gabling Task, and an AGL Task). These results supported the assumption that the higher-order representations gradually improve with the learning progress of the first-order system. Using the Post-Decision Wagering Task (Persaud et al., 2007), the authors showed that the higher-order networks gradually changed from classifying random answers as being correct to classifying only correct responses as being correct (by giving high wagers on correct responses).

There are debates how exactly the relation between first-order knowledge and a meta-cognitive learning mechanism should be modeled, with most of the suggested models being based on bottom-up signal-detection theories (Barrett et al.,2013; Fleming & Lau, 2014; Maniscalco & Lau, 2012, 2016). What these models have in common is the gradual development of higher-order knowledge. The conscious state of a representation changes from guessing, which represents being unconscious about a first-order representation, to knowing, which represents being conscious about a first-order representation (Dienes & Scott, 2005; Sandberg et al., 2010).

A rather simple higher-order learning mechanism, as proposed by Cleeremans (2014), might indeed provide an important basis for a cognitive system to determine what first order state it currently is in. The assumption that a meta-cognitive learning mechanism plays a significant role in gaining conscious insight into otherwise unconscious information processing is very promising, as it describes the brain's ability to learn not only about external information, but also about its internal states. However, there are a few open points that should be considered in future research.

The higher-order learning process informs the system that knowledge has been acquired, but knowing that one knows (instead of guessing) the correct response is not equal to knowing that there is an underlying sequence or even knowing what exact rules constitute this sequence (Scott & Dienes, 2010). It could be argued that under Higher-Order Thought Theories this aspect is less important than under the Global Workspace Theory, because consciousness is defined as possessing a higher order representation of the first-order contents. Further functional properties, such as being able to verbalize the sequence, or being able to flexibly transfer this knowledge to new, different situations are of less importance. However, the question remains: Even if consciousness relies on gradual metacognitive learning processes, how is that learning mechanism connected to explicit knowledge of the underlying sequence?

Another question to ask is if correctness of the first-order performance is the only or most important target of higher-order learning systems. Knowing that one knows might not only rely on assessing the correctness of the behavioral output. For example, noticing premature responses before the next stimulus occurs (Haider & Frensch, 2009), sudden changes in the sequential structure which lead to slower reaction times (Rünger & Frensch, 2008), or changes in the perceived fluency of the task performance (Esser & Haider, 2017) also could be the target of metacognitive learning processes. Thus, it is open to further investigation whether such a higher-order process would also be able to learn about different metacognitive judgements (e.g. fluency). The mechanism described by Cleeremans and colleagues (Pasquali et al., 2010) has so far only been tested in situations where a person is directly asked to evaluate the correctness of their responses. This leads to the question whether metacognitive learning only occurs for intentionally attended dimensions (e.g. correctness), or whether learning about first-order performance is automatic and can happen in parallel for multiple dimensions (e.g. correctness, fluency, speed, etc.) when there is no external instruction to do so (as there is by subsequently presenting the post-decision wagering task).

Research on implicit learning implies that implicit learning processes can occur in parallel (Goschke & Bolte, 2012; Haider et al.2012; Haider et al., 2014, Haider et al., 2018; Mayr, 1996). Yet, it is not granted that higher-order learning processes can happen in parallel for all implicit learning processes. It might be that higher-order learning processes rely on intention, respectively selective attention, to evaluate one specific behavioral output (correctness, speed, fluency, etc.). If this were the case, it needed to be explained how the system decides which first-order representations are accessed to develop a higher-order representation. Moreover, a large number of the implicit learning studies cited here, imply sensitivity for sudden changes, which so far have not been addressed by theories suggesting gradual metacognitive learning processes.

Therefore, a learning process involving expectations, predictions, and violations thereof should be considered rather than gradual associative strengthening. On an empirical side, this is supported by the above-mentioned studies, which used different manipulations for balancing the associative strength between conditions but manipulated whether small or large violations of expectations occurred. For example, Esser and Haider (2017) showed differences in the emergence of explicit knowledge when the structure of the task led to noticeable differences in the fluency of processing the task material. Noticeably, the number of regular and irregular sequential trials was equal for both groups. Therefore, it needs to be addressed how a metacognitive mechanism that gradually learns to evaluate first-order performance would detect the differences between both learning conditions, even though the first-order signal strength is matched. A gradual bottom-up higher-order learning mechanism does not include the size of prediction errors (here, the sudden changes in fluency) as a signal. In the following section, we will propose a tentative model which includes ideas of the Higher-Order Thought Theories and the Global Workspace Theory in order to respond to the formerly described problems.

Metacognitive Learning Mechanisms and Unexpected Events

We have reviewed two different views on the development of explicit knowledge in an implicit learning situation: The Unexpected Event Hypothesis (Frensch et al., 2003; Haider & Frensch, 2005) and the metacognitive Radical Plasticity Hypothesis from Cleeremans (2008, 2011). We have argued why the Unexpected Event Theory makes assumptions and provides empirical evidence that fits with a Global Workspace Theory of consciousness, while the Radical Plasticity account of Cleeremans theoretically and empirically fits well with Higher-Order Thought Theories (see Figure 1).

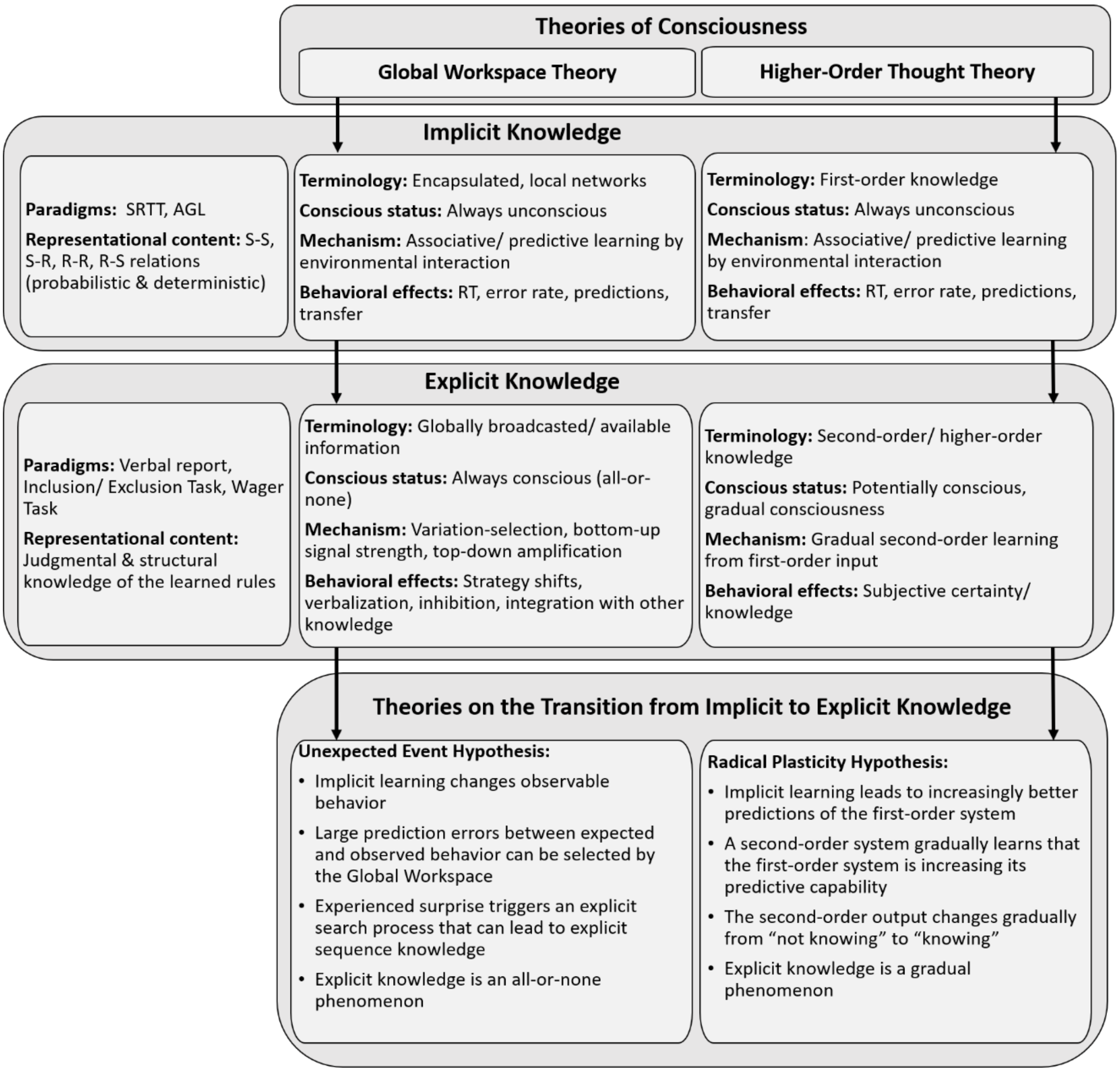

Fig. 1. Implicit and Explicit Learning Viewed under the Global Workspace and the Higher-Order Thought Theory

First, both theoretical viewpoints do not differ in the conceptualization of the process of implicit learning itself. Implicit learning is viewed as a first-order learning mechanism (as it might be called under a Higher-Order Thought Theory viewpoint), that creates localized, encapsulated representations (as the Global Workspace Theory would put it). These first-order learning processes can be described as internal perception-action loops that use prediction errors to enable learning. In such models, learning of actions or action-sequences is controlled by an interaction of feedback and feedforward loops (McNamee, & Wolpert, 2019; Wolpert & Ghahramani, 2000). A forward model predicts, given a specific motor command, sensory or proprioceptive consequences of an action. The predictive forward model can be trained by comparing predicted and actual sensory feedback and using the resulting error signal to make increasingly more accurate predictions. It has been demonstrated that such internal perception-action loops are relevant for implicit motor and perceptual sequence learning (Janacsek et al., 2020; Lutz et al., 2018; Ruttle et al., 2021; Ziessler & Nattkemper, 2001).

Both theoretical viewpoints, the Unexpected Event Hypothesis and the Radical Plasticity Hypothesis state that it is not these implicit internal models themselves that are consciously accessible. Instead, both theoretical approaches raise the question how a secondary explicit or higher-order learning mechanism represents that the contents of the first-order models have changed and, finally, how consciousness about unconsciously learned sequences arises.

Both viewpoints have their strengths and weaknesses. The metacognitive Radical Plasticity account has the strength of explaining conscious access by clearly definable connectionist learning mechanisms that rely on the same predictive learning principles as the first-order system does. However, this does not readily explain how structural explicit knowledge develops (knowing what exactly the sequence is; Dienes & Scott, 2005). Furthermore, it does not account for the role of expectancy violations from several distinct metacognitive sources (accuracy, speed, fluency, conflict, etc.).

The seemingly biggest difference between the Radical Plasticity account and the Unexpected Event Hypothesis is that the former assumes that consciousness develops gradually along with lowering the prediction error of the second-order system; when there is no surprise about the ongoings in the first-order system left, one knows that one knows. The Unexpected Event Hypothesis instead proposes that conscious awareness is triggered by large prediction errors; when an individual thought they did not know but apparently know. The Unexpected Event Theory captures the important aspect of violations of consciously accessible expectations and explains why a rather sudden insight (comprising the time span from a consciously accessible surprise to explicit structural knowledge) into the implicitly learned representations seems to develop. Its weakness is that, so far, it did not take metacognitive learning mechanisms into account that could provide a clearer prediction when expectancies will be violated or how these expectancies arise in the first place.

Thus, we aim to elaborate the processes behind the original proposal of the Unexpected Event Theory and to point to open questions which should be addressed by future research. Even though the Unexpected Event Hypothesis generally follows the Global Workspace Theory, we assume that higher-order learning is an important mechanism to consider, when trying to explain how an unexpected event can become consciously accessible. What is needed is a mechanism which allows a comparison between the expected metacognition (e.g., "How correct, fast, fluent should my response be?") an individual has in a given situation and the experienced metacognition ("How correct, fast, fluent was my response?"). Important questions in this regard are: What is the relation between the first-order internal perception-action models and the second-order or metacognitive model? How could a metacognitive learning process explain how a consciously accessible surprise occurs? Currently, there are several different models aiming to explain the relation between the first-order signal (here, implicit knowledge) and the second-order metacognitive evaluation of these signals. Mostly, these models rely on signal-detection theory (Galvin et al., 2003; Lau & Rosenthal, 2011; Del Cul et al., 2009). The problem with these models is that they are often pure bottom-up models that do not take important top-down factors into account which have shown to influence metacognitive decisions. This includes, for example, the use of heuristic cues (e.g. fluency, luminance), which have no direct relation to the first-order signal the metacognitive judgement is relating to (Hoyndorf & Haider, 2009; Koriat, 2007; Wilbert & Haider, 2012). It further includes the role of previous metacognitive experiences, with similar situations, successes and failures, or general knowledge about one's own performance capacities.

There are, however, theories that model the relationship between first-order knowledge and metacognitive judgements with Bayesian learning (Fleming & Daw, 2017; Sherman et al.,2015). One advantage is that Bayesian models allow metacognitive learning via predictive coding (Clark, 2013; Friston, 2010). The evaluation of one's own behavior, respectively knowledge leads to a first hypothesis of what metacognitive experience is expected in the next, similar situation. This prediction is compared with the current experienced metacognitive judgement and, in turn, the resulting error-signal is used as a bottom-up learning signal for the next, more precise metacognitive prediction.

For implicit learning and the development of explicit knowledge, this means that any individual has a certain expectation about their own performance in an SRTT, based on previous experiences with similar situations. Thus, in the beginning an individual has a certain metacognitive model how fast, fluent, correct, etc. their behavior should be when responding to stimuli that appear seemingly randomly on a computer screen. The sequential material inherent in the task can lead to behavior different from the expected behavior. These deviations from the expected performance will be used to adjust the metacognitive model. In order to develop explicit knowledge, the participant has to recognize that their performance does not match their expectations to an extent that is not compatible with their model of the current situation. We assume that what is important for a change of the metacognitive model and thereby the development of explicit knowledge, is the size of the metacognitive prediction error and the strength of the a-priori hypothesis.

Smaller deviations from the expected metacognitive judgement of the situation can easily be used to adjust the model via this bottom-up error signal. For example, faster responses, fewer errors, and increasing fluency are compatible with mere practice effects and only slight, gradual adjustments of the metacognitive models are the result. However, large prediction errors would more likely lead to a stronger change of the metacognitive model. In this case, it might be functional to evaluate whether a new, different model should be applied to the situation, instead of making rather drastic changes to the current model. The strength of the a-priori hypothesis could also play a significant role. If there is a very strong a-priori hypothesis and the current data strongly contradict this model, it might be less functional to make drastic changes to the current, well-established model. Instead, it might be more advisable to test whether a different model should be applied, respectively built for the new situation. During an SRTT-training, a large deviation might occur, if the well-practiced sequence is suddenly removed and replaced by random responses. In this case, it might be less functional to assume that the own performance capacity has declined and instead it might be useful to check whether the task has properties that were not considered before (e.g. that there used to be a sequence, which is now missing). If, however large metacognitive prediction errors are encountered while the participant only has a rather weak a-priori hypothesis, there is no need to replace the current metacognitive model with a new one. Instead, the participant will make adjustments to their expectations, without developing explicit sequence knowledge. Taken together, we propose that both factors, metacognitive prediction error and strength of the metacognitive a-priori hypothesis determine whether an individual will change their metacognitive model and thus enable explicit sequence learning.

As mentioned before, there are models for metacognitive learning that support our notion that metacognitive evaluations not only rely on current first-order performance signals, but also on earlier metacognitive judgements (e.g. Fleming & Daw, 2017). So far, there is not much research on how metacognitive models are selected in a given situation and under which circumstances a model is replaced with a new or different one or when instead the current model will be adjusted. Collins and Frank (2013) suggested a Bayesian “context-task set” model. In this model, an inference is made in every single learning trial about whether the current task-set is still applicable to the current situation or whether there are yet unknown rules that should influence the task-set and therefore, a new model should be applied. This model also uses arbitrary context cues to determine whether the current situation is indicating a new, unknown task context or whether previously acquired metacognitive models should be used and adjusted.

Importantly, with regard to the Global Workspace Theory, we do not assume that such metacognitive representations are per se conscious or that their strength has any relation to a gradual change in consciousness. Instead, we propose that metacognitive representations are generally good candidates for being accessed by the global workspace (see Shea & Frith, 2019, for an opinion why metacognitive learning is important to the Global Workspace Theory). Assuming that there are multiple parallel higher-order learning processes, their content might be entirely implicit as well. They send and receive information from the global workspace, just like implicit first-order information could. In the global workspace multiple meta-cognitive representations could be integrated into a hierarchically higher metacognitive representation of the current situation. This integrated information could for example involve metacognitive knowledge about one’s own performance like fluency, accuracy of processing, precision or confidence. In addition, different weights to the underlying metacognitive representations are assigned, according to their current relevance.

These original assumptions of the Unexpected Event Theory and the additional assumptions about a predictive metacognitive learning processes proposed here, could solve some of the formerly described problems behind the explanations based solely on the Global Workspace Theory or Higher-Order Thought Theories. Concerning the Global Workspace Theory, the Unexpected Event Theory does not need to explain how an implicitly acquired first-order representation can gain a signal-strength strong enough to win the competition against all other unconscious modules or how top-down amplification can be directed to this encapsulated knowledge. This problem is solved because it is the conflict between the expected and the experienced metacognitive judgements which gains access to the global workspace. It is the representation of this conflict (or surprise) which has a high likelihood of winning the competition against other parallel processes for entering the global workspace.

Concerning the Higher-Order Thought Theory based explanation of the emergence of explicit from implicit knowledge, the here proposed addition to the Unexpected Event Theory account lies in the assumption how implicit, first-order knowledge and higher-order knowledge are related. An account where metacognitive judgements depend on a predictive learning process which does not only base its predictions on the first-order bottom-up signal, but also on heuristic cues, previous knowledge, and experiences with similar situations, can help to explain different empirical findings. This includes, for example, premature responses (Haider & Frensch, 2009), changes in the underlying sequence (Schwager et al., 2012) and changes in the experienced fluency (Esser & Haider, 2017; Rünger & Frensch, 2008). All these results are difficult to explain with a pure bottom-up mechanism relying on gradual strengthening of the first-order learning signal. Furthermore, large prediction errors and the processes, they are assumed to trigger, fit the data suggesting that explicit knowledge seems to develop in a sudden moment of insight (Haider et al., 2011; Rose et al., 2010; Schwager et al. 2012; Wessel et al., 2012), rather than developing gradually.

Conclusion and Future Directions

We assume that the predictive metacognitive model a person has about their own behavior in a given situation (e.g. how fast, how precise, how difficult or fluent a task should be) adapts to the task by comparing the predicted and the experienced metacognitive judgement in any given situation. The behavioral changes resulting from implicit learning may not fit the current metacognitive model (i.e. responses might suddenly be much slower than expected when the sequence is exchanged with new, random material). If so, this violation has a high chance to enter the Global Workspace and serve as a trigger to evaluate whether a new metacognitive model of the situation should be applied.

Nevertheless, there are questions that should be addressed by future considerations. This relates to theoretical assumptions in need of further elaboration: Are all implicit learning processes continuously monitored by parallel metacognitive learning processes? How can previous knowledge and external cues, like task fluency, influence these metacognitive processes? Is metacognitive learning a pure bottom-up process? Does the metacognitive prediction error play a role in the emergence of explicit knowledge? Furthermore, there is need for a model that can account for current empirical findings: Why does explicit knowledge seem to emerge in a sudden insight process? Why do alternative explanations for the behavioral change prevent such insight?

We assume that the Unexpected Event Theory and the here proposed extensions about expectancy violations resulting from Bayesian metacognitive learning processes can provide a promising step into answering these questions. The prediction of a metacognitive judgement is compared to the currently experienced metacognitive judgement about one's own behavior. Its prediction-error signal is then used as a learning signal for developing a more accurate metacognitive model of the current situation. Small prediction errors might lead to a gradual change of the model. Yet, large prediction errors in combination with strong a-priori hypotheses can serve as a signal that the current model is not suitable for the given situation and a different model should be applied. Within such a framework, it can be modelled that not only the current bottom-up first-order signal but also top-down factors, such as heuristic cues and previous experiences with similar situations, are the basis for a prediction of the metacognitive judgement in a given situation.

Our proposal of integrating the role of metacognitive learning processes in the Unexpected Event Theory needs further experimental investigation: First, it should be tested whether the predicted metacognitive judgements can be manipulated not only by the strength of the first-order signal, but also by differences between the expected and the actual experienced metacognitive judgment. Second, the size of the prediction error of metacognitive judgements as well as the strength of the a-priori hypothesis should be manipulated to test its relation to the emergence of explicit knowledge. Third, we proposed that large prediction errors serve as a consciously accessible signal to trigger explicit search processes. These search processes are assumed to lead to a new explicit representation, independent of the implicit representation.

A better understanding of the transition from implicit to explicit sequence knowledge can provide interesting contributions to the broad and difficult field of consciousness theories itself. Implicit learning paradigms create the unique experimental situation where unconscious knowledge does not need to be induced by week signal strength or inattention. The development of metacognitive knowledge is a concern of many different and often separated research fields which all provide different contributions. For example, research on decision-making or on perception is governed by bottom-up signal-detection models (Galvin et al., 2003; Pleskac & Busemeyer, 2010), cue-utilization is prominent in memory research (Koriat, 2000, 2012, 2015) and models of evidence accumulation are often found in research on error-monitoring (Yeung & Summerfield, 2012). Implicit sequence learning paradigms can augment this research by providing additional opportunities (to the predominant priming paradigms) to manipulate the first-order signal strength, external cues, as well as the role of prior expectations and how these expectations develop over the course of learning.

6. Study 1: Response-Effects trigger the development of explicit knowledge

Abstract

In implicit learning, task-redundant response-effects can enhance the development of explicit knowledge. Here, we investigated whether learning a fixed sequence of effects (stimuli occurring immediately after the participant's keypress, but are not mapped to the identity of the respective response) influence the development of explicit rather than implicit knowledge when these effects are afterwards mapped to the identity of the responses. We tested first, whether participants would learn a fixed sequence of effects in a serial reaction time task when these effects were not mapped to the identity of the responses. Next, we tested whether learning this effect sequence in advance would facilitate the development of explicit knowledge about a contingently mapped sequence of responses.

The results showed that participants acquired implicit knowledge when confronted with only the effect sequence. Moreover, the further findings suggest that learning the effect sequence in advance led to the development of primarily explicit knowledge about a subsequently added response-location sequence. We interpret these results in light of the Unexpected-Event hypothesis: A sudden feeling of sense of agency is unexpected and triggers inference processes.

Introduction[1]

General Findings in implicit learning

Implicit learning is assumed to be one fundamental learning process enabling humans to exploit sequential structures inherent in the environment. It is assumed to take place without any intention or additional effort and even without the learner's conscious awareness about that they learn or what they actually learn.

One of the most frequently utilized paradigms in the field of implicit learning is the Serial Reaction Time Task (SRTT; Nissen & Bullemer, 1987). In this standard SRTT, participants see four horizontally marked locations on the screen, which are mapped to spatially corresponding keys. Participants are instructed to press the appropriate response key whenever an asterisk occurs at a certain location. Unbeknownst to the participants, the locations of the asterisk follow a regular sequence. After several blocks of practice, the regular sequence is replaced by a random sequence. This leads to a performance decrement that disappears almost immediately when the original sequence is reintroduced. At the same time, direct tests like, for instance, verbal report (Eriksen, 1960; Rünger & Frensch, 2010), or

[1] In the original manuscript, the headings are numbered. To avoid confusions, the numbering was taken within this thesis.

the generation tasks (Dienes & Seth, 2010; Haider, Eichler & Lange, 2011; Shanks & Johnston, 1999; Wilkinson & Shanks, 2004) suggest that the participants are not able to explicate their acquired knowledge when asked to do so. This latter finding, albeit hotly debated (e.g., Shanks & St. John, 1994), is interpreted as a hint that the participants are not aware of having acquired knowledge about the experienced sequence (see, Frensch & Rünger, 2003).

Beside other important results, several authors in this field have provided evidence that implicit learning is enhanced when response-effect contingencies are realized within the SRTT (Esser & Haider, 2017a; Haider, Eberhardt, Esser & Rose, 2014; Hoffmann, Stöcker & Sebald, 2001; Stöcker, Sebald & Hoffmann, 2003; Stöcker & Hoffmann, 2004; Tubau, Lopéz-Moliner & Hommel, 2007; Ziessler, 1998; Ziessler & Nattkemper, 2001; Zirngibl & Koch, 2002). However, the debate how and why response-effects may amplify implicit learning is still going on. Furthermore, some recent findings suggest that response-effects not only enhance implicit learning, but, in particular, foster the development of explicit knowledge (Esser & Haider, 2017a; Haider, Eberhardt, Kunde & Rose, 2012; Haider et al., 2014; Zirngibl & Koch, 2002). The goal of the current study was to further contribute to the question whether response-effect learning (R-E learning) affects implicit or explicit knowledge.

Response-Effect Learning in the Serial Reaction Time Task

Basically, response-effect learning[2] has its root in the field of voluntary actions and control. The central assumption, the ideomotor principle (Greenwald, 1970; James, 1890; Prinz, 1992, 1997; Shin, Proctor & Capaldi, 2010), states that voluntary actions are initiated by anticipating its action-effect (its perceivable change in the environment). Learning to know which effect in the environment is provoked by which action is thus crucial for voluntary actions (Kunde, 2001). In their two-stage model, Elsner and Hommel (2001) proposed that one first learns to associate actions with their effects. These action-effect relations can then be used bidirectional. This means that the activation of these effects, be it by intentionally

[2] For the purpose of our investigations, we understand the term response-effect as synonymously with action-effect. The term action-effect stems from the research on voluntary actions and control and refers to an action and its perceivable change in the environment (its action-effect; e.g. Elsner & Hommel, 2001; Greenwald, 1970; Hommel et al., 2001). In contrast, researchers in the implicit sequence-learning literature prefer the term response-effect (e.g. Esser & Haider, 2017a; Haider et al., 2014; Hoffmann et al., 2001). As our focus is on implicit learning, we primarily use the term response-effect here.

imagining the effect of an action or by externally presenting the effect, facilitates the action-selection process (Elsner & Hommel, 2001; Hommel, Müsseler, Aschersleben, & Prinz, 2001; Kunde, Schmidts, Wirth & Herbort, 2017).

In the field of implicit learning, Ziessler (1998) and Ziessler and Nattkemper (2001) were among the first who focused on the role of response-effect (R-E) learning. They realized response-effect contingencies as response-stimulus relations because in their studies the response always predicted the imperative stimulus of the next trial. In their study, Ziessler and Nattkemper (2001) varied the complexity of either the R-E (R-S in their terminology) or the S-S relations. Their results showed that the manipulation of the R-E complexity had a much larger impact on the latencies, and thus on learning, than the S-S complexity. In addition, when inspecting their results of their post-experimental knowledge tests the data suggests that R-E learning seems to not only amplify implicit learning but also increased the rate of explicit knowledge.

Slightly different, Hoffmann et al. (2001) as well as Stöcker, et al. (2003) introduced task-redundant tones as response-effects. Task-redundancy in this context means that the effect-tones themselves were not needed to perform the task. In their study, the participants responded to the location of an asterisk just as in the Nissen and Bullemer task. Subsequently, they heard one of four different tones. In the experimental condition, each response key was mapped to one single tone. In two further control conditions, the participants either did not hear any tone or the tones were not contingently mapped to the response keys (i.e. each of the four tones was mapped to the four asterisk locations depending on their serial position). Their results revealed that the effect tones facilitated implicit learning when they were contingently mapped to the responses, but not when the mapping between the response locations and the tones was arbitrary. Furthermore, the data from the knowledge test in their study also suggest that the contingent response-tone mapping led to more explicit sequence knowledge.

Recently, Esser and Haider (2017a) built on these findings. In their study, the participants responded in three different conditions to the same response sequence and also heard the same sequence of task-redundant effect tones. The crucial manipulation concerned the response-tone contingencies. The tones were either consistently mapped to the responses (Contingent-Tone condition) or consistently mapped to only the repeating colors of the stimuli (No-Contingent-Tone condition). Through the use of additional trials, the response and the color sequence were uncorrelated. Since this difference between the two conditions affects, as a side effect, the response-tone-stimulus contingencies, the authors introduced a third condition.

Basically, this condition was identical to the Contingent-Tone condition. The only difference was that, here, the tones sounded always concurrently with the appearance of the stimuli (Stimulus-Tone condition) rather than immediately after the participant's responses. The results of this study showed that the participants in the Contingent-Tone condition had significantly more knowledge than the participants in the Stimulus-Tone condition, whereas the No-Contingent Tone condition and the Stimulus-Tone condition did not differ. This difference in the amount of knowledge almost entirely vanished after removing all participants with predominately explicit knowledge from the analyses. Thus, the findings clearly speak for an effect of response-effect learning on the generation of explicit knowledge within the SRTT (for similar findings see also Haider et al., 2012, 2014; Zirngibl & Koch, 2002).

Taken together, the reported results converge to two main findings regarding R-E learning: First, implicit R-E learning seems to progress faster than other forms of implicit learning, like, for instance stimulus-stimulus or stimulus-response learning. Second, R-E learning seems to foster the development of explicit knowledge. To account for these findings, at least, three different explanations within the field of implicit learning have been proposed. As we will show below, the accounts mainly differ in their assumptions whether R-E learning enhances implicit or explicit learning processes.

Assumptions about the development of explicit knowledge in implicit R-E-Learning

A very famous model in the field of implicit learning is the Dual-System model proposed by Keele, Ivry, Mayr, Hazeltine, and Heuer (2003). The central assumption of this model is that implicit learning can rely on either the unidimensional or the multidimensional system. The former system comes into play when the to-be-learned sequence refers to only one single dimension, like, for instance, a sequence of response locations. The information of the single dimension is assumed to be processed in encapsulated modules, which are not accessible for the information processing of other dimensions. Furthermore, learning is thought to be independent of selective attention and leads to implicit knowledge because the single dimension is already sufficient to select the information. If, however, the sequence requires integrating information across different dimensions, learning is assumed to be based on the multidimensional system. According to Keele et al., only in the multidimensional system learning needs selective attention and also can contribute to the development of explicit knowledge. Thus, a plausible explanation for the increased rate of explicit knowledge in experiments based on R-E learning is that, in contrast to, for instance, a response location sequence, learning of an R-E sequence requires integrating the response location information and the perceptual information of the effects. Probably therefore, the learning process underlying R-E learning relies on the multidimensional rather than on the unidimensional system (Abrahamse, Jimenéz, Vervey & Clegg, 2010). However, an important objection concerning this assumption is that some empirical results suggest that the increased rate of explicit knowledge is specific for R-E learning. It is not found when only the imperative stimulus is enriched by other redundant information (Abrahamse, van der Lubbe, Verwey, Szumska, & Jaskowski, 2012; Esser & Haider, 2017a).

A second explanation, proposed by, for instance, Stöcker and Hoffmann (2004) is that response-effects facilitate the chunking processes (see, Hoffmann & Koch, 1997; Koch & Hoffmann, 2000). Based on the ideomotor principle (Greenwald, 1970; Prinz, 1992), these authors assume that the response selection process in R-E learning is facilitated. This, in turn, amplifies the chaining of succeeding responses and their effects resulting in longer subsequences (or chunks) of key-presses.

This assumption is consistent with the finding of increased implicit learning effects. However, it does not explain why the response-effects seem to boost primarily the development of explicit knowledge. To additionally account for the increased rate of explicit knowledge (see, Zirngibl & Koch, 2002), it needs the extra assumption that the longer the chunks the higher the probability that these chunks are explicitly represented. This, however, might be arguable given the empirical findings in the field of skill acquisition. These findings suggest that over the course of practice chunking occurs. However, the accessibility of these acquired chunks underlying such skills decreases rather than increases (Anderson, 1993; Logan, 1988; Newell & Rosenbloom, 1981). One solution to this problem is the assumption that the development of explicit knowledge is characterized by a shift from stimulus-based to plan-based action control (e.g. Tubau et al., 2007). Furthermore, for instance, Tubau et al. (2007) propose that phonetically coded stimuli as such enable plan-based action control. For phonetically coded response-effects, like, for instance, auditory effects (e.g., Esser & Haider, 2017a; Hoffmann et al., 2001), or symbolic visual effects (e.g. letters; Ziessler & Nattkemper, 2001), this account might explain why R-E learning fosters plan-based performance, hereby leading to a higher rate of explicit knowledge. However, as a necessary precondition for such a plan-based performance the underlying sequence needs to be integrated in its entirety (e.g., Haider et al. 2011; Rose, Haider & Büchel, 2010).

The third explanation is based on the Unexpected Event Hypothesis (UEH; Frensch et al., 2003). The UEH deviates from the former explanations in that it is based on the premise that implicit and explicit knowledge crucially differ with respect to their representational format. Like Keele et al. (2003), implicitly acquired knowledge is thought to be represented in encapsulated modules of the unidimensional system (Dehaene & Naccache, 2001; Eberhardt, Esser & Haider, 2017). Due to this encapsulated nature of such representations this knowledge is not consciously accessible.

The crucial difference between the UEH and the Dual-System model of Keele concerns the way how such implicitly acquired knowledge can become conscious. According to the UEH, the multidimensional nature of an R-E sequence is not sufficient. Rather, it is assumed that an independent explicit inference mechanism is needed. It is, for instance, triggered when a person recognizes an unexpected event; when, for example, in an implicit learning experiment the participant realizes that she responded correctly without already having seen the stimulus (e.g., Haider & Frensch, 2005, 2009). This, in turn, activates explicit inference processes to figure out why it should be possible to produce such premature responses in the given situation. The result of this inference process is it what then becomes consciously represented (Esser & Haider, 2017b; Scott & Dienes, 2008; Whittlesea, 2002, 2004).

According to this account, R-E learning may not alter the implicit learning processes that much. Rather it may primarily foster the development of explicit knowledge (Esser & Haider, 2017a; Haider et al., 2012, 2014). By definition, R-E learning leads to observable changes in the environment as a consequence of one’s own actions (Elsner & Hommel, 2001; Hommel et al., 2001; Nattkemper, Ziessler & Frensch, 2010). On the one hand, this should increase the likelihood that participants encounter an unexpected event during the course of learning. In R-E learning not only the observation of the characteristics of one’s own performance could violate the participants’ expectations, but also the produced changes in the environment. For instance, the participants may not only notice that their performance feels more fluent than expected, but also that they have a hunch what the next response-effect will be already before it occurs. On the other hand, it is also conceivable that the effects of one’s own actions, because they appear as consciously perceivable changes in the environment, are rather salient and as such direct the explicit inference processes towards the sequence built into the task.

Taken together, the aforementioned empirical findings show that implicit R-E learning leads to stronger learning effects and often also to more explicit knowledge about the underlying sequence than other implicit learning processes. Even though these findings seem to be rather robust, they are based on only a few studies. Furthermore, no conclusive explanation exists yet. While the first explanation mainly addresses the multidimensional characteristic of R-E learning, the latter two explanations build on the ideomotor principle of intentional action and control. In addition, the three explanations differ with regard to one core assumption of implicit learning namely that practicing a sequence strengthens the associations between the elements of the sequence. This then causes larger implicit learning effects on the one hand and more explicit knowledge on the other hand. In particular, the second explanation seems to be rooted in this tradition of thinking. By contrast, the UEH explicitly states that implicit knowledge is never consciously accessible. Rather, the explicit insight that a sequence is built into the implicit learning task needs the generation of new consciously accessible representations. Such representations result from additional explicit search processes which are triggered by the experience of an unexpected event.

Goal of the Current Study

The current study aimed at investigating the assumption derived from the UEH that response-effect contingencies may boost the development of explicit knowledge within an implicit learning situation. In extending the study of Esser and Haider (2017a), we focused more on the role of unexpected events in the context of R-E learning.

Several findings from our lab (Eberhardt et al., 2017; Haider et al., 2012, 2014; Haider, Esser & Eberhardt, 2018) suggest that participants are able to implicitly learn sequences of events even when these are not mapped to responses like, for instance, a visual sequence of colors or a visual-spatial sequence presented only in the stimuli. Therefore, we assumed that participants may also learn a fixed sequence of effects when these effects are not bound to the identity of the responses; that is, when the effects follow a sequence (e.g. A-B-C-D-E-F), but the responses producing them are random. We use the term effect sequence because pressing a response key consistently produced an event (the effect: tones in Experiment 1 and visual shapes in Experiment 2).

On the empirical side, the findings of Lepper, Massen and Prinz (2008) already support that learning of such an effect sequence is indeed possible. They showed that participants are able to separately learn a sequence of either transformation rules (S-R-E rules) or of effects that were not consistently mapped to the responses. When replacing the fixed effect sequence by a random sequence, participants showed a small, but significant deceleration of 12ms in Experiment 1 and 28ms in Experiment 2.

Thus, if participants were able to learn such a fixed sequence of effects without knowing how to produce them, they should experience a surprising sense of agency when, without further instruction, the responses are later mapped to these effects (Beck, Di Costa & Haggard, 2017; Haggard, Clark & Kalogeras, 2002; Moore, Lagnado, Deal & Haggard, 2009). This unexpected feeling of a sense of agency then should, according to the UEH, trigger explicit inference processes and as such should boost the development of explicit knowledge about the response-location sequence.

In order to test these assumptions, we first investigated in Experiment 1a whether participants would implicitly learn a fixed sequence of task-redundant effects without knowing how to produce them because these effects were not contingently mapped to the identity of either the stimuli or the responses. The only additional contingency here was the timing between the responses and the effects. Whenever the participant had pressed a key, the effect appeared immediately. In order to assess implicit learning, the fixed effect sequence was replaced in the second to last block and was reintroduced in the last block (Lepper et al., 2008).

To foreshadow our findings of Experiment 1a, the participants indeed learned the fixed sequence of effects. This allowed us to test our main question if experiencing such an effect sequence in advance will enhance the acquisition of explicit knowledge about a response-location sequence. Accordingly, in Experiment 1b, we trained the participants either with or without a fixed auditory effect sequence in the first half of the training. The presentation of this effect sequence was the only difference between the two conditions. Without any further instruction, all participants were then transferred to a test phase. Here, they received the fixed effect sequence together with a correlated response-location sequence, meaning that the participants now could learn how to generate the response-effects. Immediately after this test-phase, the participants' knowledge about the response-location sequence was assessed. In order to additionally test whether learning a sequence in advance generally facilitates the learning of a response-location sequence, we introduced a third condition. Instead of presenting a fixed effect sequence, the sequence here was built into the imperative visual color stimuli in the first half of the training. In the second half, this stimulus sequence was then mapped contingently to the response locations.

If our considerations were true, we should find more explicit knowledge in only that condition in which the participants already received the fixed effect sequence in the first phase. The goal of Experiments 2a and 2b then was to replicate these findings with visual (response-) effects instead of auditory effects in order to test whether these findings are specific for the auditory modality.

Experiment 1a

As already said, the goal of the first experiment was to test whether the participants would learn a fixed auditory effect sequence, even when (a) they were not informed about the existence of such a sequence, (b) the effects were not contingently mapped to the responses or stimuli, and (c) the effect tones were redundant for fulfilling the task. The participants were only told that they will hear a tone after each response.

Implicit learning of the effect sequence was assessed by replacing the sequence in Block 5 by a pseudo-random sequence and reintroducing it in Block 6. Based on the findings of Lepper et al. (2008), we expected that the participants will implicitly learn such an effect sequence even though they do not know how to produce it. If the participants interpret the effects as response-effects, they should anticipate them which in turn should facilitate the initiation of the responses (Kunde, Koch & Hoffmann, 2004; Pfordresher, 2003, 2005). Thus, removing the effect sequence should lead to an increase of latencies.

Methods

Participants

Forty students (4 men) of the University of Cologne with a mean age of 22.40 years (*SD*=2.77) participated the experiment either for course credit or for payment (3€).

In all experiments, we replaced the data of participants when, first, we could not ensure that they already had taken part in former SRTT-experiments or, second, when their mean error rate was higher than 15% errors over all blocks. In Experiment 1a, two participants reached this error criterion.

Apparatus and Stimuli

We used the version of the SRTT, originated by Haider et al. (2012). In this version, a target square (3 x 3 cm) appeared in the upper third of the 17-inch screen. It contained one of six colors (green, yellow, cyan, pink, blue, or red). In the lower two thirds of the screen, six response squares (3 x 3cm) were presented in a triangular space. They contained the six possible target colors. The arrangement of the six colors of the response squares changed from trial to trial (see, Figure 2).

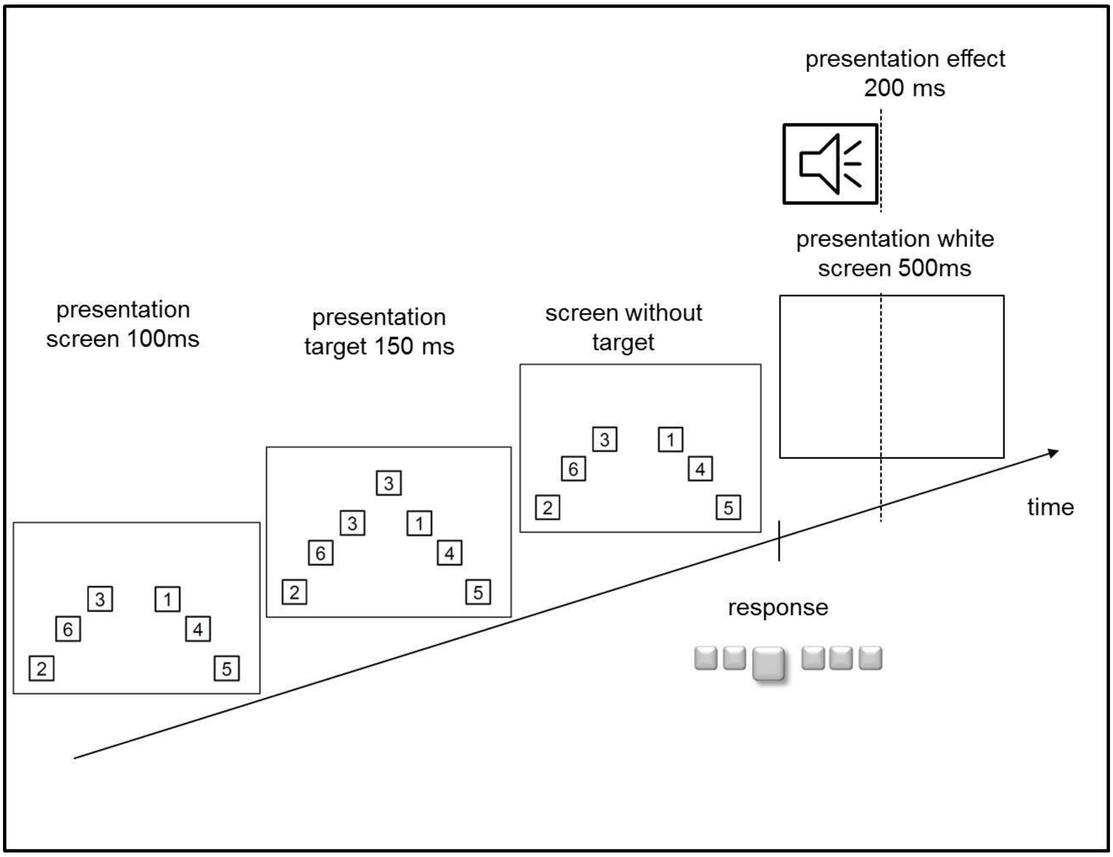

Fig. 2[3]. Succession of the events within one SRTT-trial. The numbers represent the different colors of the stimuli. 1=yellow, 2=red, 3=pink, 4=cyan, 5=blue, 6=green.

The locations of the response squares were spatially mapped to the response keys Y, X, C, B, N and M keys on a German QWERTZ-keyboard. The participants were instructed to let their index-, ring-, and middle-fingers rest on the six response keys for the entire experiment and to press the key spatially corresponding to the location of the response square containing the target color. Independent of the correctness of the response, every key press produced a sine tone of 400 Hz, 600Hz, 800Hz, 1000Hz, 1200Hz, or 1400 Hz. Only these effect tones followed a first order 6-elements sequence (1200Hz, 600 Hz, 1400 Hz, 800 Hz, 1000 Hz,

[3] In the original manuscript, the figures are numbered differently. To avoid confusions, the numbering is adapted to the consecutive numbering of this thesis.

400Hz). In Block 5, the tone sequence was replaced by a pseudo-random sequence. The only constraint of this pseudo-random sequence was that it did not contain immediate repetitions of tones.

Procedure

The experiment started with computer-based instructions. The participants were informed what their task was. In particular, they were told that after each keypress they will hear a tone which should help them to keep concentrated. Since the tones were task-redundant, we used this cover story to ensure that the participants would attend to the tones as part of the task-set (e.g., Dreisbach & Haider, 2009; Eberhardt et al., 2017). Afterwards, all participants performed six blocks of the SRTT, with 90 trials each. They were allowed to make a short break of a freely chosen length after each block.

Each trial started with the presentation of the six colored response squares. After 100ms the target was displayed in the upper third of the screen for 150ms. The participants had to find as quickly as possible the response square containing the target's color and to press the corresponding key. Immediately after the keypress, one of the six sinus-waved tones sounded as the response-effect for 200ms and the response squares disappeared. The next trial started after a response stimulus interval (RSI) of 500ms.

After having finished the last block of the SRTT, the participants were asked about their knowledge about the fixed effect sequence in a post-experimental interview and were then debriefed about the experiment.

Results and Discussion

Mean error rates and mean RTs in the SRTT were computed separately for each participant and each block. For the mean RTs, erroneous trials (5.32% of all data) were excluded. Furthermore, negative RTs and RTs one standard-deviation above the mean RTs (6.82% correct trials)[4] were excluded. We used such an unusual strict exclusion criterion because we assumed that the presentation of the pure tone-sequence will affect only the response-initiation processes (Kunde et al., 2004), and thus will lead to very small RT-increases when replacing the effect-tone sequence in the random block. Lepper et al., (2008) found a 12ms difference between the sequenced and the random response-effect trials. To ensure that the

[4] Since the single RTs were not normally distributed, the amount of excluded trials is lesser than 16%.

expected small effects were not overshadowed by random RT fluctuations, we decided to reduce the noisy influence of outliers.

In all experiments, we adopted an α-level of 5%. If Mauchly's test of sphericity reached significance, we report Greenhouse Geisser corrected p-values together with the original degrees of freedom.

Error rates and latencies in the SRTT

An one-way repeated measures ANOVA with Block as independent variable and mean error-rates as dependent variable just failed to reach the level of significance ($F(5,195) = 2.43$, $p = .056$, $\eta_p^2 = 0.06$). As can be seen from Table 1, the participants made only few errors ($M = .053$; $SD = .031$).

The analogous ANOVA with mean RTs as dependent variable revealed a significant effect of Block ($F(5,195) = 10.34$, $p < .001$, $\eta_p^2 = 0.21$). As shown in Table 1, mean RTs decreased over time.

Table 1. Mean percent error rates and RTs per block in Experiment 1a. The respective standard deviations are presented in brackets.

Block	Mean Error rates	Mean RTs
Block 1	.053(.041)	699.55(70.32)
Block 2	.043(.026)	676.06(71.95)
Block 3	.052(.039)	679.61(74.71)
Block 4	.055(.045)	671.11(70.64)
Block 5 (random)	.056(.031)	677.67(74.22)
Block 6	.060(.045)	663.79(70.97)

Test-block

To investigate our main hypothesis, we compared the mean RTs in the random Block 5 with the mean RTs of the two adjacent sequential Blocks 4 and 6. The replacement of the fixed effect sequence led to an RT-increase of 10ms (Figure 3). A t-test revealed a significant effect

($t(39)=3.42$, $p=.001$, $d=0.55$) [5]. As can be seen in Figure 3, RTs were slower in the random than in the sequential blocks.

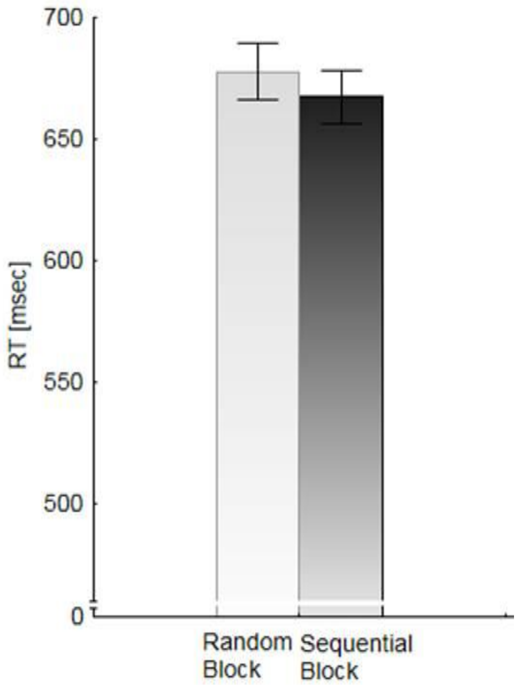

Fig. 3. Mean RTs for Bock 5 and the means of Block 4 and 6 in Experiment 1a. Error bars reflect the standard errors of the means.

The participants' sequence knowledge was assessed in a post-experimental interview. In this post-experimental interview only twelve of the forty participants (30%) reported to have noticed any system in the tone effects, but were not able to describe the sequence.

Thus, we found a significant increase of 10ms in the random block likely indicating implicit learning of the fixed effect sequence. However, since this learning effect was so small one might suspect that it simply resulted from accidental fluctuations in the RTs. This is further supported when one carefully inspects the mean RTs in Table 1. As can be seen, the mean RTs slightly decelerate from Block 2 to Block 3. However, when conducting the analogous t-test between Block 3 and the two adjacent blocks, this increase just failed the level of significance ($t(39)=2.01$, $p=.051$, $d=0.32$)[6]. Thus, albeit this weakness of Experiment 1a, we used the findings as a basis for now asking whether the implicitly learned effect sequence

[5] Note that the same pattern of RTs over the six blocks was found when no errorless RTs were excluded as outlier. The difference between the random Block 5 and the mean of the two adjacent Blocks 4 and 6 just failed to reach significance ($t(39)=1.79$, $p=.081$, $d=0.29$).

[6] The comparison between Block 3 and the adjacent Blocks 2 and 4 reached significance when no errorless RTs were excluded as outlier. Thus, one can say that the difference between the Block 5 and the adjacent Blocks reached significance by excluding extreme RTs as outlier whereas the significant difference between Block 3 and the mean of Blocks 2 and 4 disappeared.

would amplify the development of explicit knowledge about a response-location sequence. Furthermore, we replicated the findings in Experiment 2a.

Experiment 1b

Based on these findings, we turned to our main question whether implicitly learning an auditory effect sequence in advance would facilitate the (explicit) acquisition of a later presented response-location sequence. In addition, we also tested if such a facilitation effect is unique for an effect sequence.

For this purpose, the experiment comprised three conditions: The participants in the Tone-Effect condition received in the acquisition phase (3 blocks) the fixed sequence of effect tones while the response locations and the stimuli were randomly presented. In the No-Tone Control condition, the participants did not hear the effect-tones during the acquisition phase. In an immediately following test phase (3 blocks, without any further instruction), both conditions then received the effect tones which now were consistently mapped to the response locations. The additional Stimulus-Color condition was in principal identical to the Tone-Effect condition. However, here the colors of the imperative target stimulus followed a fixed sequence during the acquisition phase while the effect-tones and the response locations were random. In the test-phase the visual stimulus sequence was then mapped to the response locations and the effect tones were randomly presented. Thus, in all three conditions the tone-effect and the following next stimulus were always randomly associated (Ziessler & Nattkemper, 2001).

In order to assess the participants’ knowledge about the response-location sequence, all participants received the post-decision wagering-task (PDWT, Persaud, McLeod & Covey, 2007) immediately after having finished the last block of the test phase (see, Haider et al., 2011, 2012, 2014). The PDWT resembles the generation task (Shanks & Johnston, 1999; Wilkinson & Shanks, 2004). In certain trials, the participants receive no target stimulus and are asked to guess the next response by themselves. Additionally, they have to put a high or low wager on their response, reflecting their confidence in the correctness of their answer. According to the zero correlation criterion (Dienes, Altmann, Kwan & Goode, 1995), participants with implicit knowledge are assumed to give more correct responses than to be expected by mere guessing while concurrently showing no correlation between the correctness of their response and their confidence-ratings. By contrast, participants with explicit knowledge should rate their correct responses with high certainty and their incorrect responses with low certainty. Including the subjective metacognition about the correctness of one's own responses allows to investigate whether a manipulation primarily influences implicit or explicit knowledge (Haider et al., 2012).

If only learning the effect sequence in advance facilitates the acquisition of (explicit) knowledge about a later presented response-location sequence, we should find more knowledge about the response-location sequence in the Tone-Effect than in the No-Tone Control condition or the Stimulus-Color condition. If, however, this facilitation effect is true irrespectively of whether the in-advance learned sequence consists of stimuli or of effects, the participants in the Tone-Effect and the Stimulus-Color conditions should possess the same amount of knowledge about the response-location sequence.

Method

Participants

Participants were 91 new students (17 male) of the University of Cologne with a mean age of 23.33 years (*SD*=3.55). Thirty participants were assigned to the Tone-Effect condition, 30 to the No-Tone Control condition and 31 to the Stimulus-Color condition. All participants again received either course credit or payment (3€) for their participation.

According to our exclusion criteria (see Experiment 1a), the data of two participants (1 in the Tone-Effect condition, 1 in the Stimulus-Color condition) were replaced because they did not reach our accuracy criterion. In addition, the data of 14 participants (7 in the Tone-Effect condition, 2 in the No-Tone Control condition, 5 in the Stimulus-Color condition) were replaced due to technical problems or because they already had participated in a former SRTT-experiment.

Apparatus and Stimuli

Apart from the introduction of the response-location sequence (described below), the materials were the same as in Experiment 1a.

Procedure

The experiment started for all participants with the computer-based instructions. Stimulus-presentation was as in Experiment 1a. The SRTT training was divided in an acquisition phase (3 blocks) and a test phase (3 blocks). For the Tone-Effect condition, the acquisition phase was identical to the procedure of Experiment 1a. The No-Tone Control condition did not receive the auditory effect sequence. The participants in the Stimulus-Color condition received a repeated six-elements first-order sequence of the color targets (red, cyan, yellow, green, pink, blue), whereas the tones and the response locations were presented randomly. For all participants, the test phase started after the third block without any further announcement. Here, they all received a six-elements first-order response-location sequence which was perfectly correlated with either the auditory effect sequence (Tone-Effect condition, No-Tone Control condition; response-location effect-tone mapping was: 5-1200Hz, 2-600Hz, 6-1400Hz, 3-800Hz, 4-1000Hz, 1-400Hz) or the visual stimulus sequence (Stimulus-Color condition; response location stimulus color mapping was: 5-red, 2-cyan, 6-yellow, 3-green, 4-pink, 1-blue combined with randomly presented effect tones).

After completing the SRTT, all participants were administered to the PDWT (Persaud et al., 2007) to assess their knowledge about the response-location sequence. With three exceptions, this task was identical to the SRTT training. The participants responded again to the location of the response square containing the target's color. However, they received (a) no effect tones (b) target colors were randomly presented for all participants, and (c) 36 trials of the overall 150 trials were so-called wager trials. In these trials, instead of the target stimulus a question mark occurred at the target position. Participants then had to guess the next response location by pressing the respective response key. Subsequently, two wager options (1 Cent vs. 50 Cents) appeared on the screen and the participants had to indicate by pressing either the "a"-key (1 Cent) or the "l"-key (50 Cents) how certain they were that their answer was correct. If the entered response location was correct according to the sequence, the participants won the amount of the wager. If it was incorrect, they lost it. Importantly, the participants received no immediate feedback about the correctness of their responses. After 18 and 36 wagering trials, they only were informed about their gains.

Based on former experiments, we terminated the wager-task when a participant had reached a maximum of 8€ extra money. The earliest point to reach this criterion was after 16 correct wager trials if the participant always wagered 50 Cent. This procedure helps identifying participants with predominantly explicit knowledge. An early termination of the wager task highly correlates with the ability to verbally reproduce the sequence (Haider et al., 2011).The experiment ended again with the post-experimental interview and the debriefing of the participants.

Results and Discussion

Again, we computed mean error rates and RTs separately for each participant and each block. For the analysis of mean RTs, we also excluded erroneous trials (4.83% of all data) and RTs one standard deviation above the mean of correct trials as outliner (7.68% of correct trials). Our report of the results is divided into three parts. We first describe the findings concerning the acquisition and the test phases and then report the findings of the wagering-task.

Error rates and latencies in the SRTT

Mean error rates and mean RTs are presented in Table 2. For the mean error rates as dependent variable, the 3 (Condition: Tone-Effect, No-Tone Control, Stimulus-Color) x 6 (Block) ANOVA did not show any significant effect (Block: $F(5,440)=2.28$, $p=.080$, $\eta_p^2=0.03$; Condition: $F(2,88)=.36$, $p=.696$, $\eta_p^2=0.01$; Interaction: $F(10,440)=1.16$, $p=.326$, $\eta_p^2=0.03$). As in Experiment 1a, the participants made overall only few errors ($M=.048$, $SD=.033$; see Table 2).

The analogous ANOVA with mean RTs as dependent variable revealed a significant main effect of Block ($F(5,440)=72.04$, $p<.001$, $\eta_p^2=0.45$). The main effect of condition ($F<1$) as well as the interaction between Condition and Block ($F(10,440)=1.96$, $p=.104$, $\eta_p^2=0.04$) were not significant[7]. As can be seen from Table 2, the participants became faster over training.

In order to test more precisely our main question whether learning the effect sequence, but not a visual stimulus sequence in advance fosters the acquisition of the response location sequence, we conducted planned interaction contrasts between Blocks 3 vs. 6 X Condition. We focused on the comparison between the Tone-Effect versus the No-Tone Control conditions and the Tone-Effect versus the Stimulus-Color conditions. The first planned interaction contrast reached our level of significance ($F(1,88)=4.14$, p=.044, $\eta_p^2=0.04$) whereas the second did not ($F<1$). In addition, we computed Bayes-Analyses with JASP (JASP Team, 2018; Morey & Rouder, 2015) for ANOVA designs (Rouder, Morey, Speckmann & Province, 2012) to test the plausibility that the participants in the Tone-Effect condition did not differ significantly from participants in the Stimulus-Color condition. As we originally expected a difference in RTs between these two conditions, we calculated a Bayes-Factor BF_{10} to test the alternative hypothesis (H_1) to be favored over the null hypothesis (H_0). This Bayes-Factor was $BF_{10} = 0.325$. According to Jeffreys (1939/1961), a Bayes-Factor < 0.33 indicates that the alternative hypothesis cannot be favored over the null hypothesis.

[7] Note that we found rather the same pattern of results when no errorless RTs were excluded as outlier. The main difference was that the Block X condition interaction was significant.

In sum, after introducing the response-location sequence the mean RTs in the Tone-Effect condition accelerated significantly quicker than in the No-Tone Control condition. The Stimulus-Color condition did not differ significantly from the Tone-Effect condition.

Table 2. Mean percent error rates and RTs per block in Experiment 1b. The respective standard deviations are presented in brackets.

Condition						
	Tone-Effect		No-Tone Control		Stimulus-Color	
Block	Error rates	RTs	Error rates	RTs	Error rates	RTs
Block 1	.049(.050)	686.37(58.02)	.042(.034)	686.17(56.09)	.042(.032)	691.41(76.70)
Block 2	.044(.040)	663.74(59.19)	.044(.034)	667.28(65.59)	.044(.029)	664.65(78.33)
Block 3	.045(.034)	660.88(51.62)	.041(.033)	669.30(64.11)	.046(.032)	664.12(76.20)
Block 4	.048(.038)	638.32(73.03)	.059(.041)	646.52(65.84)	.047(.047)	632.75(91.93)
Block 5	.051(.044)	594.46(117.78)	.067(.044)	630.66(66.33)	.048(.050)	603.16(97.53)
Block 6	.056(.090)	552.65(137.92)	.058(.038)	609.57(82.12)	.041(.038)	571.35(115.87)

Post-Decision Wagering Task

With regard to the wagering trials of the PDWT, we computed for each participant the percent correct responses and the participants' certainty judgment given by the percent high wagers under the condition of a participant's correct (high|correct) versus incorrect responses (high|incorrect; see Table 3). Concerning the mean percent correct responses, the participants in all conditions had significantly more knowledge about the response-location sequence than expected by chance (chance level = 20% correct responses; Tone-Effect condition: $t(29)=5.09$, $p<.001$; No-Tone Control condition: $t(29)=4.00$, $p<.001$, Stimulus-Color condition: $t(30)=4.23$, $p<.001$). In addition, a one-way ANOVA with Condition as independent and percent correct responses as dependent variables yielded a significant effect ($F(2,88)=4.78$, $p=.010$, $\eta_p^2=0.10$). The planned contrasts showed that the participants in the Tone-Effect condition possessed significantly more knowledge about the response-location sequence than the participants in the No-Tone Control condition ($F(1,88)=6.41$, $p=.013$, $\eta_p^2=0.07$) and in the Stimulus-Color condition ($F(1,88)=7.90$, $p=.006$, $\eta_p^2=0.08$).

To test if learning the effect sequence in advance influenced the acquisition of implicit and/or explicit knowledge, we analyzed the participants' certainty judgments. A 3 (Condition: Tone-Effect, No-Tone Control, Stimulus-Color) x 2 (Certainty Judgment: high|correct, high|incorrect) ANOVA with percent high wagers as dependent variable revealed a significant main effect of Certainty Judgment ($F(1,88)=9.41$, $p=.002$, $\eta_p^2=0.10$). The main effect of Condition ($F(2,88)=1.48$, $p=.232$, $\eta_p^2=0.03$) and the interaction were ($F(2,88)=1.04$, $p=.358$, $\eta_p^2=0.02$) insignificant. Table 3 shows that the participants in the Tone-Effect condition tended to place more high wagers when they responded correctly than when they responded incorrectly as compared to the No-Tone Control condition and the Stimulus-Color condition. However, this was only a trend.

Table 3. Means and standard deviations of percent correct responses and high wagers for correct vs. incorrect responses for Experiment 1b.

Condition	Correct responses	High|Correct	High|Incorrect	N
Tone-Effect Condition	49.43(30.71)	57.13(34.32)	42.36(30.87)	30
No-Tone Control Condition	34.33(19.63)	40.72(34.87)	33.72(27.72)	30
Stimulus-Color Condition	37.80(16.81)	44.04(34.02)	38.33(29.00)	31
Without participants with predominantly explicit knowledge				
Tone-Effect Condition	30.78(11.87)	41.55(29.12)	40.68(27.19)	21
No-Tone Control Condition	30.26(12.27)	36.49(32.05)	34.34(27.48)	28
Stimulus-Color Condition	30.56(11.46)	42.54(33.55)	40.30(28.57)	30

As mentioned before, the PDWT allows identifying those participants who had acquired predominantly explicit knowledge about the response location sequence (Haider et al., 2011). These participants gain the maximum of wagers (8€) and are terminated before the end of the wagering task. In the Tone-Effect condition, nine out of 30 participants (30%) were classified as possessing predominantly explicit knowledge. By contrast, we found only 2 such participants (6.7%) in the No-Tone Control condition and one in the Stimulus-Color condition 1 (3.2%). A chi-square confirmed that the three conditions differed significantly ($\chi^2(2)=11.21$; $p=.003$, $\varphi=0.35$).

Taken together, the findings of the PDWT revealed that the participants in the Tone-Effect condition had more knowledge about the response-location sequence than the participants in the No-Tone Control and the Stimulus-Color conditions. In addition, it seems that this was true particularly for explicit sequence knowledge.

Since our main question was whether the effect tones fostered the development of only the explicit knowledge, we reanalyzed the data without the twelve participants showing predominantly explicit knowledge in the PDWT. The results are shown in the lower part of

Table 3. As can be seen from Table 3, the numerical difference between the conditions almost entirely vanished and also the higher amount of high wagers under the condition of correct responses disappeared. We computed a Bayes-analysis for ANOVA designs (Rouder et al., 2012) to test whether the participants with predominantly implicit knowledge did not differ in placing high wagers for correct vs. incorrect responses. We calculated a Bayes-Factor BF_{01} to test the plausibility of the null hypothesis (H_0; Certainty Judgement does not differ for correct vs incorrect responses) to be favored over the alternative hypothesis (H_1; Certainty Judgement differs). The Bayes factor was BF_{01} = 3.832 and thus larger than 3, which supports according to Jeffreys (1939/1961) the here tested null hypothesis.

For means of illustration, Figure 4 also depicts the mean RTs of the acquisition and the test phases separately for the participants with predominantly explicit knowledge versus those with predominantly implicit knowledge. Three findings are noteworthy: First, only the participants with predominantly explicit knowledge exhibit a steep decrease of the mean RTs after the response-location sequence had been added in the test phase. In contrast, the participants with mostly implicit knowledge do not show such a strong decrease of RTs. Second, on the descriptive level, the decrease was strongest for the explicit participants in the Tone-Effect condition and already appeared in Block 4. In the No-Tone Control condition the strongest decrease of RTs is found from Blocks 5 to 6. Third, when looking on only those participants with predominantly implicit knowledge (left panel), one can see that the advantage of the Tone-Effect condition almost completely vanished. If at all, the fastest condition here is the Stimulus-Color condition.

In addition, we computed a Bayes-Factor BF_{10} for ANOVA designs (Rouder et al., 2012) for the formerly significant interaction contrast (Block 3 vs. 6 X Effect-Tone vs. No-Tone Control conditions). We tested the plausibility of the H1 hypothesis that there is a significant interaction contrast for the Tone-Effect versus No-Tone Control condition. We found a Bayes-Factor of BF_{10}= 0.29, which speaks against the tested alternative hypothesis (Jeffreys, 1939/1961).

Thus, the above reported difference between the conditions was almost entirely driven by the participants with explicit knowledge. This further supports our former findings that R-E learning in an SRTT is more likely to influence the development of explicit rather than implicit knowledge (Esser & Haider, 2017a; Haider et al., 2014).

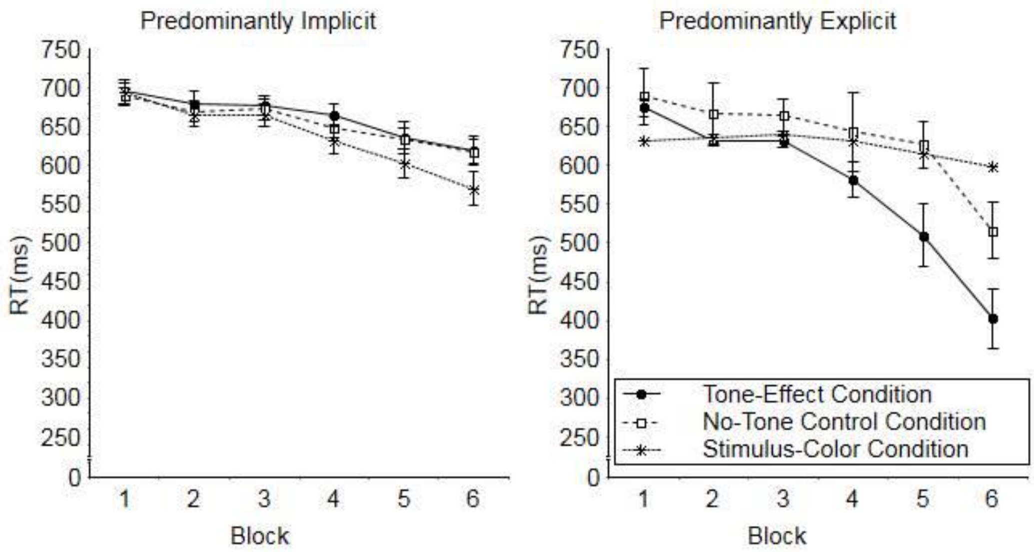

Fig. 4. Mean RTs for each block for the participants with predominantly implicit (left panel) and predominantly explicit knowledge (right panel) in Experiment 1b. Error bars reflect the standard errors of the means.

In combination, the findings of Experiments 1a and 1b suggest that a sequence of task-redundant auditory effects can be learned implicitly even if it is not mapped to either stimuli or response locations (Lepper et al., 2008). When, after a learning phase this learned effect sequence becomes contingently mapped to the response locations, it amplifies the acquisition of mainly explicit rather than implicit knowledge about this new sequence. In contrast, learning a visual stimulus sequence in advance does not yield such an advantage of learning a response-location sequence. This is in line with the assumptions that response-effects in the SRTT might influence primarily explicit rather than implicit learning processes (see, Esser & Haider, 2017a). Two not mutually exclusive arguments against this interpretation should be considered. First, one might argue that the participants had not learned the visual stimulus sequence during the acquisition phase when it was not contingently mapped to the responses. However, Haider et al. (2012, 2014) and also Eberhardt et al. (2017) already showed that participants do learn such a sequence of stimuli. Second, the effect tones might have been more salient than the visual stimuli leading participants to attend more to the response-location sequence. Also this argument is less likely because the effect tones were irrelevant for performing the main task whereas the visual stimuli were needed in order to respond correctly.

In Experiments 2a and 2b, we replaced the auditory effects by visual effects in order to replicate and generalize the current findings. For instance, Ziessler and Nattkemper (2002) pointed out that auditory stimuli may be more salient than other stimuli. Therefore, an open question is whether the participants will learn a fixed sequence of task-redundant visual effects when these effects are not mapped to specific responses.

Experiment 2a

The main goal of Experiment 2a was to replicate the findings of Experiment 1a with a fixed sequence of visual effects and thus, to test the generalizability of the current results. We used a slightly different variant of the SRTT. The main difference was that instead of presenting the target stimulus above the response stimuli, the target position in each trial was announced by changing the color of one single response stimulus. After each response all six response squares changed into another shape as the visual effects. As we did in Experiment 1a, we replaced the fixed effect sequence in Block 5 and re-introduced it in Block 6. In addition, subsequently after training the participants were administered to the PDWT to assess their knowledge about the fixed effect sequence.

Methods

Participants

Thirty new students (8 male) of the University of Cologne with a mean age of 23.57 years (*SD*=6.34) served as participants in Experiment 2a. They either received course credit or payment (3€).

Apparatus and Stimuli

Six white response squares with black frames appeared in a triangular space in the lower two thirds of the screen. Instead of presenting the target square above the response squares, the target location here was announced by changing the color of one response square from white to gray. Immediately after the participants' response, the shapes of all six response squares changed into triangles, stars, circles, hearts, moons, or clouds as the visual effect. Only these shapes followed a six-elements first-order sequence (moon, star, heart, circle, cloud, triangle), In Block 5, this sequence was replaced by a pseudo-random sequence and was re-introduced in Block 6. No other sequence was implemented into the task.

Procedure

With the two exceptions just described, the procedure of the SRTT training was identical to that of Experiment 1a. Subsequently after the last block of the SRTT, the participants were administered to a short version of the PDWT to assess their knowledge about the fixed shape-effect sequence (see Experiment 1b). Testing this knowledge required to slightly change the design of training task. The response squares were replaced by the six different shapes, presented in the same triangular space. Furthermore, here one shape served as the target shape. It appeared in the upper third of the screen 100ms after the response shapes and disappeared after 150ms. The participant's task was to find the target shape among the six shapes of the response squares and to press the spatially corresponding response key.

In the wager trials, a question mark occurred instead of the target shape. Participants then had to guess the shape they believed would occur next by pressing the respective key mapped to the predicted shape. Subsequently, they were asked again to indicate how certain they were about the correctness of their response by placing a wager of either 1 Cent or 50 Cent. The PDWT contained only 55 trials with overall 13 wager trials. The participants could earn 5€ when they were able to maximize their wagers. The experiment ended again with the post-experimental interview concerning the participant's knowledge about the visual shape-effect sequence followed by a short debriefing.

Results and Discussion

First, we calculated the mean error rates and mean RTs separately for each block and each participant. According to our exclusion criteria, erroneous trials (3.70% of all trials) and RT-outliers were excluded from the analysis of RTs (8.76% of correct trials). We first report the SRTT results followed by the findings of the PDWT.

Error rates and latencies in the SRTT

The mean error-rates and mean RTs are listed in Table 4. Table 4 shows that the error rates are rather low (M = .037; SD=.018) and slightly increase over the course of training. A one-way repeated measures ANOVA with mean error-rates as dependent variable revealed a significant effect of block ($F(5,145)=4.02$, $p=.004$, $\eta_p^2=0.12$).

Table 4. Mean percent error rates and RTs per block in Experiment 2a. The respective standard deviations are presented in brackets.

Block	Error rates	RTs
Block 1	.036(.023)	526.45(63.89)
Block 2	.024(.018)	510.16(65.57)
Block 3	.034(.025)	509.36(58.15)
Block 4	.039(.027)	502.79(51.29)
Block 5	.045(.031)	504.46(52.17)
Block 6	.044(.031)	494.24(55.03)

The analogous ANOVA for mean RTs as dependent variable yielded a significant main effect of Block ($F(5,145)=10.65$, $p<.001$, $\eta_p^2=0.27$). RTs decreased over time, but slightly increase in the Test Block 5 (see, Table 4).

Test-block

As in Experiment 1a, we compared the mean RTs of the random Block 5 with the mean RTs of the two adjacent sequence Blocks 4 and 6 (Figure 5). A t-test revealed that the RT-increase of 6ms in Block 5, albeit small, is significant ($t(29)=2.20$, $p=.035$, $d=0.41$)[8]. Thus, the removal of the effect sequence led to an increase in mean RTs. However, the effect was weaker than that of Experiment 1a (d = 0.55 in Experiment 1a).

[8] In contrast to Experiment 1a, we did not find an almost significant increase of RTs in Block 5, when no correct RTs were excluded as outlier from further analysis.

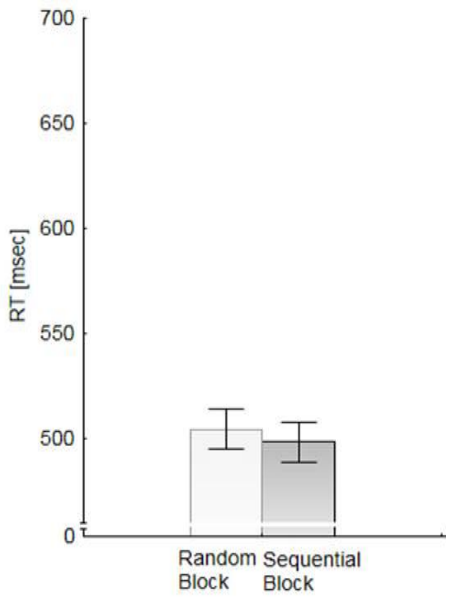

Fig. 5. Mean RT for Bock 5 and the means of Block 4 and 6 in Experiment 2a. Error bars reflect the standard errors of the means.

Post-Decision Wagering Task

In the PDWT, the participants responded correctly in 25.13% (*SD*=10.29%) of all wagering trials. This amount of knowledge is significantly above the chance level ($t(29)=2.73$, $p=.011$). To assess the extent of explicit knowledge, we also analyzed the Certainty judgments of the PDWT. The percentages of high wagers under the condition of correct responses (*MW*=33.29%; *SD*=35.59%) were numerically smaller than that under the condition of incorrect responses (*MW*=37.07%; SD=27.86%). A t-test indicated no significant difference ($|t|< 1$). We additionally computed a Bayes-Factor for t-tests (Rouder, Speckmann, Sun, Morey & Iverson, 2009) BF_{01} to test the plausibility of the null hypothesis to be favored over the H1. For the calculation of Bayes-Factors for t-tests with JASP (JASP Team, 2018), we used the Cauchy distribution as prior distribution with a Cauchy-scale of .0707. We found a Bayes factor of BF_{01} = 3.78, which according to Jeffreys (1939/1961) speaks for the here tested null hypothesis. Thus, the lacking difference between the Certainty judgements of the PDWT indicates that the acquired knowledge, albeit above the chance level, seems to be unconscious. There also was no participant with primarily explicit knowledge.

In sum, the findings of Experiment 2a replicated the findings of Experiment 1a with a visual shape-effect sequence. The influence of the shape-effects on performance was weaker than that of the auditory tone-effects in Experiment 1a. Nevertheless, the performance in the training and the PDWT converge to the point that the participants had acquired some knowledge about the fixed effect sequence. Importantly, this knowledge seems to be entirely implicit.

Experiment 2b

After having shown that the participants could learn the fixed shape-effect sequence, we tested whether we could replicate also the findings of Experiment 1b. Again, we focused on the question of whether learning the fixed effect sequence in advance would facilitate the acquisition of knowledge about the response-location sequence when the responses were contingently mapped to the shape-effects. For this purpose, we compared the visual Shape-Effect condition, in which the participants received the fixed effect sequence already during the acquisition phase, with a Control condition in which the participants did not receive the shape effects in the acquisition phase.

Method

Participants

Sixty new students (7 male) of the University of Cologne participated in Experiment 2b, respectively 30 in the Shape-Effect and the Control conditions. Their mean age was 22.65 years (*SD*=4.39). Participants again received either course credit or payment (3€).

Three participants (one in the Shape-Effect condition; two in the Control condition) were replaced by new participants because their error rates exceeded our error criterion.

Apparatus and Stimuli

The materials were the same as in Experiment 2a.

Procedure

The training procedure was the same as in Experiment 2a, except that the contingently mapped response-location sequence was introduced in Blocks 4 to 6. The PDWT was almost identical to that described for Experiment 2a but consisted of 150 trials and 36 wager trials. The second difference of the here used PDWT was that we tested the response-location sequence instead of the shape-effect sequence. Thus as in the SRTT, participants had to respond to the marked stimulus positions instead of to the target-shape.

Results and discussion

Again, we first computed mean error rates and mean RTs separately for each block and each participant. For the analysis of mean RTs in the SRTT, erroneous trials (4.55% of all data) and RTs with more than one standard deviation over the mean RTs were excluded (5.17% of correct trials).

Error rates and latencies in the SRTT

Table 5 shows the mean error rates and the mean RTs. A 2 (Condition: Shape-Effect vs. Control) x 6 (Block) ANOVA with error rates as dependent variable did not reveal any significant effect (Condition: $F(1,58)=2.43$, $p=.124$, $\eta_p^2=0.29$; Block: $F(5,290)=.79$, $p=.495$, $\eta_p^2=0.12$; interaction: $F(5,290)=1.28$, $p=.282$, $\eta_p^2=0.02$).

The same ANOVA conducted for the mean RTs as dependent variable showed a significant main effect of Block ($F(5,290)=73.90$, $p<.001$, $\eta_p^2=0.56$) and a significant Condition x Block interaction ($F(5,290)=7.01$, $p=.001$, $\eta_p^2=0.11$). The main effect of Condition was not significant ($F(1,58)=1.02$, $p=.317$, $\eta_p^2=0.02$).

As Table 5 shows, the participants in both conditions became faster over time. More important, though, in comparison to the Control condition, the participants in the Shape-Effect condition exhibited again a steeper decrease in RTs in the test phase. This was confirmed by a significant planned interaction contrast (Block 3 vs. 6 X Condition: $F(1,58)=8.08$, $p=.006$, $\eta_p^2=0.12$)[9].

However and against our expectation, Table 5 also reveals that the mean RTs in Blocks 1 to 3 largely differ between the two conditions. The participants in the Shape-Effect condition responded approximately 45ms slower than the participants in the Control condition. This suggests that the shape effects, as long as they were not contingently mapped to the responses, may have impaired the performance in the SRTT. Yet, as soon as the participants experienced the consistent relation between their responses and the appearance of the respective shape effect, they seem to overcome this potential interference and, as in Experiment 1b, stronger speed-up responding than the participants in the Control condition. Thus, despite this unexpected difference between the two conditions in the acquisition phase, the pattern of the performance data resembles that of Experiment 1b.

[9] Note that we found the same results when no correct RTs were excluded as outlier from further analysis.

Table 5. Mean percent error rates and mean RTs per block for the Shape-Effect and the Control conditions of Experiment 2b. The respective standard deviations are presented in brackets.

	Shape-Effect Condition		Control Condition	
Block	Error rates	RTs	Error rates	RTs
Block 1	.042(.034)	535.24(67.15)	.046(.033)	490.86(37.93)
Block 2	.044(.034)	524.63(65.35)	.051(.040)	479.61(39.42)
Block 3	.041(.033)	519.55(59.36)	.058(.051)	482.41(36.35)
Block 4	.059(.041)	487.75(83.62)	.059(.040)	480.18(63.30)
Block 5	.067(.044)	426.13(103.52)	.047(.032)	439.31(70.85)
Block 6	.058(.038)	370.50(117.45)	.044(.034)	397.51(84.43)

Post-Decision Wagering Task

Table 6 shows the mean percent correct responses and the mean percent high wagers for the high|correct and high|incorrect categories of the Certainty Judgment. Concerning the mean percent correct responses, the participants in both conditions had significantly more knowledge about the response-location sequence than expected by mere guessing (Shape-Effect: $t(29)=4.25$, $p<.001$; Control: $t(29)=3.44$, $p<.001$). However and in contrast to Experiment 1b, the amount of knowledge did not differ significantly between the two conditions ($t(58)=1.47$, $p=.147$, $d=0.19$). As this finding was against our hypothesis, we calculated a Bayes-Factor BF_{10} for t-tests (Rouder et al., 2009) to test the plausibility of favoring the alternative hypothesis (H_1) over the Null Hypothesis. The resulting Bayes factor of $BF_{10} = 0.63$ indicates insensitivity (Jeffreys, 1939/1961). Thus, the insignificant t-test was due to the insensitive data rather than to a plausible lacking difference between the two conditions. The comparison of the mean percent high wagers under the condition of correct versus incorrect responses also did not indicate any significant difference between the two conditions. The 2 (Condition: Shape-Effect, Control) x 2 (Certainty Judgement) ANOVA with percent high wagers as independent variable revealed only a significant main effect of Certainty Judgment ($F(1,58)=10.19$, $p=.002$, $\eta_p^2=0.15$). Neither the main effect of Condition nor the interaction was significant (both Fs<1). As in Experiment 1b, the participants in the two conditions differed only on a numerical level in their way of placing high wagers (see, Table 6).

Table 6. Means and standard deviations of percent correct responses and high vs. low wagers for correct responses for Experiment 2b.

Condition	Correct responses	High|Correct	High|Incorrect	N
Shape-Effect Condition	56.94(28.47)	56.00(36.47)	34.68(31.29)	30
Control Condition	46.98(23.93)	55.32(28.78)	44.17(30.74)	30
Without participants with predominantly explicit knowledge				
Shape-Effect Condition	41.27(16.93)	40.37(31.68)	38.51(28.13)	21
Control Condition	37.38(14.24)	47.38(25.28)	48.27(26.03)	25

Using the same criterion as in Experiment 1b, we additionally tested whether the amount of participants who possessed predominantly explicit knowledge (explicit subgroup) differed between the two conditions. We identified nine such participants in the Shape-Effect condition (30%) and six in the Control condition (20%). Against our hypothesis, also this difference was not significance ($\chi^2(1)=1.87$; $p=.380$, $\varphi=.158$, Fishers exact p is reported). Compared to Experiment 1b, the explicit subgroup in the Control condition was much larger (6.7% in Experiment 1b versus 20% in the current experiment).

The reanalyses conducted in Experiment 1b have shown that the overall observed differences between the two conditions were almost entirely driven by the explicit subgroup. This implies that finding significant differences in the PDWT primarily depend on the amount of participants with explicit knowledge and how they are distributed across the two conditions. In the lower part of Table 6, the wagering results are shown again without the participants with primarily explicit knowledge. Also here, this eliminates the former effect of the Certainty Judgment. We also computed a Bayes-Factor BF_{01} for ANOVA designs (Rouder et al., 2012) to test the null hypothesis that there is no significant effect for Certainty Judgement. The Bayes factor of $BF_{01} = 4.53$ confirms the plausibility of the tested null hypothesis. In addition, Table 6 also shows that the numerical difference between the two conditions decreased after excluding the explicit subgroup. Thus, even in the current experiment, the observable differences between the conditions were almost entirely due to the explicit subgroup.

As shown in Figure 6, this was also true for the performance differences in the SRTT. Figure 5 depicts the mean RTs of the acquisition and the test phases separately for the implicit and explicit subgroups. As in Experiment 1b, primarily the participants with predominantly explicit knowledge showed a stronger decrease of RTs when the response-location sequence was added. And again, the explicit participants of the Shape-Effect condition showed this decrease already from Blocks 3 to 4 while the Control condition showed it from Blocks 4 to 5.

The 2 (Condition: Shape-Effect, Control) x 6 (Block) ANOVA for only participants with predominantly implicit knowledge revealed a significant main effect of Block ($F(5,220)=47.61$, $p<.001$, $\eta_p^2=0.52$) and a significant interaction ($F(5,220)=4.08$, $p=.020$, $\eta_p^2=0.08$). The main effect for condition was insignificant ($F(1,44)=2.92$, $p=.948$, $\eta_p^2=0.06$)[10]. The planned interaction contrast between Block 3 versus Block 6 X Condition reached significance ($F(1,44)=4.78$, $p=.035$, $\eta_p^2=0.10$). However, these differences between the two conditions for participants with predominantly implicit knowledge seem to be caused by the overall longer mean RTs in the experimental condition in the acquisition phase of Experiment 2b.

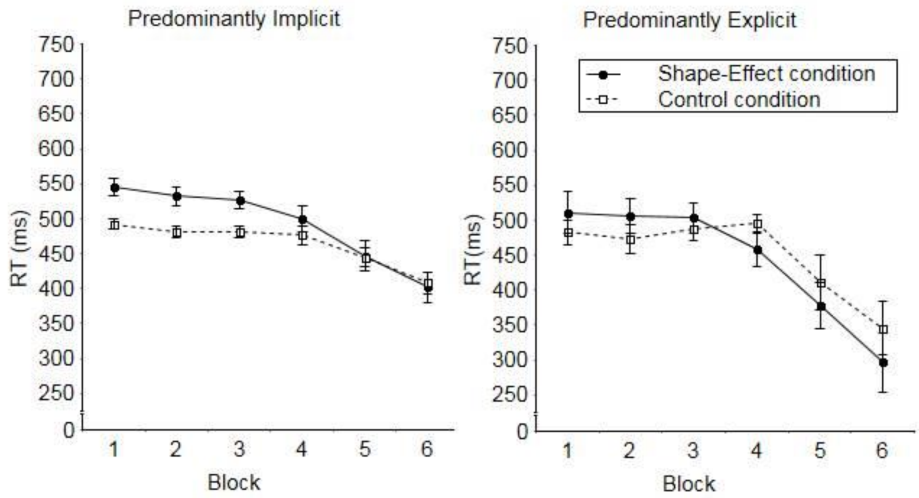

Fig. 6. Mean RT for each block separated for the participants with predominantly implicit and predominantly explicit knowledge in Experiment 2b. Error bars reflect the standard errors of the means.

Overall, the findings of Experiments 2a and 2b suggest that the visual shapes as effects are less effective than the auditory effects in Experiment 1. We could replicate the results of Experiment 1a by showing that the participants implicitly learned the fixed effect sequence even when they did not know how to produce these effects.

The findings of Experiment 2b are less clear. We could replicate the steeper RT-decrease in the Shape-Effect condition after introducing the response sequence in the test phase. However, due to the higher rate of the participants with explicit knowledge in the Control condition, our manipulation was not successful with regard to the acquired knowledge about

[10] Note that the main effect for condition reached significance even when no correct RTs were excluded as outlier from the analysis. In addition, the planned interaction contrast just failed to reach significance.

the response-location sequence. Nevertheless, the results and, in particular, the separate analysis of the two subgroups, suggest again that R-E learning in the SRTT seems to specifically affect the development of explicit knowledge (Esser & Haider, 2017a; Haider et al., 2014).

General Discussion

The main goal of the current study was to better understand why R-E associations amplify learning within an SRTT paradigm. As shown in the introduction, the theoretical explanation mainly differs with regard to the question of whether response-effects influence implicit or explicit learning processes and how. In particular, we focused on the assumption derived from the UEH that response-effect contingencies may boost mainly the development of explicit knowledge within an implicit learning situation (Esser & Haider, 2017a). The current study extends the findings of Esser and Haider in two points. First, we manipulated here the likelihood of experiencing an unexpected event (an unexpected feeling of sense of agency). Second, we also investigated whether the influence of the effect sequence learning applies for other than the auditory modality.

As a requirement to test this question, we first showed that participants could learn a fixed sequence of effects without knowing how to produce them. In line with the findings of Lepper et al. (2008) this was true for the auditory effects (Experiment 1a) and, to a lesser degree, also for the visual shape effects (Experiment 2a). The post-experimental interviews as well as the findings of the PDWT additionally confirmed that the acquired knowledge was predominantly implicit (for similar findings, see Eberhardt et al., 2017).

With regard to our main question, our findings are less clear. Experiment 1b showed that the participants in the Tone-Effect condition expressed more knowledge about the response location sequence than the participants in the No-Tone Control condition or the Stimulus-Color condition. This difference in the amount of knowledge was mostly driven by participants with predominantly explicit knowledge about the response-location condition. After excluding this explicit subgroup, the three conditions possessed rather the same amount of knowledge. With the visual effects in Experiment 2b, the qualitative pattern of results was similar. However, the amount of knowledge assessed in the PDWT did not differ between the two conditions, neither in terms of correct responses nor in terms of the Certainty Judgment.

Probably, three difficulties of Experiment 2b might have contributed to the insignificant difference between the experimental and control condition. First, the visual shape effects were presented within the same modality as the stimuli of the SRTT and therefore might have interfered with the main task of searching for the announced location in the SRTT. The longer mean RTs in the experimental condition of Experiment 2b are in line with this interference account. Second, we slightly changed the design of the SRTT in Experiments 2a and 2b. Instead of presenting the target stimulus above the response stimuli (as was the case in Experiment 1a and 1b), the target position here was announced by changing the color of one of the response squares. In the test phase of Experiment 2b, the participants thus saw a rather salient location sequence on the screen which was mapped to the response-keys on the one hand and to the visual response-effects on the other hand. This might have facilitated the detection of the response-location sequence in both conditions. Third, the fixed shape-effect sequence was learned to a lesser degree than the tone-effect sequence (see the comparison of Experiments 1a and 2a) presumably because the task-redundant visual effects are less salient (Ziessler & Nattkemper, 2002). This lesser learning of the fixed visual shape-effect sequence might have reduced the felt surprise when, through the introduction of the response-location sequence, the participants suddenly experienced how they could produce the shape effects. Given these weaknesses, the question remains open for further research whether only tone-effects play a special role in the development of explicit knowledge (Tubau et al., 2007).

Nevertheless, the results of Experiments 1b and 2b clearly suggest that the sudden introduction of the response-location sequence that was correlated to the effect sequence amplified the acquisition of primarily explicit knowledge. This finding nicely fit with the former findings of Esser and Haider (2017a) as well as Hoffmann et al. (2001), Haider et al. (2014), or Zirngibl and Koch (2002). In addition, our results suggest that this seems to occur in an all-or-none fashion. After excluding the participants with primarily explicit knowledge, the differences between the experimental and the control conditions almost entirely disappeared.

In the first part of this article, we had proposed three different accounts able to explain why response-effects might increase sequence learning. The first account of Keele et al. (2003) assumes that implicit learning is based on either the unidimensional or the multidimensional system. According to the Dual-System account, only knowledge acquired in the multidimensional system can become consciously accessible as it integrates information across different dimensions and requires selective attention. Thus, learning in our experiments should have involved the multidimensional system, at least in the test-phase of Experiments 1b and 2b. However, the finding that the participants in the Stimulus-Color condition of Experiment 1b did not possess more explicit knowledge seems to be at odds with this assumption because here a formerly learned visual stimulus sequence was mapped to a response-location sequence. Thus, also learning within this condition should have involved the multidimensional system. One way to incorporate this finding is to assume that the visual stimulus sequence was less salient than the tone-effect sequence. This then led to slower learning of the response-location sequence when the responses were mapped to either the visual stimuli or the tone-effects. This might have been the case given the findings of Experiment 2b. Yet, in Experiment 1b, the visual stimuli were task relevant whereas they were task-redundant when presented as response-effects in Experiment 2b. Therefore, it is not very likely that a difference in the salience might have moderated the higher amount of explicit knowledge in the Tone-Effect than in the Stimulus-Color conditions of Experiment 1b.

The second account, we introduced above, is the chunking assumption proposed by, for instance, Hoffmann and Stoecker (2004). According to this assumption, response-effects facilitate the chunking process of responses leading to longer subsequences (e.g. Stöcker & Hoffmann, 2004). As already stated, this assumption fits with the findings of increased implicit learning effects. In combination with the assumptions of Tubau et al. (2007), one could argue that the response-effect sequence enhanced the shift from stimulus-based to plan-based performance (Tubau et al., 2007). This assumption might be in line with our findings of Experiment 2b that the introduction of the response-effect sequence increased the acquisition of explicit knowledge in both conditions. However, it does not explain why, in Experiment 1b, learning the tone-effect sequence in advance facilitated the shift to plan-based performance.

Thus, at least the findings of Experiment 1b seem to best fit with the assumptions of the UEH (Frensch et al., 2003). According to this account, the sudden emergence of the response-location sequence might have led the participants to experience a strange feeling of sense of agency as they now could (voluntarily) produce the formerly perceived effects by their responses. This unexpected feeling triggers searching processes, whose content then becomes consciously available (Esser & Haider, 2017b; Haider & Frensch, 2005, 2009; Scott & Dienes, 2008; Whittlesea, 2002, 2004). This might explain why only the participants with predominantly explicit knowledge in the experimental conditions of Experiment 1b and 2b showed a steep decrease of RTs in the test-phase. As soon as they became aware of the response- location sequence they could speed-up their responding (e.g., Haider et al., 2011). It is conceivable that also in the control conditions the sudden emergence of the response-effect sequence might have surprised the participants which, in turn, had triggered search processes. Yet, the descriptive separate analyses of the implicit and explicit subgroups suggest that the participants in the control group might have needed some extra time to recognize the response-effect sequence. Overall, our results are in accord with the findings of Esser and Haider (2017a), as they once again showed that the response-effects boosted primarily the development of explicit knowledge.

As also mentioned in the introduction section, R-E-sequence learning is closely linked to the assumptions that response or action-effects play an important role for voluntary actions, as for instance the ideomotor principle (Greenwald, 1970; James, 1890; Prinz, 1992, 1997; Shin et al., 2010). According to the two-stage model proposed by Elsner and Hommel (2001), bi-directional associations between actions and their effects are built at a first stage when randomly executed actions lead to sensory perceivable changes in the environment. At the second stage, these associations can be used then for goal-directed behavior. Taken this fixed order of the formation of bidirectional associations between actions and effects for granted, this model does not seem suitable to explain our findings in its entirety. However, it seems reasonable to argue that the two stages overlap in processing (e.g., Drost, Rieger, Brass, Gunter & Prinz, 2005). Considering this assumption, the model could explain our finding that a fixed sequence of effects can be learned in advance. The implicitly acquired knowledge about this sequence then allows anticipating of the following effects. When a contingent motor sequence is added, the respective response- or action-effect associations are formed. These associations in turn facilitate learning the sequence of response-locations.

An alternative model for the building of response- or action-effect associations is proposed by Ziessler and Nattkemper (2002; Nattkemper et al., 2010; Ziessler et al., 2004). In this model, goals are cognitively represented in form of desired effects, which can activate a suitable motor program to achieve the respective goal. Anticipated effects are initially abstract and become stepwise specified with practice. If the desired and the anticipated effect differ, the motor program is modified. In sum, Ziessler and colleagues (Nattkemper et al., 2010; Ziessler & Nattkemper, 2002; Ziessler et al., 2004) suppose that effect-anticipation and action-planning occur in parallel. To account for our results of Experiment 1b and 2b, one could assume that during the acquisition phase the implicitly learned fixed effect sequence should lead to an anticipation of subsequent effects. Yet, as long as the responses were randomly mapped to the effects, no suitable motor program could be activated. Since effect-anticipation and action-planning occur in parallel, anticipation of subsequent effects might be accompanied by a permanent testing of a suitable motor program. If then the response sequence is mapped to the sequence of effects, the system might very fast learn which action fits the desired goal. This in turn, might have fostered the learning of the response sequence.

Taken together, in addition to the findings of Esser and Haider (2017a), the findings of our experiments provide evidence that response-effects facilitate the development of explicit knowledge about a response-location sequence in a SRTT. Moreover, this is also true for effects that were initially not contingently mapped to the responses. A sudden sense of agency, which is generated by the introduction of a correlated response-sequence, seems to be a suitable manipulation to provoke the experience of an unexpected event and, in turn, to trigger searching processes. However, this sense of agency seems to be stronger in case of salient effects, like, for instance, tones. Thus, the relation between salience and effectiveness of response-effects for the development of explicit knowledge in an implicit learning task seems to be a crucial subject for further research.

7. Study 2: The interplay between unexpected events and behavior in the development of explicit knowledge in implicit sequence learning

Abstract

Some studies in implicit learning investigate the mechanisms by which implicitly acquired knowledge (e.g. learning a sequence of responses) becomes consciously aware. It has been suggested that unexpected changes in the own behavior can trigger search processes, of which the outcome then becomes aware. A consistent empirical finding is that participants who develop explicit knowledge show a sudden decrease in reaction times, when responding to sequential events. This so called RT-drop might indicate the point of time when explicit knowledge occurs.

We investigated whether an RT-drop is a precursor for the development of explicit knowledge or the consequence of explicit knowledge. To answer this question, we manipulated in a serial reaction time task (SRTT) the timing of long and short stimulus onset asynchronies (SOA). For some participants the different SOAs were presented in blocks of either long or short SOAs, while for others the SOAs changed randomly. We expected the participants who were given a blocked presentation to express an RT-drop because of the predictable timing. In contrast, randomly changing SOAs should hamper the expression of an RT-drop.

We found that more participants in the blocked-SOA condition than in the random-SOA condition showed an RT-drop. Furthermore, the amount of explicit knowledge did not differ between the two conditions. The findings suggest that the RT-drop does not seem to be a presupposition to develop explicit knowledge. Rather, it seems that the RT-drop indicates a behavioral strategy-shift as a consequence of explicit knowledge.

Introduction

Implicit learning is a fundamental process in humans. It is said to occur without any intention to learn and, albeit debatable (Shanks, 2006), to lead to knowledge usually not consciously accessible. It enables humans to adapt to regularities inherent in the environment. Examples of implicit learning processes are learning the mother tongue or social behavior in children (Cleeremans, 2008, 2011).

One frequently used paradigm to investigate implicit learning is the Serial Reaction Time Task (SRTT; Nissen & Bullemer, 1987). In this task, the participants see a number of marked locations on the screen (often four or six locations) which are mapped to spatially corresponding keys. In every trial, an asterisk appears at one of the locations and the participants' task is to press the corresponding key as fast as possible. Unbeknownst to the participants, the positions of the asterisk and thus the to-be-pressed keys follow a regular sequence. To assess implicit learning, the regular sequence is replaced by a random sequence after several blocks of training and is re-introduced thereafter. The usual finding is that the reaction times (RTs) increase when the regular sequence is replaced by a random sequence and decrease when the regular sequence is re-introduced. Despite of this performance decrement, the participants are often unable to name the sequence. This led many researchers to conclude that learning in the SRTT is implicit and does not necessarily lead to explicit knowledge (Nissen & Bullemer, 1987). A common finding is, however, that a few participants indeed show at least some explicit knowledge about the practised sequence.

Given that there are participants who acquire explicit knowledge in an implicit learning situation, an important question arises: Which cognitive mechanisms underlie this interplay between implicit and explicit knowledge. One possibility to account for the acquisition of conscious knowledge in an implicit learning situation is to assume that learning leads to a gradual strengthening of the acquired representations (Cleeremans & Jiménez, 2002).

A different proposal is made by the Unexpected-Event Hypothesis (UEH; Frensch et al., 2003; Esser et al., 2021). The central assumption of the UEH is that, during SRTT training, participants acquire unconscious knowledge about the sequence inherent in the task. This unconscious knowledge is not assumed to become conscious by mere strengthening of the underlying representation. Rather, implicit knowledge is assumed to lead to consciously perceivable behavioral changes, such as the feeling of fluency or the urge to respond even though the stimulus is not entirely processed. Since implicit knowledge is unconscious, such changes are likely to be unexpected and might violate participants' expectations about their task performance (Koriat, 2012; Whittlesea, 2004). Therefore, if a participant detects such an unexpected behavioral change, it will likely trigger an explicit search process for the causes of this experienced change. This search processes in turn might lead to the detection of the regular sequence and thus to conscious sequence knowledge. Thus, the central assumption of the UEH is that an explicit representation of the sequence results from a second inferential process, rather than only from strengthening the unconscious representation. A second assumption of the UEH is derived from the Global Workspace theory (Dehaene & Nacchache, 2001; Marti & Dehaene, 2017). According to the Global Workspace theory, a conscious representation results from bottom-up strengthening, on the one hand, and, on the other hand, from a neurological "ignition", a sudden, strong top-down activation of a vast variety of cortical and subcortical regions (Dehaene & Naccache, 2001; Del Cul et al., 2009). Taking this into account, it is assumed in the UEH that the insight into the sequence occurs abruptly in an all-or-none manner (Esser et al., 2021).

The assumptions of the UEH fit well with similar assumptions that are proposed by the Event Segmentation Theory (EST; Zacks et al., 2007) about the human experience of events. According to the EST, active neuronal processing causes memory representations of events that lead to a sense of the current moment and in sum build the content of working memory (the event model). These memory representations enhance perception as well as prediction of future events. Analogous to the UEH, Zacks et al. assume that processing is always monitored by error detection processes. If an unexpected event occurs that does not fit with the predictions, a prediction error is detected. Different to the UEH, the EST does not assume that search processes are started as a consequence of the detected prediction error. In contrast, the EST assumes that the current event model will be updated by building a new set of memory representations. Thus, the EST describes more globally how changes in the perception of events might occur, whereas the UEH focuses more on the special situation of insight processes in implicit learning.

In the last years, the UEH gained support from various studies. For example, Haider and Frensch (2005, 2009) tested the role of premature responses (responses that are entered before the stimulus is entirely encoded) as an instantiation of unexpected events. Their findings suggest that increasing the probability of such premature responses either through long RSIs or through computer-generated premature responses led to more explicit knowledge.

Rünger and Frensch (2008) generated unexpected events by disrupting the sequence within an SRTT experiment. After some training blocks, the participants were transferred to either a different sequence, a random sequence, or a dual task situation. The findings revealed that only the interruption through a different sequence led to increased amounts of explicit knowledge. They concluded that merely interrupting the participants, as was the case when a dual task situation was introduced, was not sufficient to increase the amount of explicit knowledge. Instead, only interruptions that do not disrupt search processes lead to higher amounts of reportable knowledge.

Recently, Esser and Haider (2017) provided further evidence that mere strengthening of the underlying sequence representation is not sufficient to explain the emergence of explicit sequence knowledge. They trained participants with a sequence interspersed with deviants. These deviants were either presented in a blocked or a random order. Even though the number of sequence trials was identical in both conditions, the participants in the blocked order condition developed more explicit knowledge than the participants in the random order condition. According to the UEH, the participants in the blocked order condition could experience an unexpected variation in fluency while performing the task. This violation of expectancies led to attribution processes and thus to the development of explicit knowledge. By contrast, the participants in the random order condition might have experienced the task as constantly fluent and therefore might not have experienced a violation of their expectations.

In sum, these studies fit with the UEH by showing that introducing discrepancies between the expected and the actual behaviour increases the amount of explicit knowledge. Furthermore, most of these findings cannot be explained by mere strengthening of the implicit knowledge because the amount of training was kept constant in all mentioned studies.

Another line of empirical studies focuses on the second assumption of the UEH that explicit knowledge occurs abruptly in an all-or-none manner. These studies found high correlations between the ability to report the entire sequence in a post-experimental knowledge test, and an abruptly occurring decrease in RTs during the training phase (Frensch et al., 2003; Haider & Rose, 2007). This sudden decrease of RTs, the so called RT-drop, is thought to indicate a qualitative shift of the processing strategy from stimulus-driven to top-down driven processing (Haider & Frensch, 2005; Haider et al., 2005; Haider et al., 2011; Rose, et al., 2010; Wessel et al., 2012). If participants explicitly know the entire sequence, they do not need to process the imperative stimulus in order to produce the required response (Haider et al., 2011; Koch, 2007; Tubau & López-Moliner, 2004; Tubau et al., 2007).

Going beyond the findings of behavioural experiments, Rose et al. (2010) as well as Wessel et al. (2012) conducted fMRI and EEG studies within a sequence learning paradigm. They analysed the data time locked to the individual appearance of the RT-drop. The findings revealed that the RT-drop was accompanied by a strong increase in neuronal activity (fMRI) and in high-frequency coupling (EEG) between distant brain areas (prefrontal, parietal, and occipital). Similar results came from Schuck and colleagues (2015), who investigated strategy-shifts in a rather simple task and found also that the participants abruptly speeded up responding when they became aware of a more efficient strategy than that they had used before. Right before this changing point, Rose et al. (2010) as well as Schuck et al. (2015) found increased activity in the medial prefrontal cortex (for similar results see Lawson et al., 2017). These activation patterns are associated with the transition from unconscious to conscious representations (Del Cul et al., 2009).

Given the results from the neurophysiological studies as well as the high correlations between the occurrence of RT-drops and verbal reports, it seems a reasonable conclusion that the RT-drop might define the point in time when explicit knowledge occurs. The findings suggest that, at this point in time, the participants change their processing strategy from stimulus-driven to top-down (Haider et al., 2011; Schuck et al., 2015; Tubau & López-Moliner, 2004; Tubau et al., 2007; Wessel et al., 2012).

However, the exact role of the RT-drop for the development of conscious knowledge about the sequence has been unclear, yet. On the one hand, the RT-drops may occur as a side-effect of having developed a consciously accessible representation of the sequence. On the other hand, they might be a precursor for conscious awareness about the sequence. Shedding light on this issue will help to better understand how explicit knowledge and performance are linked. Furthermore, knowing about the exact role of RT-drops might provide some methodological implications for further investigations.

Overview of the Current Study

The goal of the current study was to investigate the specific role of RT-drops for the development of explicit knowledge. In the above mentioned studies, the occurrence of the RT-drop was highly correlated with the ability to verbally report the entire sequence (Haider & Frensch, 2005; Haider et al., 2011; Haider & Rose, 2007; Wessel et al., 2012). Together with the neuroimaging studies, this suggests that the RT-drop might indicate the point in time when an explicit representation of a given rule in an implicit learning paradigm arises. However, it has never been tested experimentally, whether the RT-drop really indicates the transition from implicit to explicit knowledge or whether it only indicates the strategy-change that results from having acquired explicit knowledge.

If the RT-drop were a precursor for the development of a conscious representation of the sequence, preventing the participants from changing their strategy should reduce the amount of verbally reportable knowledge (Cleeremans, 2006; Haider & Frensch, 2009). By contrast, if an RT drop indicates the translation of an otherwise acquired explicit representation into performance, preventing participants from changing their strategy should not have any effect on the development of explicit knowledge per se but only on the behavioural expression of what has been learned explicitly.

For this purpose, we conducted an SRTT with a first-order 6-elements motor sequence and manipulated the likelihood that an RT-drop would occur. This was done in two ways: First, we used two different stimulus onset asynchronies (SOAs) between the presented locations on the screen (the trial onset) and the appearance of the target stimulus, a short (200ms) and a long (800ms) one. These two SOAs were presented in either a blocked order (a series of 60 trials of each of the respective SOAs in each training block; blocked-SOA condition), or randomly (random-SOA condition). Second, responses were only valid if they were entered after the target onset to hinder participants from responding prematurely (Haider & Frensch, 2009). Furthermore, we used a rather short response window of 500ms starting with the target onset. This was done to force the participants to wait – in particular, when the SOA was long – for the appearance of the target and then to immediately enter their response. Consequently, with increasing practice, these long SOAs should lead to a strong urge to respond right before the stimulus appears (see, e.g., Grosjean et al., 2001). If the stimulus then occurs and corresponds to the anticipated response, this should come along with a surprise. In an implicit learning situation, the participants are not informed about the existence of a sequence. Thus, knowing the response in advance to the presentation of a target should likely violate the participants' expectations. This surprising experience together with not being aware about the cause (the sequence), should, in turn, trigger explicit search processes in both conditions.

Therefore, our manipulations should not affect the experience of an unexpected event per se because all participants should experience a possibly surprising urge to respond after a long SOA. Rather, the manipulation of the SOA-order should affect the likelihood that the participants will detach their responding from the appearance of the target and thus will rely their performance on only the explicit sequence representation (Gaschler et al., 2019; Shin, 2008; Willingham et al., 1997). In the blocked-SOA condition, the two different SOAs are predictable such that the participants could anticipate the point in time when the target will occur. In contrast, the SOAs in the random-SOA condition vary accidentally making it rather impossible to anticipate the exact stimulus onset in a given trial (e.g., Niemi & Näätänen, 1981). Therefore, the participants in this latter condition are forced to respond after the target has been presented. This should reduce the likelihood that the participants will rely their performance on their explicit representation of the sequence, and thus, will not show an RT-drop (the switch from a data-driven to a top-down processing strategy). If the RT-drop were a precursor for conscious awareness about the sequence to develop, the amount of explicit knowledge within this condition should be reduced.

Methods

Participants

One-hundred forty-one volunteers (81 men) with a mean age of 28.29 years (*SD* = 9.74) participated in the experiment either for course credit or for payment (4€ respectively 3.75£). These volunteers were acquired online among students from the University of Cologne (*N* = 23) as well as German participants recruited on the online platform Prolific Academic (https://www.prolific.co; *N* = 118). The participants were randomly assigned to the blocked-SOA (*N* = 67) and random-SOA (*N* = 74) conditions.

The data from participants were excluded if their mean accuracy rate was lower than 60%. We used a liberal accuracy criterion because the fixed response window had increased the likelihood to miss a response. Thirteen participants in the blocked-SOA (with *M* = 34% misses (*SD* = 18%) and *M* = 30% errors (*SD* = 12%)) and twelve participants in the random-SOA conditions (with *M* = 24% misses (*SD* = 19%) and *M* = 36% errors (*SD* = 20%)) were excluded[11].

Materials

The experiment was programmed with Psychopy and uploaded as JavaScript in Pavlovia (https://pavlovia.org/). We used a variation of the standard SRTT from Nissen and Bullemer (1987). In our version of the SRTT, six black response squares (2 x 2 cm) appeared on a horizontal line in the middle of the screen. The response squares were spatially mapped to the six response keys Y, X, C, B, N, and M on a German QWERTZ-keyboard. In every trial, the picture of a mouse appeared as the target in one of the response squares and the participants were instructed to press the corresponding key as fast as possible to catch the mouse. The locations of the target and thus of the response keys followed a repeated 6-elements first order sequence (526341)[12]. After each trial, the participants were informed whether they responded too early or too late (“fail!”), whether their response was incorrect (“error!”) or whether they responded correctly and in the accurate timing (“correct!”). At the end of the experiment, the

[11] Of the remaining participants, four participants in the blocked-SOA condition and five participants in the random-SOA condition missed more than 15% of the responses. Furthermore, six participants in the blocked-SOA condition and four participants in the random-SOA condition responded incorrectly in more than 15% of all trials. Due to our liberal exclusion criteria, these participants were not excluded.

[12] For the purpose of our study, we decided to use this short first-order sequence. Using a more complex sequence, like, for instance, a second order sequence with more than six positions might reduce the likelihood that the participants acquire explicit knowledge about the entire sequence since the number of elements cannot be held concurrently in working memory increasing the difficulty to integrate the sequence.

participants were interviewed via socscisurvey (SoSci Survey GmbH. https://www.soscisurvey.de).

Procedure

The experiment started with computer-based instructions. Then, the participants in both conditions completed the SRTT that consisted of six blocks with 120 trials, each. After each block, the participants were allowed to take a short break.

Each trial started with the presentation of the six response squares on the screen. After either a short (200ms) or a long SOA (800ms), the target stimulus appeared at one of the six positions for 150ms. Starting with the target-onset, the participants had a response window of 500ms to press the respective key. Each trial ended with the feedback about the actual performance (fail, error, or correct), which was displayed for 300ms. After the feedback, the next trial started immediately with the presentation of the six response squares.

The participants in both conditions received 60 short and 60 long SOAs in each block. In the blocked-SOA condition, the participants started with the 60 short SOAs followed by the 60 long SOAs. For the participants in the random-SOA condition, the SOAs changed unpredictably from trial to trial.

After having finished the SRTT training, the participants completed 150 trials of a post-decision wagering task (Persaud et al., 2007) to assess their sequence knowledge. The wagering task is identical to the SRTT with the only exception that in 36 trials a question mark appears at one of the six positions instead of the target. In these trials, the participants had to guess the position of the next target by pressing the corresponding key. Afterwards, they were asked to place a high (50 Cent/Pence) or low (1 Cent/Penny) wager, depending on their confidence in the correctness of their response. The rationale behind this task is that the participants with predominantly implicit sequence knowledge should not be able to maximize their gains and, therefore, the correctness of their responses should not correlate with their wagers (Dienes & Seth, 2010). However, the participants with predominantly explicit knowledge are highly confident in the correctness of their responses and should place more high wagers when responding correctly. In the current experiment, the participants could win a maximum of 3€/2.5£. The task did not have any time restrictions and had a fixed SOA of 400ms. The participants were informed about the changes in this part of the experiment and about the possibility of the bonus payment. However, they were not informed about the existence of a motor sequence and did not receive any feedback about the correctness of their responses.

Finally, the participants were forwarded to an online-survey. Participants were asked whether they had noticed any sequence or whether they thought that there was no system at all. Afterwards, the participants were informed that a motor sequence has been built into the task and were asked to reproduce the complete sequence as best as they could.

RT-drop Analysis

The procedure for determining the RT-drop was similar to that of Haider and colleagues (2011; see, also, Wessel et al., 2012). The authors adapted the method from Haider and Rose (2007; Rose et al., 2010), which was developed to detect the RT-drop in a slightly different task, the number reduction task. The procedure is based on the assumption that sequential knowledge in a SRTT is formed in chunks (Schlaghecken et al., 2000). This means that not all transitions between two succeeding sequence elements drop at the same time. Rather, they may drop one after the other with an individual time delay. Based on this rationale, we initially calculated the RT-drop for each single transition of the sequence.

Calculation of the RT-drops of each single transition was as follows: First, we removed all erroneous and missing responses and replaced them by the mean of the preceding and the subsequent trials. Second, we computed a median filter per transition with a lag of 3 subsequent repetitions to reduce noise in the data. Third, a minimum function was computed (Haider & Rose, 2007). That is, the value of this function changed only when the current median RT was shorter than the previous one. In all other cases, the value of this function remained unchanged. This produces a monotonously decreasing RT-function for each sequence transition. Fourth, to define the RT-drop for each single SRTT transition, we computed confidence intervals including all remaining transitions (as long as they have no diagnosed RT-drop). For each transition, we defined the point in time when the minimum function of this respective transition dropped below the confidence interval of the combined minimum function of the remaining transitions. If an RT-drop was detected for one transition, no further RT-drop was searched for this single transition. In addition, the RT of the minimum function in the trial immediately before the diagnosed drop of this transition was used to compute the combined minimum functions for the RT-drop detection of the remaining transitions. Fifth, to avoid the detection of too early pseudo-RT decreases due to familiarization with the task, we defined a threshold of 200ms. Only if the minimum function of one transition fell below the confidence interval of the combined minimum functions of the remaining transitions and below the threshold of 200 ms, it was defined as an RT-drop. Finally, the overall RT-drop for one individual participant was defined as that point in time when at least 4 out of 6 sequence transitions had dropped (Haider et al., 2011). Figure 7 shows an example for the RT-drop detections for the six transitions for one participant.

It is important to note, that using the minimum function to determine the RT-drops within participants should reduce the likelihood that the probably more varying RTs in the random-RSI-order condition affect the detection of RT-drops.

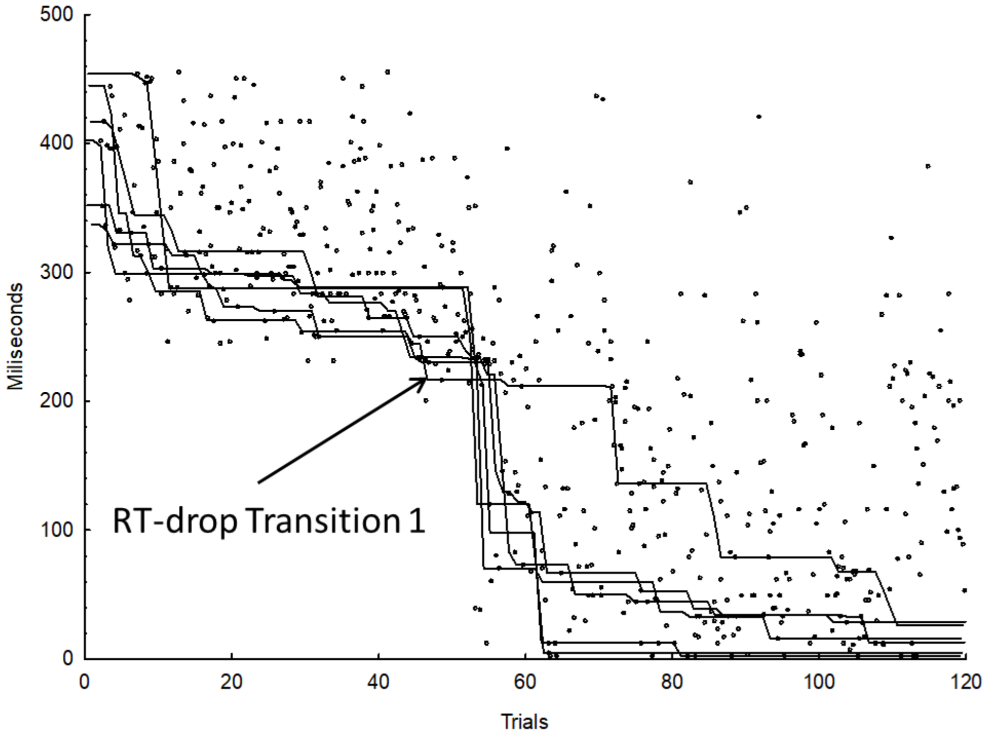

Fig. 7. Minimum functions per transition (lines) and raw RTs (dots) as illustration for the RT-drop detection. The detected RT-drop for Transition 1 is marked as an example.

Results

For each participant and each block, median RTs and mean error rates were computed separately. For the RTs, erroneous trials (7.5% of all data) and missings (8.85% of all data) were excluded. As the response window was restricted, there were no RT-outliers to exclude. For all statistical analyses, an α-level of 5% was adopted. If Mauchly's test of sphericity reached significance, Greenhouse Geisser corrected p-values are reported together with the original degrees of freedom.

Error rates and latencies in the SRTT

Figure 8 shows the mean error rates per block and condition separately for long and short SOAs. For the mean error rates as dependent variable, a 2 (SOA-condition: blocked-SOA, random-SOA) x 2 (SOA-length: short, long) x 6 (Block) ANOVA with repeated measures for the last two factors revealed a significant main effect of Block ($F(5,570) = 73.39$, $p < .001$, $\eta_p^2 = 0.39$). As can be seen from Figure 2, the error rates in both SOA-conditions decreased over time. The main effect of SOA-condition ($F(1,114) = 1.15$, $p = .285$, $\eta_p^2 = 0.00$) as well as the interaction between SOA-condition and Block ($F(5,570) = 1.17$, $p = .319$, $\eta_p^2 = 0.01$) were not significant. The main effect of SOA-length ($F(1,114) = 26.59$, $p < .001$, $\eta_p^2 = 0.18$) and also the two-way interactions between SOA-condition and SOA-length ($F(1,114) = 17.93$, $p < .001$, $\eta_p^2 = 0.13$), between SOA-length and Block ($F(5,570) = 3.15$, $p = .014$, $\eta_p^2 = 0.02$), as well as the triple interaction ($F(5,570) = 7.94$, $p < .001$, $\eta_p^2 = 0.06$) were significant. Figure 2 shows that the participants made more errors when the SOA was short than when it was long. This difference was more pronounced for the participants in the blocked-SOA condition.

Separate post-hoc tests for each SOA-condition show that the difference in error rates for long and short SOAs was significant for the participants in the blocked-SOA condition (short SOA: M = 9.59%; SD = 6.49%; long SOA: M = 6.67%; SD = 5.40%; $t(53) = 6.36$, $p < .001$), but not for the participants in the random-SOA condition (short SOA: M = 7.28%; SD = 4.02%; long SOA: M = 6.99%; SD = 5.10%; $t(61) = 0.68$, $p = 0.498$). Furthermore, the two SOA-conditions did not differ with regard to the long SOAs (t(114) = 0.32, p = 0.746). However, considering only the short SOAs, the error rates were significantly higher in the blocked-SOA condition than in the random-SOA condition (t(114) = 2.34, p = 0.020). This was particularly the case at the beginning of training.

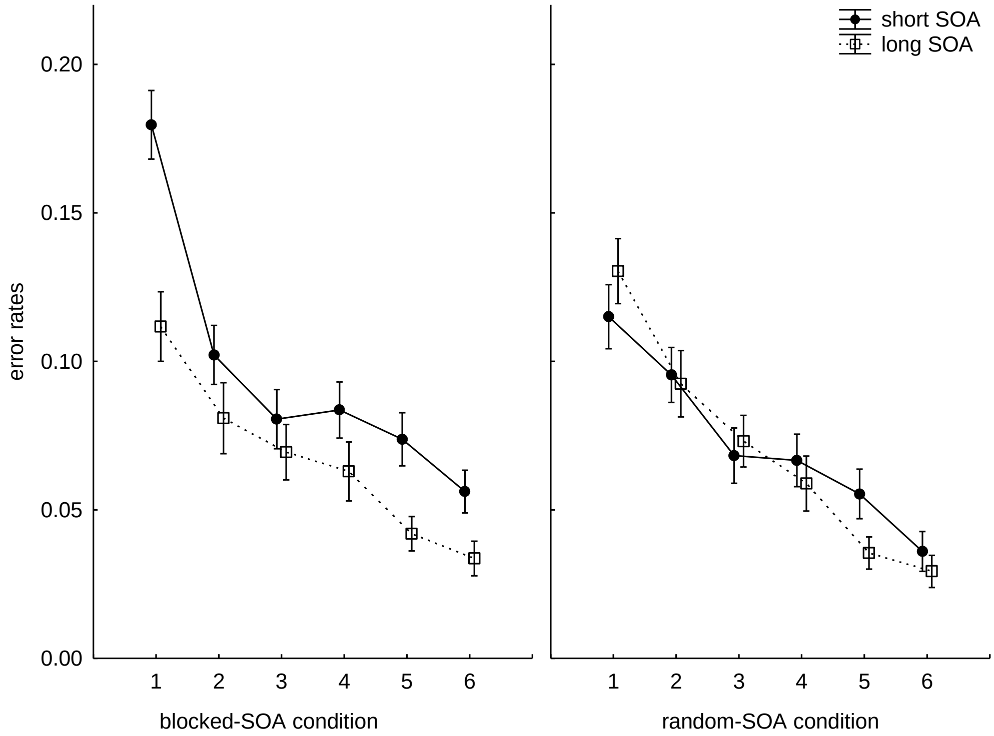

Fig. 8. Mean percent error rates per block and SOA. Error bars reflect the standard errors of the means.

The mean RTs are shown in Figure 9[13]. The 2 (SOA-condition: blocked-SOA, random-SOA) x 2 (SOA-length: short, long) x 6 (Block) ANOVA with RT as dependent variable showed a significant main effect of Block ($F(5,570) = 164.34$, $p < .001$, $\eta_p^2 = 0.59$). Here, the main effect of SOA-condition ($F(1,114) = 5.61$, $p = .019$, $\eta_p^2 = 0.04$) as well as the SOA-condition x Block interaction ($F(5,570) = 7.85$, $p < .001$, $\eta_p^2 = 0.06$) were significant. As can be seen from Figure 3, the RTs of the participants in both SOA-conditions decreased over time. Furthermore, the participants in the blocked-SOA condition responded overall faster than the participants in the random-SOA condition. This RT-difference increased with training. In addition, the main effect of SOA-length ($F(1,114) = 24.15$, $p < .001$, $\eta_p^2 = 0.17$), the two-way interactions between SOA-condition and SOA-length ($F(1,114) = 20.42$, $p = .012$, $\eta_p^2 = 0.15$),

[13] We additionally analyzed if the RT data in the random-SOA condition differ for trials in which the SOA repeated versus changed. We found slightly longer RTs in short SOA trials when these followed a long compared to following a short SOA-trial. However, for two reasons it is rather unlikely that these slower RTs in the random-SOA condition should have affected our following RT-drop analysis. First, we median filtered the RTs in both SOA-conditions with a lag of 3 subsequent trials. This median filter cancels RT-outliers and reduces noise in the data. Second, on the basis of this median filtered data, we computed the minimum functions. This implies that the values of this function changed only when the current median RT was shorter than the previous one. This minimum function then built the basis for the RT-drop analysis. Thus, the 25% of potentially slower RTs due to the long-short SOA sequences likely should not have affected the minimum functions.

and between SOA-length and Block ($F(5,570) = 3.10$, $p = .008$, $\eta_p^2 = 0.02$), as well as a the triple-interaction ($F(5,570) = 3.64$, $p = .002$, $\eta_p^2 = 0.03$) were significant. As the data in Figure 9 suggest, the participants in the blocked-SOA condition did not show any systematic difference between long and short SOAs. However, the participants in the random-SOA condition were significantly faster with the long than with the short SOAs. The interaction between SOA-length and block and the triple interaction did not indicate a clear trend.

To further investigate the differences between the SOA-conditions, we also conducted separate 2 (SOA-length: short, long) x 6 (Block) ANOVA for the two SOA-conditions. For the blocked-SOA condition, the main effect of Block ($F(5,265) = 87.17$, $p < .001$, $\eta_p^2 = 0.62$) as well the interaction between SOA and Block ($F(5,265) = 3.31$, $p = .010$, $\eta_p^2 = 0.05$) reached significance (for SOA ($F(1,53) < 1$). For the random-SOA condition, the main effect of SOA-length ($F(1,61) = 104.45$, $p < .001$, $\eta_p^2 = 0.63$) and for Block ($F(5,305) = 74.57$, $p < .001$, $\eta_p^2 = 0.55$) were significant, while the SOA-length x Block interaction was not ($F(5,305) < 1$). The data in Figure 3 shows that, for the random-SOA condition, the difference between the long and short RSI-lengths remain rather constant over all blocks.

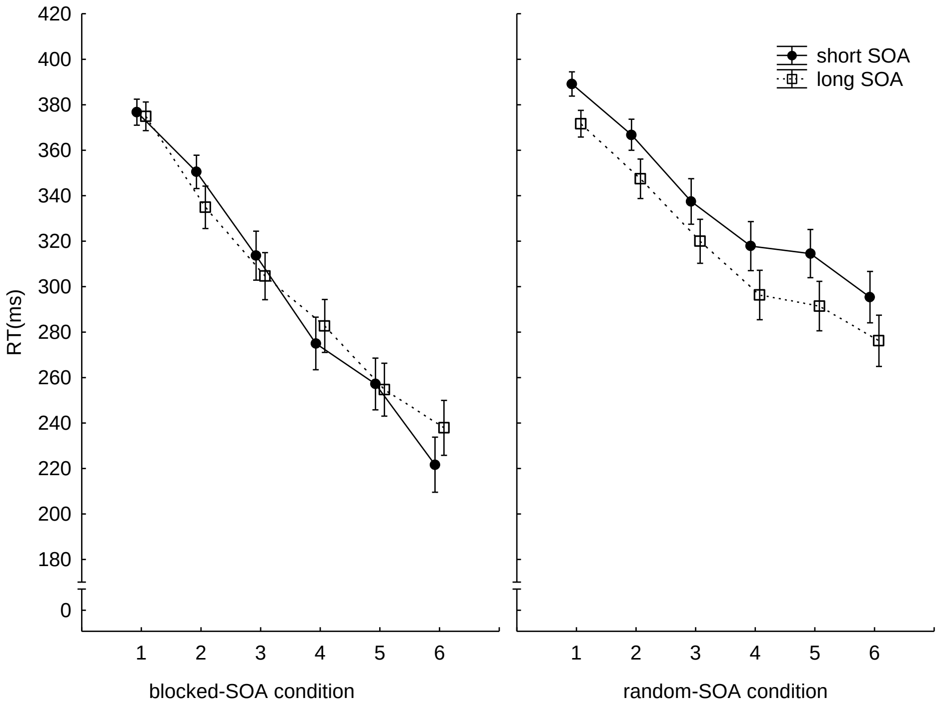

Fig. 9. Mean RTs per block and SOA. Error bars reflect the standard errors of the means.

To sum up, our manipulation of the order of the SOAs affected both, the error rates and the RTs. In the blocked-SOA condition, the participants reacted faster than the participants in the random-SOA condition. However, optimizing response speed seems to come at costs, since error rates for the short SOA trials were higher in the blocked-SOA condition.

In contrast, in the random-SOA condition, the pattern of results suggests slower responses for short than for long SOA trials. Whether this was accompanied by an increased accuracy is ambiguous. On the one hand, the error rates did not systematically differ between the two SOA-types. On the other hand, the participants in the random-SOA condition made fewer errors than the participants in the blocked-SOA condition. Since a baseline is missing, it remains unclear whether the error rate was increased in the blocked-SOA condition or reduced in the random-SOA condition.

Overall, the error rates and the RT findings together suggest that our SOA-manipulation was successful. The slower responses in the random-SOA compared to the blocked-SOA condition suggest that the exact timing of the responses was more difficult in the former SOA-condition. However, a not mutually exclusive explanation for the overall shorter RTs in the blocked-SOA condition might be that more participants exhibited an RT-drop in this condition than in the random-SOA condition, since these participants were better able to adapt and optimize their performance according to the respective SOAs. Thus, the next important question concerns the effect of our SOA manipulation on the rate of RT-drops.

RT-drop

Based on the procedure described in the methods section, we conducted the RT-drop analyses for each individual participant. In the blocked-SOA condition, 39 participants (72%) showed an RT-drop as defined above, whereas only 26 participants (41%) in the random-SOA condition did so. This difference between the two conditions was significant ($\chi^2(1) = 10.75$; $p = .001$, $\phi = 0.30$). Thus, our manipulation of the SOA arrangement not only affected RTs and error rates, but also the probability that the participants in the random-SOA condition exhibited an RT-drop. On the basis of these findings we investigated our main hypothesis concerning the relationship between RT-drops and explicit knowledge.

Post-Decision Wagering Task (PDWT)

In the PDWT, as the main test for explicit knowledge, only the wagering trials were considered. For these trials, we computed percent correct responses as a measurement of overall acquired sequence knowledge. In addition, we also computed the percent correct responses under high (correct|high) versus low wagers (correct|low) as a measurement of the subjective certainty to have responded correctly and thus as a measure of explicit knowledge. The results are presented in Table 7.

Table 7. Percent correct responses and certainty judgements for the blocked-SOA and random-SOA condition. The respective standard deviations are presented in brackets (C|H refers to correct| high; C|L to correct |low).

SOA-Condition	correct (%)	C|H	C|L	N total	N explicit (%)
Blocked-SOA	76.28(24.19)	79.30(28.00)	49.02(33.09)	54	37(68.51)
Random-SOA	75.31(25.97)	77.49(28.97)	46.41(35.86)	62	38(61.29)

Concerning the percent correct responses, the participants in both SOA-conditions had more sequence knowledge than expected by chance (chance level: 20%; blocked-SOA condition: $t(53) = 7.21$, $p < .001$; random-SOA condition: $t(61) = 7.03$, $p < .001$). More important, the two SOA-conditions did not differ with regard to their amount of correct responses ($t(114) = 0.20$, $p = .835$). To further bolster this insignificant difference between the two SOA-conditions, we computed a Bayes-Analysis with JASP (JASP Team, 2018; Morey & Rouder, 2015) for t-tests (Rouder et al., 2009). We calculated a Bayes-Factor BF_{01} to test the null hypothesis (H_0) to be favored over the alternative hypothesis (H_1). We used the Cauchy distribution as prior distribution with a Cauchy-scale of .0707 for the calculation of Bayes-Factors for t-tests with JASP (JASP Team, 2018). Our Bayes factor was $BF_{01} = 4.96$, which, according to Jeffreys (1939/1961), provides evidence for the null hypothesis.

The certainty judgements served to test the amount of explicit sequence knowledge. A 2 (SOA-condition: blocked-SOA, random-SOA) x 2 (Certainty Judgment: correct|high, correct|low) ANOVA with percent correct responses as dependent variable was conducted. A significant main effect of Certainty Judgment ($F(1,114) = 61.40$, $p < .001$, $\eta_p^2 = 0.35$) was found. As can be seen in Table 1, the participants in both SOA-conditions placed more high than low wagers for correct responses. Importantly, there was again no main effect of SOA-condition (F(1,114) = 0.25, p = .617, η_p^2 = 0.00) and no SOA-condition x Certainty Judgement interaction (F(1,114) = 0.01, p = .919, η_p^2 = 0.00). Thus, the two SOA-conditions did not differ regarding their confidence in the correctness of their responses and therefore in their amount of explicit knowledge. Additionally, we computed a Bayes-Analysis for ANOVA designs (Rouder et al., 2012) to validate the finding that the two SOA-conditions did not differ in the amount of their explicit knowledge. The calculated Bayes-Factor was BF_{01} = 5.72 indicating that the null hypothesis can be favored over the alternative hypothesis.

In addition, we identified those participants who had acquired predominantly explicit sequence knowledge. The participants were classified as possessing predominantly explicit knowledge when they responded correctly in at least 70% of all wagering trials and placed primarily high wagers for correct answers. The number of participants with explicit knowledge in each SOA-condition is shown in Table 7. Again, the two SOA-conditions did not differ (χ^2(1) = 0.66; p = .416, ϕ = 0.075).

In sum, the results of the PDWT show that the participants in both conditions had acquired a comparable amount of sequence knowledge. Furthermore, this knowledge seems to be primarily explicit as the participants placed high wagers on their correct responses and thus, showed a high confidence in the correctness of their responses. Approximately 70% of the participants in each condition were classified as having full explicit knowledge about the motor sequence. Thus, even though there were fewer RT-drops in the random-SOA condition, the amount of explicit knowledge about the sequence did not differ from that of the blocked-SOA condition.

However, two alternative explanations concerning the discrepancy between the amount of explicit knowledge and the occurrence of an RT-drop in the random-SOA condition need to be discussed. First, it might be that our procedure used to diagnose the RT-drops was not a valid method. Second, it is conceivable that the PDWT did not validly assess what the participants knew about the sequence. To handle these considerations, we conducted two post hoc analyses.

Post-hoc Analyses

In order to validate our RT-drop analysis, we divided both SOA-conditions into two subgroups, one with the participants with (drop-subgroup) and one with the participants without an RT-drop (no-drop subgroup). In addition, we tested whether the RT-drops were indicative for the amount of explicit knowledge. If our procedure to diagnose the RT-drops was valid, we should find a stronger RT-decrease in both drop subgroups. Furthermore, the drop subgroups should show more explicit knowledge about the sequence. The mean RTs for the first analysis are shown in Figure 10.

We conducted a 2 (SOA-condition: blocked-SOA, random-SOA) x 2 (Subgroup: RT-drop, no-drop) x 6 (Block) ANOVA with RT as dependent variable and found a significant main effect of Block ($F(5,560) = 136.31$, $p < .001$, $\eta_p^2 = 0.54$) and of Subgroup ($F(1,112) = 54.95$, $p < .000$, $\eta_p^2 = 0.32$). The main effect of SOA-condition ($F < 1$) as well as the SOA-condition x Block interaction ($F(5,560) = 1.91$, $p = .090$, $\eta_p^2 = 0.01$) were not significant. The interaction between Subgroup and Block ($F(5,560) = 23.46$, $p < .001$, $\eta_p^2 = 0.17$) the interaction between SOA-condition and Subgroup ($F(1,112) = 5.86$, $p = .017$, $\eta_p^2 = 0.04$) as well as triple-interaction ($F(5,560) = 2.30$, $p = .043$, $\eta_p^2 = 0.02$) were significant. Figure 4 shows that within both SOA-conditions the RTs of the participants in the drop subgroup decreased faster with practice than in the non-drop subgroup. This difference was greater in the blocked-SOA condition than in the random-SOA condition. However, the qualitative picture was rather similar in both SOA-conditions. This suggests that the RT-drop analysis reliably identified the participants with an RT-drop in both SOA-conditions. It is rather unlikely, that the rate of miss-classified participants in the random-SOA condition is much higher than that in the blocked-SOA condition. If this had been the case, we should have found no significant RT-differences between the two subgroups within the random-SOA condition.

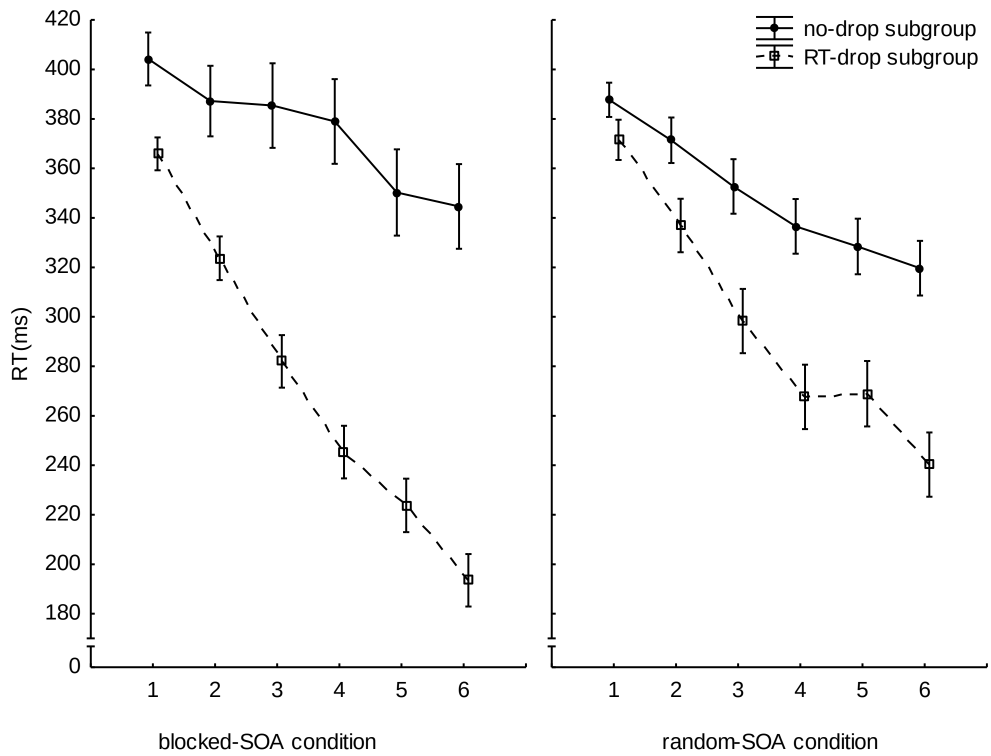

Fig. 10. Mean RTs per block for the participants with and without RT-drop. Error bars reflect the standard errors of the means.

To test whether the RT-drops are indicative for explicit sequence knowledge (Haider & Rose, 2007), we also analyzed, how many participants in the two subgroups (drop, no-drop) within both SOA-conditions (blocked-SOA and random-SOA) showed predominantly explicit knowledge. The results are shown in Table 8. In both SOA-conditions, approximately 80% of the participants in the drop subgroup possessed predominantly explicit knowledge. A chi²-test conducted separately for the two SOA-conditions confirmed that this rate was significantly higher than that in the two non-drop subgroups (blocked-SOA condition: $\chi^2(1) = 11.92$; $p = .006$, $\phi = 0.46$; random-SOA condition: $\chi^2(1) = 7.16$; $p = .007$, $\phi = 0.33$). Only the effect size in the random-SOA condition was smaller than that in the blocked-SOA condition. Thus, in both SOA-conditions the RT-drop seems to be indicative for explicit knowledge to occur.

Despite of this difference between the two subgroups, the number of participants with predominantly explicit knowledge in the no-drop subgroup of the random-SOA condition is higher than that in the blocked-SOA condition. This latter finding is in line with our assumption that the SOA manipulation affected the RT- drop rate in the random-SOA condition, but not the development of explicit knowledge.

Table 8. Participants with predominantly explicit knowledge in the blocked-SOA and random-SOA condition

	SOA-Condition			
	Blocked-SOA		Random-SOA	
Subgroup	N total	N explicit (%)	N total	N explicit (%)
Drop	39	32(82.05)	26	21(80.77)
No-drop	15	5(33.33)	36	17(47.22)

Taken together, the two post-hoc analyses that we conducted to verify our RT-drop analysis revealed that the behavioural patterns of the two subgroups (drop, no-drop) were qualitatively similar within both SOA-conditions (blocked-SOA, random-SOA). The main difference between the two SOA-conditions was that, probably due to the higher number of participants with RT-drops in the blocked-SOA condition, the performance differences between the two subgroups were more pronounced within this condition. Given these results, it seems justified to assume that our main finding of more RT-drops in the blocked-SOA condition and comparable explicit knowledge in both SOA-conditions, was not due to a lesser likelihood to correctly diagnose the RT-drops in the random-SOA condition.

Post experimental Interview

For our second question, whether the PDWT validly assessed participants' explicit knowledge, we analyzed the data of our post experimental interview. For each participant, we counted how many transitions of the sequence were reproduced correctly. We then used the method from Rünger and Frensch (2008) to compute individual knowledge scores. The calculation of these scores takes into account how likely it is to report a certain number of transitions by mere guessing. The individual knowledge score is calculated from 100 minus the guessing probability. The score is thus adjusted for the guessing probability and shows how much the participant knows about the sequence. The participants in the blocked-SOA condition had a mean score of 92.22 (SD = 13.02). In the random-SOA condition, one participant did not complete our online survey. From the remaining 61 participants, the mean score was 90.37 (SD = 13.33). Comparably to the results from the PDWT, there was no difference between the two SOA-conditions ($t(113) = 0.74$, $p = .456$). In addition, the Bayes-Analyses for t-tests confirmed the plausibility of the Null-hypothesis ($BF_{01} = 3.93$).

To sum up, the data of the post experimental interview revealed a similar picture as the data of the PDWT: The participants in the two conditions did not differ concerning their amount of explicit sequence knowledge.

Discussion

The goal of our study was to better understand the concrete role of RT-drops for the development of explicit knowledge in implicit sequence leaning. The initial point of the study was twofold: First, former findings have shown high correlations between RT-drop and verbal report (Haider & Rose, 2007; Haider & Frensch, 2009; Haider et al., 2011; Schwager et al., 2012). Second, the neuroimaging studies (Rose et al., 2010; Schuck et al., 2015; Wessel et al., 2012) revealed that the RT-drops were accompanied by a strong increase in neuronal activity in certain brain areas. These two findings led to the reasonable conclusion that the RT-drop reflects the point in time when a person gains explicit sequence knowledge. However, based on these correlations, it remains open whether RT-drops are a precursor for or a consequence of the generation of explicit sequence knowledge.

In the current study, we therefore focused on the manipulation of the ease of producing such an RT-drop during the SRTT training. If the RT-drop were a precursor for sequence awareness, increasing the difficulty to exhibit such an RT-drop should reduce the development of explicit knowledge. If, however, the RT-drop were a behavioral consequence of awareness, the manipulation should affect the probability of RT-drops but not the amount of explicit knowledge.

The study yielded three main results: First, task performance was affected by our manipulation. While the blocked order of the SOAs led to overall shorter RTs as well as higher error rates for the short than for the long SOAs, the random order of SOAs led to shorter reaction times with long SOAs than with short SOAs. This pattern of results is in line with the assumption that the participants in the blocked-SOA condition were better able to anticipate the onset of the next target and thus could optimize their speed of responding easier than the participants in the random-SOA condition. Second, significantly more participants in the blocked-SOA condition than in the random-SOA condition showed an RT-drop. Both findings together confirm the success of our SOA manipulation since it suggests that the participants in the random-SOA condition were less able to predict the onset of the target and, probably due to this uncertainty, showed less RT-drops. Third, despite of the finding that the two conditions differed in the number of participants who showed an RT-drop, the amount of explicit sequence knowledge was comparable in both conditions. This was also confirmed by the verbal knowledge assessed in the post-experimental interviews. Thus, our SOA-manipulation did not affect the acquisition of (explicit) sequence knowledge.

Taken together, these findings complement the previous studies on RT-drops in one important point. Up to now, the high correlations between RT-drops and the amount of explicit knowledge suggested, together with the neuroimaging findings, that the RT-drop might indicate the occurrence of conscious sequence representations. Yet, since brain activity was analyzed depending on the point in time when the RT-drops appeared, it was only possible to analyze the brain activity of the participants with an RT-drop. These analyses could therefore not answer the question whether the RT-drop is the point in time when awareness occurs (as a precursor of the development of explicit knowledge) or whether it is the point in time when the participants decide to use this knowledge (as a consequence of the development of explicit knowledge). The results only showed that the brain activity was increased right before an RT-drop occurred and that the participants with an RT-drop had more explicit sequence knowledge than those without an RT-drop.

In the current study, we also found such a difference in the amount of explicit knowledge. In both conditions, more participants who showed an RT-drop developed predominantly explicit sequence knowledge than those who did not show an RT-drop. However, this correspondence was larger for the participants in the blocked-SOA condition than for those in the random-SOA condition. Thus, the current comparison of the blocked and the random-SOA conditions suggest that the emergence of a conscious representation of the sequence is, at least, partly independent of RT-drops. This independence between conscious sequence knowledge and its usage in performance allows for two alternative conclusions: First, it is conceivable that the randomly presented SOAs in the random-SOA condition increased the variability of the RTs making it harder to detect an RT-drop with our RT drop analysis. This would be a methodological problem and would confound our results. Second, the difference in the number of participants with an RT-drop between the two conditions is reliable and reflects a result of our manipulation.

Our post-hoc analyses seem to support the second conclusion. The results show that the behavioral pattern of the participants with and without RT-drop was rather similar for the two conditions. In the blocked-SOA condition as well as in the random-SOA condition, the participants who showed an RT-drop responded faster than the participants who did not. This difference increased over practice, albeit this increase was stronger in the blocked-SOA than in the random-SOA condition. However, if the varying SOAs had reduced the probability to detect an RT-drop in the random-SOA condition, we should have found a small or even no difference between these two subgroups. Thus, the results seem to be better in line with the assumption that the difference between the two conditions in the number of participants with an RT-drop is a result of our manipulation.

Regarding our main question, namely whether the RT-drop is a precursor for or a consequence of the development of explicit knowledge, our findings indicate that the RT-drops are a consequence of the development of explicit knowledge rather than a precursor. Our findings suggest that an RT-drop marks the point in time when participants rely on their explicit sequence representation and shift from stimulus-driven to plan-based performance (Haider & Rose, 2007; Hoffmann & Koch, 1997; Nattkemper & Prinz, 1997; Tubau & López-Moliner, 2004; Tubau et al., 2007). It does not seem to indicate the development of an explicit representation itself. Once developed, most of the participants in the blocked-SOA condition relied on this knowledge and used it to optimize their performance producing the sudden decrease of reaction times. In the random-SOA condition, by contrast, far lesser participants showed this sudden decrease despite having developed explicit knowledge. A possible explanation is that these participants may have gained awareness in the same manner as the participants in the blocked-SOA condition. They just were not willing or were not able to rely on this newly generated representation because the unpredictable order of the SOAs together with the short response window made this strategy rather error-prone. This might have prevented them to change the strategy from stimulus-driven to plan-based performance (Tubau & López-Moliner, 2004; Tubau et al., 2007).

This explanation seems plausible in the light of other studies that investigated voluntary strategy shifts. Haider and colleagues (2005) showed that a switch to a new strategy is not an automatic consequence of task processing. Rather, a strategy shift occurs voluntary and controlled when the new strategy is efficient for further task performance (Gaschler et al., 2019). Further studies have shown that interfering manipulations in learning tasks often do not impair learning per se but the expression of what has been learned. In particular, impaired timing in form of, for instance, changed RSI patterns (Miyawaki, 2006; Willingham et al., 1997) or uncertainty due to random RSIs (Tubau et al., 2007), seem to influence task performance although the learning process itself seemed unaffected.

Our findings also fit well with the assumptions of the UEH (Frensch et al., 2003). According to the first assumption, we assume that in both conditions, a strong urge to respond due to the locked response window might have triggered search processes that led to the development of explicit sequence knowledge. Furthermore, the sudden decrease of reaction times, which was found in our study, is in line with the notion that sequence awareness occurs abruptly in an all-or-nothing manner. The former interpretation of the RT-drop was that it shows the point in time when explicit knowledge occurs (e.g., Rose et al., 2010). We assume that this interpretation is still correct. However, due to our results, it has to be refined by the assumption that the usage of the newly generated explicit sequence representation for performance seems to underlie voluntary control.

Despite the promising results, two limitation of our study are noteworthy. First, it was conducted online and, thus, we could not control for all confounding factors in the same manner as in lab-studies. Furthermore, it was impossible for the participants to contact the investigator in the case that they did not understand the instructions. However, the wordings in our instructions have been adapted to the situation. Moreover, neither the data nor the post experimental questionnaires gave any hints that the participants had problems to correctly perform the task. Furthermore, there already are studies in implicit sequence learning that were conducted online (e.g., Sævland & Norman, 2016) and which results were comparable to lab experiments. Lastly, potential confounding variables should have affected both the blocked-SOA and random-SOA condition.

Another limitation might be that we did not include a random block at the end of the training to assess sequence learning. Consequently, we could not distinguish sequence learning from pure practice effects on the basis of our training results. We decided not to use a random block since we were particularly interested in the relation between the expression of RT-drops and the development of explicit sequence knowledge. Such an interspersed random block bears the danger to attenuate this relation because it might hinder the participants to shift their strategy late in practice. Furthermore, the sudden decrease of RTs and the explicit sequence knowledge expressed in the PDWT cannot be explained by pure practice effects.

Finally, explicit sequence knowledge was assessed only after the training phase and we did not measure the exact point in time when the participants became aware of the sequence. Thus, it is conceivable that the RT-drop and sequence awareness developed independently. In this case, our conclusion that the RT-drop is a behavioral consequence of explicit knowledge might have been incorrect. However, Rose and colleagues (2010) had assessed explicit knowledge online during training. The participants were interrupted after the expression of fast responses and were asked about the reasons for these fast responses. Their findings suggest that three fast responses indicate explicit knowledge and are expressed right before an RT-drop. Taken these finding of Rose and colleagues (2010) into account, the conclusion, that the RT-drop is a behavioral consequence of awareness, is reasonable.

In sum, we conclude that the focus of our study on the functional role of RT-drops in implicit sequence learning and the results of the manipulation in our experiment allows the interpretation of the RT-drop as indicative for a strategy shift from stimulus-based to plan-based responding rather than the generation of an explicit knowledge representation, per se. This explicit representation comes along with the possibility to control whether one will or will not use this explicit knowledge to optimize performance.

8. General Discussion

The term implicit learning refers to an unconscious learning process that takes place without any intention to learn. In implicit learning paradigms as, for instance, the SRTT participants show behavioral learning effects but often do not express any explicit knowledge about what has been learnt. This empirical finding led to the conclusion that learning in this case might have resulted in implicit knowledge (e.g., Nissen & Bullemer, 1987).

Research in implicit learning led furthermore to the development of various theories about how explicit knowledge arises in implicit learning situations and paradigms such as the SRTT. There exists supporting evidence for both, the single system as well as the multiple system theories as two classes of theories. Thus, among the scientists who deal with this topic in their research, there is still little consensus about the underlying processes and the characteristics of implicit and explicit knowledge representations.

The aim of this thesis was contribute to this debate about how explicit knowledge develops in implicit learning. After the introduction of the important theories of consciousness and implicit learning, a review and two empirical studies were presented, that shed light on different core aspects of the main research question.

The review focused on the theoretical view of how conscious representations are developed in an implicit learning situation. The goal of this review was to introduce and to discuss two different points of views on how conscious knowledge arises, namely the strengthening account (Cleeremans & Jiménez, 2002; Cleeremans, 2006, 2008, 2011, 2014, 2020) as well as the UEH (Frensch et al., 2003).Furthermore, these two viewpoints were related to two classical theories of consciousness, the GWT (Baars, 1997, 2003) and the HOTTs (Lau & Rosenthal, 2011; Rosenthal, 1985).

As a result of this discussion, both theories of consciousness have advantages and disadvantages in explaining how explicit knowledge arises in implicit learning. Furthermore, the assumptions of the UEH (Frensch et al., 2003) fit well with the GWT (Baars, 1997, 2003). In contrast, the strengthening account (Cleeremans, 2008, 2011) rather fits with the assumptions of the HOTTs (Lau & Rosenthal, 2011; Rosenthal, 1985). Finally, the review concluded with a suggestion of using aspects of the GWT and the HOTT to consider the advantages and disadvantages of both theories of consciousness in explaining how explicit knowledge in implicit learning arises. In combination with the UEH, problems might be solved that arise when only the GWT or HOTTs are applied. A metacognitive model about the own behavior in a given situation is assumed that is always adapted to the current task through the comparison of the expected and experienced behavior. As in implicit learning the

behavioral changes are based on FO-performance, they might not fit the current model and violate the expectations one has about the behavior. This then might trigger an evaluation process whether a new mental model is needed for the current situation.
The first empirical study was conducted to investigate single assumptions of the UEH in further detail. According to the UEH implicit knowledge representations never can become consciously aware. Rather, a secondary explicit search process is needed, which is triggered by an unexpected change in one's own behavior, which in turn leads to the generation of explicit knowledge representations. Here, the role of response-effects as unexpected events and thus, for triggering the development of explicit sequence knowledge in implicit sequence learning was examined while the strengthening of knowledge representations was held constant. We investigated in four experiments whether participants can learn a pure effect sequence implicitly when the effects are not mapped to any responses and whether the development of explicit sequence knowledge is triggered when afterwards a contingent response-effect sequence is introduced that should lead to an experience of an unexpected sense of agency (Beck, et al., 2017; Haggard et al., 2002; Moore et al., 2009) and thus to searching processes.

The main results of Study 1 can be summarized as followed. First, we were able to show that participants can learn a pure sequence of task-irrelevant auditory or visual effects even when the effects are not contingently mapped to any responses. Second, learning such an effect sequence in advance and the experience of a sudden sense of unexpected agency and the possibility to produce this effect-sequence can trigger the development of explicit knowledge about a later introduced contingent response-location sequence. In contrast, the implicit learning system seems to be unaffected by the effect sequence. Furthermore, the results of Study 1 indicate that the development of explicit knowledge occurred in an all-or-none manner. After excluding the participants with predominantly explicit knowledge, the differences between the experimental and the control conditions almost entirely disappeared. This was true particularly for salient auditory effects. The visual effects led to weaker results. Third, only participants with predominantly explicit knowledge who experienced the unexpected sense of agency showed a strong decrease of reaction times when the response-location sequence was introduced. In sum, our results provide evidence that the manipulation of unexpected events with response-effects foster the development of explicit knowledge about a response-location sequence in a SRTT while the implicit learning system remains unaffected. The findings of the first study fit well with the assumptions of the UEH (Frensch et al., 2003) whereas other accounts, as for instance the strengthening account by Cleeremans

(2006, 2008) fail to explain the results in whole. According to the UEH, the sudden sense of agency of the effect-sequence, triggered by the sudden emergence of the contingent response-location sequence, might have triggered searching processes whose content became aware (Esser & Haider, 2017b; Haider & Frensch, 2005, 2009; Scott & Dienes, 2008; Whittlesea, 2002, 2004). The steep decrease of reaction times for those participants who experienced the sudden sense of agency as an unexpected event and developed explicit sequence knowledge is in line with the assumption of the UEH that explicit knowledge arises suddenly in an all-or-nothing manner. The strengthening account (Cleeremans, 2006, 2008) cannot explain these results, as this account postulates a graded transition from implicit to explicit knowledge. The further investigation of the so called RT-drop (e.g., Haider et al., 2011) was the purpose of Study 2.

Particularly, Study 2 focused on the suddenness of insight processes that is behaviorally often characterized by a sudden decrease of RTs (RT-drop; Haider & Frensch, 2005, 2009). Here, the functional role of such an RT-drop was investigated and thus, the question whether the RT-drop is a precursor for the development of explicit sequence knowledge or whether it is a behavioral consequence of awareness. To investigate this question, we conducted a SRTT with a repeated response-location sequence and realized two conditions. For participants in the blocked-SOA condition, the timing in the SRTT and the occurrence of succeeding targets was predictable. In contrast, the timing was unpredictable for participants in the random-SOA condition who were therefore not able to anticipate the occurrence of the succeeding targets (e.g., Niemi & Näätänen, 1981). The creation of an unexpected event was equal for the participants of the two conditions. As part of the analysis of reaction times, RT-drops were determined separately for each participant of both conditions.

Our study led to two main results. We found that, first, the two conditions differed in their number of participants who showed an RT-drop, as more participants in the blocked-SOA condition compared to the random-SOA condition showed an RT-drop. Second, the amount of explicit sequence knowledge was equal for the two conditions. Thus, in the blocked-SOA condition, those participants who developed explicit sequence knowledge seem to have used this knowledge for the performance in the SRTT. In contrast, in the random-SOA condition, a comparatively larger proportion of participants without RT-drop nevertheless developed explicit sequence knowledge. These results led us conclude that the development of explicit knowledge in implicit learning seems to be independent of the expression of an RT-drop. Rather, the RT-drop seems to be a behavioural consequence of explicit knowledge. Thus, participants who develop explicit knowledge may be able to decide whether they change their strategy from stimulus-driven to top down (Haider & Rose, 2007; Hoffmann & Koch, 1997; Nattkemper & Prinz, 1997; Tubau & López-Moliner, 2004; Tubau et al., 2007), given that a strategy shift is conductive for the performance of the actual task (e.g., Gaschler et al., 2019).

The interpretations of the RT-drop as precursor as well as behavioral consequence of explicit sequence knowledge fit both with the assumption of the UEH (Frensch et al., 2003), because both interpretations do not contradict the suddenness of the occurrence of explicit knowledge representations. For our interpretation of the RT-drop as behavioral consequence of explicit sequence knowledge, one can assume that once explicit knowledge is developed, the RT-drop defines the point in time when this explicit knowledge is used for a voluntary strategy shift. This strategy shift occurs only when the new strategy is efficient for the task performance.

In sum, the results of the review and the two empirical studies support the assumptions of the UEH (Frensch et al., 2003) in different aspects from a pure theoretical view to the more detailed investigations of single predictions. In doing so, the review and the two studies contribute to the research question about the development of explicit knowledge in implicit learning.

The review discussed two prominent views in implicit learning with the result that the UEH seems to be a promising theory in explaining a broad variety of results, particularly in combination with the GWT and HOTT as two classical theories of consciousness. This new theoretical framework might build a good starting point for the reconsideration of former particularly inconsistent results as well as for further research questions.

The first study focused on the empirical investigation of the trigger for the transition from unconscious to conscious knowledge representations and the second study the behavioral consequences. The findings of these two empirical studies indicate that explicit knowledge is based on a learning system that differs from the implicit learning system and that explicit knowledge arises suddenly in an all-or-none manner.

The assumption of one single learning system, as proposed by, for instance, Shanks and colleagues (Shanks & Johnstone, 1999; Shanks & St. John, 1994) or Perruchet and Vinter (2002) seems to be inadequate to explain the results of the two empirical studies. The multiple system view that supposes the existence of at least two different learning systems that work independently from each other seems to provide a more plausible explanation.

The radical plasticity hypothesis (Cleeremans, 2011), as one of the accounts that pursue the idea that the implicit and explicit learning system differ from each other, takes the perspective that explicit meta-representations develop similar to implicit FO-representations and thus, gradually through an increase of representational quality (strength, stability in time, and distinctness) during learning. However, the results of Study 1 and Study 2 show that explicit knowledge arises in an all-or nothing manner rather than graded. Furthermore, in Study 1, knowledge differences of the respective conditions completely vanished after excluding participants with full explicit knowledge. In addition, in the first study, the strengthening of knowledge representations was held constant, while only the experience of an unexpected event was manipulated. This manipulation affected only the amount of participants with full explicit knowledge, while the implicit learning system was unaffected. Furthermore, both studies showed that participants who develop explicit knowledge show a steep decrease of reaction times. This suddenness, which was investigated in more detail in Study 2, also speaks against a gradual development of explicit knowledge representations.

The UEH (Frensch et al., 2003) can explain both aspects of the results, as its assumptions predict both, the suddenness and the all-or-non principle of explicit knowledge. In both studies, an unexpected event (either the unexpected sense of agency or the unexpected urge to respond) might have triggered searching processes that led to the development of explicit knowledge. The implicit learning system remained respectively unaffected. The development of explicit knowledge was further accompanied by a strong decrease of reaction times.

As a result of the review, the UEH in combination with aspects of the GWT (Baars, 1997, 2003) and HOTT (Lau, 2008; Lau & Rosenthal, 2011) might be promising in explaining the development of explicit knowledge in implicit sequence learning. The results of Study 1 and Study 2 can also be discussed in the context of the theories of consciousness. Regarding the GWT, the unexpected event in both studies might have gained access to the GWS and triggered searching processes that led to sudden insight, similar to the sudden ignition that is proposed by the GWT. Regarding the HOTTs, one may assume that during implicit learning, FO-knowledge is built that affects task performance. Furthermore, the task performance is constantly monitored by the explicit system. If the actual task performance, that is affected due to implicit learning processes, does not fit with the expected task performance, searching processes are triggered that lead to the development of explicit knowledge.

In sum, the review and the two empirical studies contributed on different levels, from a theoretical view to the detailed investigation of single predictions, to the debate how explicit knowledge arises in implicit learning and to what extend the UEH (Frensch et al., 2003) provides a reasonable explanation for this question.

Two points seem to be promising for further investigations in upcoming studies. First, it seems worthwhile to take a closer look on the unexpected events. In both experimental studies (Study 1 and Study 2), some of the participants in the respective conditions did not develop explicit knowledge even though the task in the respective condition was manipulated to create an unexpected event. Furthermore, in Study 1, participants in the control conditions should not have experienced an unexpected event but some of them showed explicit knowledge. Here, it could be conceivable that the occurrence of unexpected events might be individually experienced by the participants and somewhat lesser controllable as desired by the experimenter. Thus, every participant might have an individual reason to experience an unexpected event. If this were the case, such unexpected events should be difficult to manipulate in order to draw reliable conclusions from this manipulation. Here, research should focus on the understanding of the formation of such unexpected events, whether the unexpected events are equal for every individual and how it is possible to deliberately create unexpected events for the tested participants to draw reliable conclusions.

A second research field might contain the voluntary shift in performance after the development of explicit sequence knowledge. Here, it could be further investigated, when and how participants who have developed explicit sequence knowledge decide to use the acquired knowledge or under which conditions it is more efficient to maintain the old strategy (e.g., Gaschler et al., 2019).

References

Abrahamse, E. L., Jimenéz, L., Verwey, W. B. & Clegg, B. A. (2010). Representing serial action and perception. *Psychonomic Bulletin & Review, 17,* 603–623. doi:10.3758/PBR.17.5.603

Abrahamse, E. L., van der Lubbe, R. H. J., Verwey, W. B., Szumska, I., & Jaśkowski, P. (2012). Redundant sensory information does not enhance sequence learning in the serial reaction time task. *Advances in Cognitive Psychology, 8,* 109-120. doi:10.2478/v10053-008-0108-y

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48,* 35-44. doi: 10.1037/0003-066X.48.1.35

Baars, B. J. (1997). In the theatre of consciousness: Global Workspace Theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies, 4*(4), 292-309.

Baars, B. J. (2003). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience? *Progress in Brain Research, 150,* 45–53. doi: https://doi.org/10.1016/S0079-6123(05)50004-9

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences, 7*(4), 166-172. doi: 10.1016/S1364-6613(03)00056-1

Baars, B. J. (2005). Global workspace theory of consciousness: Towards a cognitive neuroscience of human experience? *Progress in Brain Research, 150*, 45-53. doi: 10.1016/S0079-6123(05)50004-9

Baars, B. J., Franklin, S., & Ramsøy, T. Z. (2013). Global workspace dynamics: cortical “binding and propagation” enables conscious contents. *Frontiers in Psychology, 4*. doi: 10.3389/fpsyg.2013.00200

Baddeley, A.D. and Hitch, G.J. (1974) Working memory. In Bower, G.H. (Ed.). *The Psychology of Learning and Motivation*. pp. 47–89, Academic Press.

Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition in signal-detection theoretic models. *Psychological Methods, 18,* 535-552. doi: 10.1037/a003326

Beck, B. Di Costa, S. & Haggard, P. (2017). Having control over the external world increases the implicit sense of agency. *Cognition, 162,* 54-60. doi:10.1016/j.cognition.2017.02.002

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience, 7,* 356-366. doi: 10.3758/CABN.7.4.356

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 642-652. doi: 10.1037//0033-295X.108.3.624

Block, N. (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral Brain Sciences, 30,* 481–548. doi:10.1017/S0140525X07002786

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences, 23,* 754–768. doi: 10.1016/j.tics.2019.06.009

Chalmers, D. J. (1995). Facing up to the problems of consciousness. *Journal of Consciousness Studies, 2*, 200-219. doi: 10.1093/acprof:oso/9780195311105.003.0001

Changeux, J. P., & Dehaene, S. (1989). Neuronal models of cognitive functions. *Cognition, 33*, 63-109. doi: 10.1016/0010-0277(89)90006-1

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences, 36,* 181-204. doi: 10.1017/S0140525X12000477

Clark, A. & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language, 8*, 487-519. doi: 10.1111/j.1468-0017.1993.tb00299.x

Cleeremans, A. (2006). Conscious and unconscious cognition: A graded, dynamic perspective. In Q. Jing, M.R. Rosenzweig, G. d'Ydewalle, H. Zhang, H.-C. Chen, & K. Zhang (Eds.). *Progress in Psychological Science around the World. Volume 1 Neural, Cognitive and Developmental Issues. Proceedings of the 28th International Congress of Psychology (1st ed.).* (pp. 401–418). London: Psychology Press. doi: 10.4324/9780203783122

Cleeremans, A. (2008). Consciousness: the radical plasticity thesis. In R. Banerjee & B. K. Chakrabarti (Eds.). *Models of Brain and Mind. Physical, Computational and Psychological Approaches* (pp. 19-33). Amsterdam: Elsevier.

Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology*, *2*: 86, doi: 10.3389/fpsy.2011.10086

Cleeremans, Axel (2014). Connecting conscious and unconscious processing. *Cognitive Science, 38,* 1286–1315. doi: 10.1111/cogs.12149

Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., Vuillaume, L., & de Heering, A. (2020). Learning to be conscious. *Trends in Cognitive Sciences. 24*. 112-123. doi: 10.1016/j.tics.2019.11.011.

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: Psychology Press.

Cleeremans, A., & Sarrazin, J.-C. (2007). Time, action, and consciousness. *Human Movement Science, 26*, 180–202. doi: 10.1016/j.humov.2007.01.009

Cleeremans, A., Timmermans, B., and Pasquali, A. (2007). Consciousness and metarepresentation: a computational sketch. *Neural Networks, 20,* 1032–1039. doi: https://doi.org/10.1016/j.neunet.2007.09.011

Collins, A. G., & Frank, M. J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychological Review, 120*, 190-229. doi: 10.1037/a0030852

Cosmelli, D., & Preiss, D. D. (2014). On the temporality of creative insight: A psychological and phenomenological perspective. *Frontiers in Psychology*, *5*, Article 1184. doi: 10.3389/fpsyg.2014.01184

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*, 51-57. doi: 10.1177/0963721409359277

Curran, T. (2001). Implicit learning revealed by the method of opposition. *Trends in Cognitive Sciences, 5, 503-504*. doi:10.1016/S1364-6613(00)01791-5

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70,* 200-227. doi: 10.1016/j.neuron.2011.03.018

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1-37. doi:10.1016/S0010-0277(00)00123-2

Dehaene, S., Changeux, J.-P., & Naccache, L. (2011). The global neuronal workspace model of conscious access: From neuronal architectures to clinical applications. In: S. Dehaene & Y. Christen (eds.). *Characterizing consciousness: From cognition to the clinic?* Heidelberg, Berlin: Springer Verlag. pp. 55–84.

Dehaene, S., Sergent, C., & Changeux. J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 8520–8525. doi: 10.1073/pnas.1332574100

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences, 10*, 204-211. doi: 10.1016/j.tics.2006.03.007

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness and could machines have it? *Science, 358*, 48-492. doi: 10.1126/science.aan8871

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain, 132*, 2531–2540. doi: 10.1093/brain/awp111

Destrebecqz, A. & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review, 8,* 343–350. doi: https://doi.org/10.3758/BF03196171

Dienes, Z., Altmann, G. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1322-1338. doi:10.1037/0278-7393.21.5.1322

Dienes, Z. & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*, 735–808. doi: 10.1017/S0140525X99002186

Dienes, Z. & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge from judgement knowledge. *Psychological Research, 69*, 338-351. doi: 10.1007/s00426-004-0208-3

Dienes, Z. & 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*, 674–681. doi:10.1016/j.concog.2009.09.009

Dietrich, A., & Haider, H. (2017). A Neurocognitive Framework for Human Creative Thought. *Frontiers in Psychology, 7,* 2078. doi: 10.3389/fpsyg.2016.02078

Dreisbach, G. & Haider, H. (2009). How task representations guide attention: Further evidence for the shielding function of task sets. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 477–486. doi:10.1037/a0014647

Drost, U. C., Rieger, M., Brass, M., Gunter, T. C. & Prinz, W. (2005). Action-effect coupling in pianists. *Psychological Research, 69*, 233–241. doi:10.1007/s00426-004-0175-8

Eberhardt, K., Esser, S. & Haider, H. (2017). The building blocks of the implicit learning system. *Journal of Experimental Psychology,* doi:10.1037/xhp0000380

Elsner, B. & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology, 27*, 229–240. doi:10.1037//0096-1523.27.1.229

Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review, 67,* 279-300. doi:10.1037/h0041622

Esser, S. & Haider, H. (2017a). Action-effects enhance explicit sequential learning. *Psychological research*. doi:10.1007/s00426-017-0883-5

Esser, S. & Haider, H. (2017b). The emergence of explicit knowledge in a serial reaction time task: The role of experienced fluency and strength of representation. *Frontiers in Psychology*. doi:10.3389/fpsyg.2017.00502

Esser, S. Lustig, C. & Haider, H. (2021). What Triggers Explicit Awareness in Implicit Sequence Learning? Implications from Theories of Consciousness. *Psychological Research.* doi: 0.1007/s00426-021-01594-3

Fedor, A., Zachar, I., Szilágyi, A., Öllinger, M. de Vladar, H., & Szathmáry, E. (2017). Cognitive Architecture with Evolutionary Dynamics Solves Insight Problem. *Frontiers in Psychology, 8*. doi: 10.3389/fpsyg.2017.00427

Fleming, S. M. (2020). Awareness as inference in a higher-order state space, *Neuroscience of Consciousness*, *1*(2020), 1-9. doi:10.1093/nc/niz020

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review, 12*, 91-114. doi: 10.1037/rev0000045

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*. doi: 10.3389/fnhum.2014.00443

Franklin, S. (2001) Automating human information agents. In: Chen Z. and Jain L.C. (Eds.). *Practical Applications of Intelligent Agents*. Springer-Verlag, Berlin.

Frensch, P.A., Haider, H., Rünger, D., Neugebauer, U., Voigt, S. & Werg, J. (2003). Verbal report of incidentally experienced environmental regularity: The route from implicit learning to verbal expression of what has been learned. In: L. Jiménez (Ed.), *Attention and implicit learning* (pp. 335-366). Amsterdam: Benjamins.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, *12*, 13-18. doi:10.1111/1467-8721.01213

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews: Neuroscience, 11*, 127-138. doi: 10.1038/nrn2787

Galvin S.J., Podd J.V., Drga V., & Whitmore J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review, 10*, 843-76. doi: 10.3758/BF03196546

Gaschler, R., Frensch, P. A., Cohen, A. & Wenke, D. (2012). Implicit sequence learning based on instructed task set. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1389–1407. doi: 10.1037/a0028071

Gaschler R., Schuck N. W., Reverberi C., Frensch P. A., Wenke D. (2019). Incidental covariation learning leading to strategy change. *PLOS ONE 14*: e0210597. doi: 10.1371/journal.pone.0210597

Gaschler, R., Marewski, J. N., Wenke, D., & Frensch, P. A. (2014). Transferring control demands across incidental learning tasks – Stronger sequence usage in serial reaction task after shortcut option in letter string checking. *Frontiers in Psychology, 5*:1388. doi: 10.3389/fpsyg.2014.01388

Goschke, T., & Bolte, A. (2012). On the modularity of implicit sequence learning: Independent acquisition of spatial, symbolic, and manual sequences. *Cognitive Psychology, 65*, 284-320. doi: 10.1016/j.cogpsych.2012.04.002

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, *77*, 73-99. doi:10.1037/h0028689

Grosjean, M., Rosenbaum, D. A., & Elsinger, C. (2001). Timing and reaction time. *Journal of experimental psychology. General*, *130*, 256–272. doi:10.1037//0096-3445.130.2.256

Haggard, P., Clark, S. & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5,* 382–385. doi:10.1038/nn827

Haider, H., Eberhardt, K., Esser, S. & Rose, M. (2014). Implicit visual learning: How the task set modulates learning by determining the stimulus-response binding. *Consciousness and Cognition*, *26*, 145–161. doi:10.1016/j.concog.2014.03.005

Haider, H., Eberhardt, K., Kunde, A. & Rose, M. (2012). Implicit visual learning and the expression of learning. *Consciousness and Cognition, 22,* 82–98. doi:10.1016/j.concog.2012.11.003

Haider, H., Eichler, A., & Lange, T. (2011). An old problem: How can we distinguish between conscious and unconscious knowledge acquired in an implicit learning task? *Consciousness and Cognition*, *20*, 658–672. doi:10.1016/j.concog.2010.10.021

Haider, H., Esser, S. & Eberhardt, K. (2018). Feature codes in implicit sequence learning: perceived stimulus location transfer to motor response locations. *Psychological Research,* doi:10.1007/s00426-018-0980-0.

Haider, H. & French, P. A. (2005). The generation of conscious awareness in an incidental learning situation. *Psychological Research*, *69*, 399–411. doi:10.1007/s00426-004-0209-2

Haider, H., & Frensch, P. A. (2009). Conflicts between expected and actually performed behavior lead to verbal report of incidentally acquired sequential knowledge. *Psychological Research, 73,* 817-834. doi:10.1007/s00426-008-0199-6

Haider, H., Frensch, P. A., & Joram, D. (2005). Are strategy shifts caused by data-driven processes or by voluntary processes? *Consciousness and Cognition, 14*, 495-519. doi: 10.1016/j.concog.2004.12.002

Haider, H. & Rose, M. (2007). How to investigate insight: A proposal. *Methods, 42*(1), 49-57. doi: 10.1016/j.ymeth.2006.12.004

Hélie, S., & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review, 117*, 994–1024. doi: 10.1037/a0019532

Hoffmann, J., & Koch, I. (1997). Stimulus-response compatibility and sequential learning in the serial reaction time task. *Psychological Research*, *60*, 87-97. doi:10.1007/BF00419682

Hoffmann, J., Sebald, A. & Stöcker, C. (2001). Irrelevant response effects improve serial learning in serial reaction time tasks. *Journal of Experimental Psychology, 27*, 470–482. doi: 10.1037//0278-7393.27.2.470

Hommel, B., Müssler, J., Aschersleben, G. & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849–937. doi:10.1007/s00426-009-0234-2

Hoyndorf, A. & Haider, H. (2009). The “Not Letting Go” phenomenon: Accuracy instructions can impair behavioral and metacognitive effects of implicit learning processes. *Psychological Research, 73*, 695-706. doi: 10.1007/s00426-008-0180-4

Hunt, R. H., & 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*, 658–680. doi: 10.1037/0096-3445.130.4.658

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory &Language*, *30*, 513-541. doi: https://doi.org/10.1016/0749-596X(91)90025-F

James, W. (1890). *The principles of psychology*. New York, NY: Holt& Co.

Janacsek, K., Shattuck, K. F., Tagarelli, K. M., Lum, J., Turkeltaub, P. E., & Ullman, M. T. (2020). Sequence learning in the human brain: A functional neuroanatomical metaanalysis of serial reaction time studies. *NeuroImage, 207*, 116387. doi: 10.1016/j.neuroimage.2019.116387

JASP Team (2018). JASP (Version 0.9.0.1) [Computer software].

JASP Team (2020). JASP (Version 0.14.1) [Computer software].

Jeffreys, H. (1939/1961). *The Theory of Probability*, 1st/3rd Edn. Oxford, England: Oxford University Press.

Jiménez, L. & Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 236–259. doi: 10.1037//0278-7393.25.1.236

Keane, M. M., Gabrieli, J. D. E., Noland, J. S., & McNealy, S. I. (1995). Normal perceptual priming of orthographically illegal nonwords in amnesia. *Journal of the International Neuropsychological Society, 1*,425–433. doi: https://doi.org/10.1017/S1355617700000527

Keele, S. W., Mayr, U., Ivry, R., Hazeltine, E. & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110*, 316–339. doi: 10.1037/0033-295X.110.2.316

Kinder, A., & Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience, 13*, 648–669. doi: 10.1162/089892901750363217

Kinder, A. & Shanks, D. R. (2003). Neurophysiological dissociations between priming and recognition: A single-system connectionist account. *Psychological Review, 110*, 316–339. doi: https://doi.org/10.1037/0033-295X.110.4.728

Kinsbourne, M. (1996). What qualifies a representation for a role in consciousness? In J. D. Cohen & J. W. Schooler (eds.), *Scientific Approaches to the Study of Consciousness* (pp. 335-355). Hillsdale, NJ: Erlbaum.

Koch, I. (2007). Anticipatory response control in motor sequence learning: Evidence from stimulus–response compatibility. *Human Movement Science, 26*, 257–274. doi:10.1016/j.humov.2007.01.004

Koch, I., & Hoffmann, J. (2000). Patterns, chunks, and hierarchies in serial reaction-time tasks. *Psychological Research*, *63*, 22-35. doi:10.1007/PL00008165

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9,* 149-171. doi: 10.1006/ccog.2000.0433

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-325). New York, NY: Cambridge University Press.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119*, 80-114. doi: 10.1037/a0025648

Koriat, A. (2015). Knowing by doing: When metacognitive monitoring follows metacognitive control. In S. D. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger III (Eds.), *Remembering: Attributions, Processes, and Control in Human Memory: Essays in honor of Larry Jacoby* (pp. 185-197). Psychology Press.

Kouider, S., de Gardelle, V., Sackur, & J., Dupoux. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences, 14*, 301-307. doi: 10.1016/j.tics.2010.04.006

Kouider, S. & Faivre, N. (2017). Conscious and unconscious perception. In S. Schneider & M. Velmans (Eds.), *The Blackwell Companion in Consciousness*, 2nd Edition (pp. 855-864). Wiley-Blackwell

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 387–394. doi:10.1037//OO96-1523.27.2.387

Kunde, W., Koch, I. & Hoffmann, J. (2004). Anticipated action effects affect the selection, initiation, and execution of actions. *The Quarterly Journal of Experimental Psychology, 57A,* 87–106. doi:10.1080/02724980343000143

Kunde, W., Schmidts, C., Wirth, R. & Herbort, O. (2017). Action effects are coded as transitions from current to future stimulation: Evidence from compatibility effects in tracking. *Journal of Experimental Psychology: Human perception and performance, 43*, 477–486. doi:10.1037/xhp0000311

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences, 10*, 494-501. doi: 10.1016/j.tics.2006.09.001

Lau, H. C. (2008). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research, 168*, 35-48. doi: 10.1016/S0079-6123(07)68004-2

Lau, H. C., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences, 15*, 365-373. doi:10.1016/j.tics.2011.05. 009

Lau, H. C. & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *PNAS, 103*, 18763–18768. doi: https://doi.org/10.1073/pnas.0607716103

Lawson, R. R., Gayle, J. O., & Wheaton, L. A. (2017). Novel behavioral indicator of explicit awareness reveals temporal course of frontoparietal neural network facilitation during motor learning. *PloS one, 12*, e0175176. doi: 10.1371/journal.pone.0175176

Lepper, M., Massen, C. & Prinz, W. (2008). What to do and how to do it: Sequence learning of action effects and transformation rules. *Acta Psychologica, 128,* 139–152. doi:10.1016/j.actpsy.2007.12.001

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527. doi:10.1037/0033-295X.95.4.492

Lustig, C., Esser, S., & Haider, H. (2021). The interplay between unexpected events and behavior in the development of explicit knowledge in implicit sequence learning [Manuscript submitted for publication]. Department of Psychology, University of Cologne.

Lutz, N. D., Wolf, I., Hübner, S., Born, J., & Rauss, K. (2018). Sleep strengthens predictive sequence coding. *The Journal of Neuroscience, 38,* 8989–9000. doi: 10.1523/JNEUROSCI.1352-18.2018

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*, 422-430. doi: 10.1016/j.concog.2011.09.021

Maniscalco, B., & Lau, H. C. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neuroscience of Consciousness, 2016*(1). doi: 10.1093/nc/niw002

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications, 8*, 1955. doi: 10.1038/s41467-017-02079-x

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 350-364. doi: 10.1037/0278-7393.22.2.350

McNamee, D., & Wolpert, D. M. (2019). Internal Models in Biological Control. *Annual Review of Control, Robotics, and Autonomous Systems, 2*, 339–364. doi: 10.1146/annurev-control-060117-10520

Moore, J. W., Lagnado, D., Deal, D. C. & Haggard, P. (2009). Feelings of control: Contingency determines experience of action. *Cognition, 110*, 279–283. doi:10.1016/j.cognition.2008.11.006

Morey, R. D., & Rouder, J. N. (2015). BayesFactor (Version 0.9.11-3) [Computer software].

Miyawaki, K. (2006). The influence of the reponse-stimulus interval on implicit and explicit learning of stimulus sequence. *Psychological Research, 70*, 262–272. doi: 10.1007/s00426-005-0216-y

Nattkemper, D. & Prinz, W. (1997). Stimulus and response anticipation in a serial reaction task. *Psychological Research, 60*, 98–112. doi:10.1007/BF00419683

Nattkemper, D., Ziessler, M. & Frensch, P. A. (2010). Binding in voluntary action control. *Neuroscience and Biobehavioral Reviews, 34,* 1092–1101. doi:10.1016/j.neubiorev.2009.12.013

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences, 37*, 1-61. doi: 10.1017/S0140525X12003214

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89*, 133–162. doi:10.1037//0033-2909.89.1.133

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1-32. doi:10.1016/0010-0285(87)90002-8

Overgaard, M. (2003). On the theoretical and methodological foundations for a science of consciousness. *Journal of Anthropological Psychology, 13*, 6-31.

Pasquali, A., Timmermans, B., & Cleeremans, A. (2010). Know thyself: Metacognitive networks and measures of consciousness. *Cognition, 117,* 182-190. doi: 10.1016/j.cognition.2010.08.010

Pavlovia. https://pavlovia.org/ (accessed on 21 July 2020).

Perruchet, P. & Vinter, A. (1998). Learning and development: The implicit knowledge assumption reconsidered . In: M. Stadler & P. Frensch (Eds.). *Handbook of implicit learning*, Sage.

Perruchet, P. & Vinter, A. (2002). The self-organizing conscious. *Behavioral and Brain Sciences, 25*, 297–388. doi: 10.1017/S0140525X02000067

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 55A*, 485–503. doi: 10.1080/02724980143000451

Perruchet, P., Gallego, J. & Savy, I. (1990) A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology, 22,* 493–516. doi: 10.1016/0010-0285(90)90011-R

Persaud, N., McLeod, P, & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience, 10*, 257–261. doi:10.1038/nn1840

Persuh, M., LaRock, E., & Berger, J. (2018). Working memory and consciousness: The current state of play. *Frontiers in Human Neuroscience, 12*:27, doi: 10.3389/fnhum.2018.00078

Peters, M. A., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLife, 3*. doi: 10.7554/eLife.09651

Pfordresher, Q. P. (2003). Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 949–964. doi:10.1037/0096-1523.29.5.949

Pfordresher, Q. P. (2005). Auditory feedback in music performance: The role of melodic structure and musical skill. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1331–1345. doi:10.1037/0096-1523.31.6.1331

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review, 117,* 864-901. doi: 10.1037/a0019737

Prather, R. W. (2012) Implicit Learning of Arithmetic Regularities Is Facilitated by Proximal Contrast. *PLoS ONE, 7*: e48868. doi: 10.1371/journal.pone.0048868

Prinz, W. (1992). Why don't we perceive our brain states? *European Journal of Cognitive Psychology, 4,* 1-20. doi:10.1080/09541449208406240

Prinz, W. (1997). Perception and action planning *European Journal of Cognitive Psychology, 9,* 129-154. doi:10.1080/713752551

Prolific | Online participant recruitment for surveys and market research. https://www.prolific.co/ (accessed on 21 July 2020).

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863. doi: https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & Memory, 1,* 217–229. doi: 10.1101/lm.1.4.217

Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience, 10*, 248–263. doi: 10.1162/089892998562681

Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The Cognitive-Evolutionary Model of Surprise: A Review of the Evidence. *Topics in Cognitive Science, 11*, 50–74. doi: 10.1111/tops.12292

Rose, M., Haider, H., & Büchel, C. (2010). The emergence of explicit memory during learning. *Cerebral Cortex*, *20,* 2787–2797. doi: 10.1093/cercor/bhq025

Rosenthal, D. M. (1985). Two concepts of consciousness. *Philosophical studies, 49*, 329–359. doi:10.1007/BF00355521

Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. Flanagan, G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 729-753). Cambridge, MA: MIT Press.

Rosenthal, D. M. (2002). How many kinds of consciousness? *Consciousness & Cognition, 11*, 653–665. doi: 10.1016/s1053-8100(02)00017-x

Rosenthal, D. (2012). Higher-order awareness, misrepresentation and function. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367*(1594), 1424-1438. doi: 10.1098/rstb.2011.0353

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356-374. doi: 10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. doi: 10.3758/PBR.16.2.225

Ruttle, J. E., Hart, B., & Henriques, D. (2021). Implicit motor learning within three trials. *Scientific Reports, 11*, 1627. doi: 10.1038/s41598-021-81031-y

Rünger, D., & Frensch, P. A. (2008). How incidental sequence learning creates reportable knowledge: The role of unexpected events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1011-1026. doi: 10.1037/a0012942

Rünger, D. & Frensch, P. A. (2010). Defining consciousness in the context of incidental sequence learning: Theoretical considerations and empirical implications. *Psychological Research, 74*, 121–137. doi:10.1007/s00426-008-0225-8

Sævland, W. & Norman, E. (2016). Studying different tasks of implicit learning across multiple test sessions conducted on the web. *Frontiers in Psychology, 7*(808), 1–16. doi: 10.3389/fpsyg.2016.00808

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition, 19*, 1069-1078. doi: 10.1016/j.concog.2009.12.013

Schlaghecken, F., Sturmer, B., & Eimer, M. (2000). Chunking processes in the learning of event sequences: electrophysiological indicators. *Memory & Cognition, 28,* 821–831. doi: 10.3758/bf03198417

Schuck, N. W., Gaschler, R., Wenke, D., Heinzle, J. Frensch, P. A., Haynes, J.-D., & Reverberi, C. (2015). Medial Prefrontal Cortex Predicts Internally Driven Strategy Shifts. *Neuron, 86*, 331-340. doi: 10.1016/j.neuron.2015.03.015

Schwager, S., & Hagendorf, H. (2009). Goal-directed access to mental objects in working memory: The role of task-specific feature retrieval. *Memory & Cognition, 37,* 1103-1119. doi: 10.3758/MC.37.8.1103

Schwager, S., Rünger, D., Gaschler, R., & Frensch, P. A. (2012). Data-driven sequence learning or search: What are the prerequisites for the generation of explicit sequence knowledge? *Advances in Cognitive Psychology, 8*(2), 132-143. doi: 10.2478/v10053-008-0110-4

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

Scott, R., & Dienes, Z. (2010). The metacognitive role of familiarity in artificial grammar learning: Transitions from unconscious to conscious knowledge. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (p. 37–61). Springer Science + Business Media.

Seager, W. (2004). A cold look at HOT theory. In R. J. Gennaro (Ed.). *Higher-order theories of consciousness: An anthology* (pp. 255–275). John Benjamins Publishing Company. doi: 10.1075/aicr.56.14sea

Shanahan, M., & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition, 98,* 157-176. doi: 10.1016/j.cognition.2004.11.00

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1435-1451. doi:10.1037/0278-7393.25.6.1435

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367-447. doi:10.1017/S0140525X00035032

Shea, N., & Frith, C. D. (2019). The global workspace needs metacognition. *Trends in Cognitive Sciences, 23*, 560-571. doi:10.1016/j.tics.2019.04.007

Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition, 35*, 53-65. doi: 10.1016/j.concog.2015.04.015

Shin, J. C. (2008). The procedural learning of action order is independent of temporal learning. *Psychological Research, 72,* 376–386. doi: 10.1007/s00426-007-0115-5

Shin, Y., K., Proctor, R. W. & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin, 136*, 943–974. doi:10.1037/a0020541

SoSci Survey GmbH. https://www.soscisurvey.de (accessed on 21 July 2020).

Stahl, C., Barth, M., & Haider, H. (2015). Distorted estimates of implicit and explicit learning in applications of the process-dissociation procedure to the SRT task. *Consciousness and Cognition, 37,* 27–43. doi: 10.1016/j.concog.2015.08.003

Stadler, M. A. (1989). On learning complex procedural knowledge. *Journal of Experimental Psychology*, *11*, 653–665. doi: 10.1037/0278-7393.15.6.1061

Stöcker, C. & Hoffmann, J. G. (2004). The ideomotor principle and motor sequence acquisition: Tone effects facilitate movement chunking. *Psychological Research, 56A*, 1–19. doi:10.1007/s00426-003-0150-9

Stöcker, C., Sebald, A. & Hoffmann, J. G. (2003). The influence of response-effect compatibility in a serial reaction time task. *The Quarterly Journal of Experimental Psychology, 68*, 126–137. doi:10.1080/02724980244000585

Tamayo, R., & Frensch, P. A. (2015). Temporal stability of implicit sequence knowledge: Implications for single-system models of memory. *Experimental Psychology*, *62*, 240–253. doi: 10.1027/1618-3169/a000293

Tubau, E., & Lopéz-Moliner, J. (2004). Spatial interference and response control in sequence learning: the role of explicit knowledge. *Psychological research, 68*, 55–63. doi: 10.1007/s00426-003-0139-4

Tubau, E., Lopéz-Moliner, J. & Hommel, B. (2007). Modes of executive control in sequence learning: from stimulus-based to plan-based control. *Journal of Experimental Psychology: General*, *136,* 43–63. doi:1037/0096-3445.136.1.43

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564. doi: 10.1037/0096-3445.134.4.552

Wessel, J. R., Haider, H., & Rose, M. (2012). The transition from implicit to explicit representations in incidental learning situations: More evidence from high-frequency EEG coupling. *Experimental Brain Research*, *217*(1), 153-162. doi: 10.1007/s00221-011-2982-7

Whittlesea, B. A. (2002). False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General*, *131*, 96-115. doi:10.1037/0096-3445.131.1.96

Whittlesea, B. A. (2004). The Perception of Integrality: Remembering Through the Validation of Expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30,* 891-908. doi:10.1037/0278-7393.30.4.891

Whittlesea, B. W., & Williams, L. D. (2000). The source of feelings of familiarity: the discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26,* 547-565. doi: 10.1037//0278-7393.26.3.547

Wilbert, J. & Haider, H. (2012). The subjective experience of committed errors and the Discrepancy-Attribution hypothesis. *Acta Psychologica, 139*, 370-381. doi: 10.1016/j.actpsy.2011.11.010

Wilkinson, L., & Shanks, D. R. (2004). Intentional Control and Implicit Sequence Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30,* 354-369. doi:10.1037/0278-7393.30.2.354

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105*, 558–584. doi: 10.1037/0033-295x.105.3.558

Willingham, D.B., Greenberg, A. R., & Cannon, T. (1997). Response-to-stimulus interval does not affect implicit motor sequence learning, but does affect performance. *Memory & Cognition, 25*, 534–542. doi: 10.3758/bf03201128

Willingham, D.B., Nissen, M. J. & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology, 15*, 1047–1060. doi: 10.1037/0278-7393.15.6.1047

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3*, 1212–1217. doi: 10.1038/81497

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences, 367*, 1310-1321. doi: 10.1098/rstb.2011.0416

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S. & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin, 133*, 273–293. doi: 10.1037/0033-2909.133.2.273

Ziessler, M. (1998). Response–effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 962-978. doi:10.1037/0278-7393.24.4.962

Ziessler, M. & Nattkemper, D. (2001). Learning of event sequences is based on response-effect learning: Further evidence from a serial reaction task. *Journal of Experimental Psychology, 27*, 595–613. doi:10.1037W0278-7393.27.3.595

Ziessler, M. & Nattkemper, D. (2002). Effect anticipation in action planning. In W. Prinz & B. Hommel (Hrsg.), *Common Mechanisms in Perception and Action. Attention and Performance XIX* (S. 645 – 672). Oxford, England: Oxford University Press.

Ziessler, M., Nattkemper, D. & Frensch, P. A. (2004). The role of anticipation and intention in the learning of effects of self-performed actions. *Psychological Research, 68*, 163–175. doi:10.1007/s00426-003-0153-6

Zirngibl, C. & Koch, I. (2002). The impact of response mode on implicit and explicit sequence learning. *Experimental Psychology*, *49*, 153–162. doi:10.1027//1618-3169.49.2.153

Published articles and contributions of the authors

Review

Esser, S. Lustig, C. & Haider, H. (2021). What Triggers Explicit Awareness in Implicit Sequence Learning? Implications from Theories of Consciousness. *Psychological Research.* doi: 0.1007/s00426-021-01594-3

Sarah Esser wrote the manuscript. Clarissa Lustig helped writing the manuscript. Hilde Haider revised the manuscript.

Study 1

Lustig, C. & Haider, H. (2019). Response-effects trigger the development of explicit knowledge. *Acta Psychologica, 194*, 87–100. doi: 10.1016/j.actpsy.2019.01.016

Clarissa Lustig and Hilde Haider developed the hypotheses and designs for the Experiments 1 a, b and 2 a, b. Clarissa Lustig programmed the experiments, analyzed the data and wrote the manuscript. Hilde Haider revised the manuscript.

Study 2

Lustig, C., Esser, S. & Haider, H. (2021). The interplay between unexpected events and behavior in the development of explicit knowledge in implicit sequence learning. *Psychological Research*. doi: 10.1007/s00426-021-01630-2

Clarissa Lustig, Sarah Esser and Hilde Haider developed the hypotheses and designs for the experiment. Clarissa Lustig programmed the experiments, analyzed the data and wrote the manuscript. Hilde Haider and Sarah Esser revised the manuscript.