The Role of Unexpected Events in the Emergence of Explicit Knowledge in an Implicit Learning Situation

Inaugural-Dissertation

zur
Erlangung des Doktorgrades
der Humanwissenschaftlichen Fakultät
der Universität zu Köln
vorgelegt von

Sarah Esser
aus Köln

Köln 2017

1. Berichterstatter: Prof. Dr. Hilde Haider (Köln)

2. Berichterstatter: PD Dr. Michael Rose (Hamburg)

Tag der mündlichen Prüfung: 12.10.2017

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Oktober 2017 angenommen.

Zusammenfassung

In einer impliziten Lernaufgabe, wie der seriellen Wahlreaktionszeit-Aufgabe, erwerben die meisten Personen unbewusst Wissen über die zugrundeliegende Regelhaftigkeit. In der Regel gibt es aber auch immer eine kleine Gruppe Personen, welcher diese Regelhaftigkeit auffällt und diese auch berichten kann. Da das Bewusstsein über eine erworbene Repräsentation entscheidend dafür ist, wie flexibel und vielfältig dieses Wissen eingesetzt werden kann, ist es von großem Interesse, Verständnis darüber zu erlangen, welche Mechanismen den Übergang von unbewusstem zu bewusstem Wissen realisieren.

In bisherigen Forschungsarbeiten haben sich zwei zentrale Strömungen ausgebildet, welche sich dieser Frage widmen. Zum einen besteht die sparsamste Annahme darin, dass unbewusste Repräsentationen durch Übung zunehmend an Qualität gewinnen und so graduell in einen bewussten Zustand übergehen (Single-System Annahme; z.B. Cleeremans & Jiménez, 2002). Dem gegenüber stehen komplexere Modelle, welche annehmen dass implizite und explizite Repräsentationen durch separate Lern- und Gedächtnissysteme gestützt werden (Multiple-System Annahme). Eines dieser Modelle ist die Unexpected Event Hypothese (Frensch et al., 2003). Diese besagt, dass implizites Lernen zu Verhaltensänderungen führt, welche in Widerspruch zu den Erwartungen der Person über ihr eigenes Verhalten in der jeweiligen Situation stehen. Diese Erwartungsverletzung löst einen Attributionsprozess aus, anhand welchem Erwartung und Erleben wieder in Kohärenz gebracht werden sollen; das Resultat kann die plötzliche Einsicht in die zugrundeliegende Regel sein.

Die vorliegenden drei Studien haben zum Ziel die Vorhersagen der Unexpected Event Hypothese zu testen und diese den Vorhersagen einer sparsameren Single-System Annahme gegenüber zu stellen. In allen drei Studien werden daher in einer impliziten Lernsituation über verschiedene Manipulationen unerwartete Ereignisse induziert. Gleichzeitig sind alle Aufgaben so entwickelt, dass die assoziative Stärke der Repräsentationen zwischen den Manipulationen nicht variieren soll.

In Studie 1 wird in drei Experimenten das subjektive Gefühl der Flüssigkeit anhand der Anordnung regelhafter und zufälliger Durchgänge als unerwartetes Ereignis manipuliert. Experiment 1 zeigt, dass die Anordnung der verschiedenen Durchgangstypen keinen Einfluss auf die assoziative Stärke der erworbenen Repräsentationen zu haben scheint. Experiment 2 zeigt, dass diese Anordnung tatsächlich das subjektive Flüssigkeitsempfinden beeinflusst. Experiment 3 zeigt abschließend, dass TeilnehmerInnen, welche größere Unterschiede im subjektiven Flüssigkeitsempfinden wahrnehmen, auch mehr explizites Wissen erwerben.

In Studie 2 und 3 werden kontingente Handlungseffekte als unerwartete Ereignisse eingesetzt. Dabei wird in Studie 2 die Handlungs-Effekt Kontingenz durch das bei den

TeilnehmerInnen induzierte Task-Set manipuliert. Im Verlauf von zwei Experimenten wird zunächst geprüft, ob die Manipulation des Task-Sets tatsächlich die Entstehung expliziten Sequenzwissens beeinflusst. Anschließend wird in einem dritten Experiment geprüft, inwiefern dieser Effekt spezifisch für explizite, aber nicht implizite Lernprozesse ist. In Studie 3 hingegen wird eine direktere Manipulation der Handlungs-Effekt Kontingenz verwendet. Hierbei sind die Effekte entweder an die Handlungen der Teilnehmer oder an eine aufgaben-irrelevante Stimulusdimension gebunden. Die Studien 2 und 3 zeigen vermehrt explizites Sequenzwissen, wenn die TeilnehmerInnen kontingente Handlungseffekte erleben. Diese Befunde werden ebenfalls im Rahmen der Unexpected Event Hypothese interpretiert.

Insgesamt zeigen alle drei Studien mehr explizites Wissen bei Personen, welche unerwartete Ereignisse erlebten. Da alle Studien darauf ausgelegt sind, die assoziative Stärke zwischen den Bedingungen gleich zu halten und nur die Wahrscheinlichkeit eines unerwarteten Ereignisses zu manipulieren, scheinen die Ergebnisse dafür zu sprechen, dass die Unexpected Event Hypothese einer einfachen Single-System Annahme vorzuziehen ist.

Abstract

In an implicit learning task like the serial reaction time task, most people demonstrate implicit knowledge about the underlying regularity. Usually, a small group of persons can be found which notices this regularity and is also able to report it. Because whether the acquired representation can be used in a flexible and diverse way crucially depends on conscious awareness of this knowledge, it is of great importance to understand which mechanisms realize the transition from implicit to explicit knowledge.

Research on this issue has led to two main theoretical streams. On the one hand, the most parsimonious account assumes that unconscious representations gain quality through practice and therefore gradually transform into explicit knowledge (single-system account; e.g. Cleeremans & Jiménez, 2002). On the other hand, there are more complex models which assume that implicit and explicit representations are supported by separable learning- and memory systems (multiple-systems account). One of these models is the Unexpected Event Hypothesis (Frensch et al., 2003). Within this model, it is proposed that implicit learning leads to behavioral changes which contradict the expectations of a person about their own behavior in the given situation. This violation of expectations triggers an attributional process which should bring expectation and experience back into coherence; a sudden insight into the underlying rule can be the result.

The three studies presented here are aimed at testing the predictions of the Unexpected Event Hypothesis and contrast these with the more parsimonious predictions of a single system account. Therefore, in all three studies, different manipulations will induce unexpected events in an implicit learning situation. At the same time, all tasks are designed in a way to match the associative strength of the representations between the manipulations.

In Study 1 in three experiments, the subjective feeling of fluency is manipulated through the arrangement of regular and random trials in order to establish an unexpected event. Experiment 1 shows that the arrangement of the different trial-types does not affect the associative strength of the acquired representational. Experiment 2 shows that the arrangement of the trial-types affects the subjective experience of fluency. Lastly, Experiment 3 demonstrates that participants who experienced greater differences in their subjective feeling of fluency exhibit more explicit knowledge.

In Study 2 and 3, contingent action-effects establish the unexpected events. In Study 2 the action-effect contingency is manipulated by the induced task-set of the participants. First, over the course of two experiments, it is tested whether the manipulation of the task-set affects the emergence of explicit sequence knowledge. Subsequently, in a third experiment, it is tested whether this effect is specific for explicit but not implicit learning processes. In Study 3, on the contrary a more direct manipulation of the action-effect contingency is used. Here, the effects are either bound to the responses of the participants or to a task-irrelevant stimulus dimension. Studies 2 and 3 show

enhanced explicit sequence knowledge when participants experienced contingent action-effects. These results are interpreted in favor of the Unexpected Event Hypothesis.

Together, all three studies demonstrate more explicit knowledge in participants who experienced unexpected events. Because all studies were aimed at keeping the associative strength equal across the conditions equal and only manipulate the likelihood of unexpected events, the results seem in favor of the Unexpected Event Hypothesis over the simpler single system account.

Content

1 Introduction

Our world is full of structured information. Sometimes this structure is rather simple and fully deterministic like the few keys on a keyboard one has to remember to type at an acceptable speed. Sometimes it is highly complex and probabilistic, like social interactions where small modifications of facial expressions or the choice of words can profoundly change the impression we leave on other people. Luckily, we are able to pick up this information and learn to predict our environment to a satisfactory certainty. This ability alleviates our day-to-day lives invaluably or, to a certain extent, enables them in the first place. In this respect it is usually unnecessary and, given the serial nature of conscious processing, often also impossible to be consciously aware of all the hidden rules that guide our behavior. Yet it remains mysterious why sometimes we become aware of the rules that structure information. Of course, sometimes we become aware of them simply because we are asked to do so, for example when someone asks us for the directions to a certain location. At other times however, we experience a sudden insight into knowledge that was entirely unconscious before. For example when we realize that a friend we sometimes play poker with becomes very quiet when they have a good hand.

What are the necessary and sufficient conditions for such insights to occur? Answering this question is not only interesting per se as it could provide important practical implications, for example in an educational context, but also to much more fundamental interests. Understanding the transition from unconscious to conscious knowledge can help with the further development of scientific theories of consciousness. How and why is neural information processing sometimes accompanied by conscious knowledge of its contents and in which qualities does this processing differ from one that is not accompanied by consciousness? How are representational quality or strength and consciousness related? What is the significance of interaction with and feedback from the external world?

From a scientific view, implicit learning research can provide very interesting insights into the necessary conditions for a transition from unconscious to conscious knowledge. Here, the typical paradigm for implicit learning, the Serial Reaction Time Task (SRTT; Nissen & Bullemer, 1987) can reflect everyday situations with deterministic and probabilistic contingencies. These contingencies can be of arbitrary complexity and can occur within various stimulus- and response dimensions (e.g. motor, visual, auditive, spatial, and temporal). It has been shown repeatedly that learning in the SRTT can result in implicit, unconscious knowledge about the hidden sequential contingencies (see Abrahamse, Jiménez, Verwey, & Clegg, 2010, for a review). The paradigm provides a vast horizon of possible manipulations, not only concerning the content of what can be learned implicitly but also concerning the factors that can influence the rate of conscious insights into the hidden structures of the task. The following studies are aimed at testing two accounts about the generation of conscious, explicit knowledge in an implicit learning situation. These two accounts are usually treated separately despite surely being compatible. The first account concerns the role of metacognitive judgements and is treated in study 1. Moreover, in this study the importance of metacognitive judgements relying on observance of one's own behavior is contrasted with a simpler single system account according to which representational strength is seen as the variable on which conscious knowledge depends. The second account is less directly related to the transition from implicit to explicit knowledge and rather finds its roots in the investigation of intentional action control as opposed to stimulus-based action control. The topic of discussion here is action-effect learning. The studies 2 and 3 both investigate the assumption that experiencing contingent action-effects is an important source of external feedback that promotes a shift from unconscious, stimulus-based to conscious plan-based control.

Before these studies are presented, the following chapters provide a short insight into the scientific theories of unconscious and conscious processing, which build the crucial basis for any further thoughts on how these two forms of processing are related (Chapter 2). Building on this, a more specified summary will be given on how the transition from unconscious to conscious processing is treated in the field of implicit learning. In that context two important theoretical viewpoints, single- and multiple-system views, will be summarized and compared (Chapter 3). Finally, an overview of the empirical evidence which has accumulated for both opposing viewpoints so far will shortly be discussed with regard to open points that are subject of the present studies (Chapter 4).

2 Scientific Theories of Consciousness

In order to investigate the transition from unconscious to conscious knowledge, it is of indispensable importance to think about the highly controversial possibilities to conceive a scientifically seizable theory of consciousness. There are two positions which, at the current point of time, share an equally strong degree of influence in the literature on consciousness. In their most reduced form, both positions can be separated by either postulating or disputing a differentiation between “easy” and “hard” problems of consciousness research. These terms go back to Chalmers 1995 who said that it is “easy” for science to identify the (neuronal) mechanisms that support cognitive functions and which have a strong relatedness to conscious processing. These include language, attention, executive functions or generally any process that is traceable with objective measures. Nevertheless, even after all cognitive functions have been explained, consciousness, according to this view, has an inherently subjective, phenomenal component, also referred to as *qualia*, that will not be explained by understanding information processing. Due to its impenetrable first-person nature, it would remain unanswered how and why these cognitive mechanisms are accompanied by a subjective feeling of being conscious; this is what is called the “hard” problem. A similar idea stands behind the separation between so-called “access”- and “phenomenal” consciousness (Block, 1995). Access consciousness refers to any information that is made available to a broad network of cognitive functions. This includes verbalization, categorization, reasoning, planning or more generally anything that subsumes under cognitive control. Phenomenal consciousness, as Block (2007, p. 487) put it, “overflows access”. Phenomenal consciousness has the property of containing richer information than what we can report or voluntarily act upon, again touching the subjective first-person characteristic of consciousness. Several neuroscientific theories that agree with this dissociative view have been put up, trying to explain phenomenal consciousness. These, for example, pronounce the relevance of local recurrency of information (Block, 2005, 2007; Lamme, 2006), assume that a holistic macro-consciousness is comprised of distributed nodes that each a create micro-consciousness (e.g. color, sound, etc.; Zeki, 2003) or take an evolutionary route by postulating that neurons form temporary coalitions to compete for access to attentional systems (Crick & Koch, 1990).

Consciousness researchers of the other camp heavily disagree with a divide between access- and phenomenal consciousness. The most prominent representative of these views, Daniel Dennett, claims that the assumption of a special property as qualia or phenomenal consciousness is “scientifically insupportable and deeply misleading” (Dennett, 2015, p. 2). The reason for this is that there never can be an experimental setting that is able to study consciousness in the absence of access and function. According to proponents of the existence of qualia, in a “perfect (thought) experiment”, isolating the perception of the color “red” of an apple from all other cognitive functions should lead to a phenomenal but inaccessible consciousness about the apples redness. Nevertheless, the scientist would find a person who verbally insures that they do not see red, that they do not feel anything they associate with this color and who is not able to act upon this perception in any possible way, because the center for color cannot communicate with any other functional area (Cohen & Dennett, 2011). Phenomenal consciousness, according to this view, cannot be falsified and would have to rely on arbitrarily chosen criteria independent of subjective and functional observations and is therefore by definition a question of believing but not of science. Functional theories of consciousness on the contrary assume that there is no brain state independent of function. Consciousness is fully explained once it has been understood how cognitive processes interact. The interaction of this multitude of processes, most importantly attention, working memory, language and decision making processes, *is* consciousness. There is no additional process that produces consciousness. The task for a science of consciousness is to specify, in a falsifiable way, which functions are necessary for consciousness and how these functions can be measured adequately.

Understanding these differences between functional approaches to consciousness and those that assume that there is an extra process creating phenomenal consciousness is vital for the further understanding of the definition and the measures of unconscious and conscious knowledge behind our hypotheses. In the following, two important functional theories behind our hypotheses about the transition from implicit to explicit knowledge will be introduced shortly.

2.1 The Global Workspace Theory

The *Global Workspace Theory* (GWT; Baars, 1997, 2005; Dehaene & Changeux, 2011; Dehaene & Naccache, 2001; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006) was one of the first scientific functional theories of consciousness that seems capable of abolishing the *homunculus problem* many other approaches have, especially those assuming a separable phenomenal consciousness.

The homunculus problem encompasses two difficulties: The first is the assumption that there are certain networks or certain neurons which *create* consciousness, and which constitute a place where consciousness happens. This place is termed the *Cartesian Theatre* by Dennett (1991). The implication here is that there is an entity (the homunculus) additional to brain states, for whom or which information needs to be presented in order to exert control, and which leads to the subjective experience of a *stream of consciousness*. Introducing such an entity necessarily leads back to the hard problem, asking why and how neuronal information becomes transformed into conscious information in the first place. The second problem with any theory implicitly comprising a homunculus is that it also has to explain how the information which gets presented is selected from the enormous pool of unconscious information and how consciousness as a final product can have a causal influence on unconscious brain processes.

The GWT aims at solving both of these problems. Most importantly, the GWT does not assume that there is any specified area in the brain where consciousness is created; rather, consciousness is equalized with global availability of information. The basic assumption of the GWT is that the brain contains a multitude of functionally highly specialized areas working in parallel. Information in these areas is unconscious, there is no micro-consciousness 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 specialization 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 information and puts it into the focus of attention. Here, the theory postulates that there is a global workspace (GWS) which provides the necessary infrastructure, neurologically mainly realized by thalamo-cortical long-distance neurons of the prefrontal and the anterior cingular cortex (see Baars, Franklin, & Ramsøy, 2013 for a detailed elaboration of the neuronal architecture). The GWS is able to select relevant information, prevents interference, allows the encapsulated modules to exchange information and flexibly establishes temporary networks between these modules (Dehaene & Naccache, 2001). The GWT uses a blackboard as a metaphor for imagining how the GWS works. When a module gets selected to enter the GWS, it can broadcast its content to any other network in the brain. Other modules can use this information from the blackboard and process it in their specified function. The information of 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, but instead is now processed in a broad context of unconscious subsystems. These subsystems include, for example, perception, language, intentions, self-concepts, expectations, memory, and also exclusive access to working-memory function (Baars, 1997, 2005; Baars et al. 2013; see Figure 1). 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).

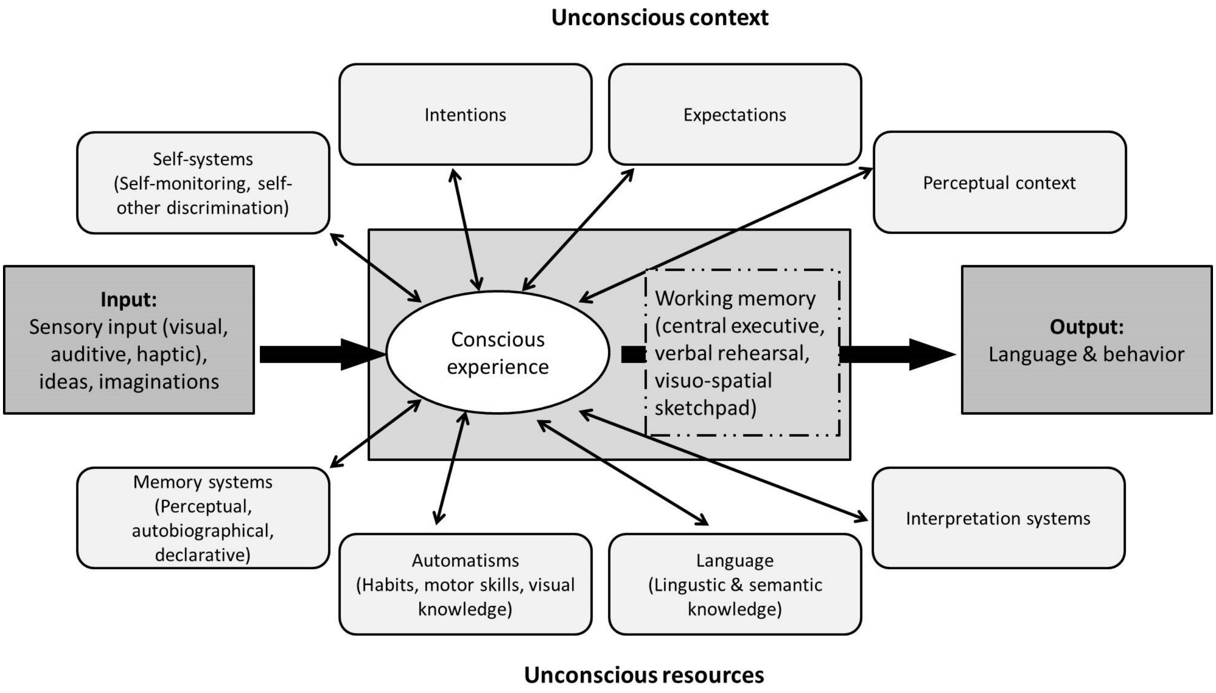
Unconscious context
Intentions
Expectations
Self-systems (Self-monitoring, self-other discrimination)
Perceptual context
Input: Sensory input (visual, auditive, haptic), ideas, imaginations
Conscious experience
Working memory (central executive, verbal rehearsal, visuo-spatial sketchpad)
Output: Language & behavior
Memory systems (Perceptual, autobiographical, declarative)
Interpretation systems
Automatisms (Habits, motor skills, visual knowledge)
Language (Lingustic & semantic knowledge)
Unconscious resources

Figure 1. Schematic global workspace model (Baars et al., 2013). Conscious experience corresponds to the information currently being represented in working memory and focused by attention. Conscious contents have access to a broad range of unconscious modules (self-systems, intentions, etc.), which set the context for the conscious information, and define the quality of subjective experience. Furthermore, the information currently in the global workspace is accessible to unconscious resources (memory systems, language, etc.) which can process this information according to their specialized function. All unconscious modules constantly send feedback to the global workspace and can enter the global workspace themselves, if their activational strength signals high relevance to the current goals. It is the function of the global workspace to converge the parallel input of the various specialized modules and allow serial behavioral output.

This conception of consciousness resembles the perfect thought experiment of Cohen & Dennett (2011) described in Chapter 2. There is neither a certain mechanism that creates consciousness, nor is there a certain place in the brain where consciousness happens. Consciousness is the global accessibility of information. The apple is consciously perceived as being red because the encapsulated, unconscious color module is allowed to communicate with all the just named subsystems. Further, a broad variety of inner thoughts and emotions, as well as options to openly act upon this perception, becomes enabled. Even though the GWS is said to rely on long-distance neurons of the PFC and the ACC, the GWS is not located in any certain area. It is a virtual space, dynamically changing with the contents being processed and the functions being involved.

As mentioned above, the GWT also aims at solving the question of how the information which is most relevant for current action planning gets selected to be distributed via the GWS. The problem of theories that comprise a homunculus assume a top-down control process where some supervising entity decides which information is the most important at any point of time. Which modules should be considered as being relevant and what would the consequences of this selection be? Obviously, a mechanism that tests all potentially relevant unconscious modules would be stuck in a combinatorial explosion, never coming to any successful decision (see the Frame-Problem, Dennett, 1984). Instead, the GWT suggests a bottom-up stochastic variation-selection mechanism (neural Darwinism, Changeux & Dehaene, 1989). Every unconscious module constantly competes for access to the GWS (variation component), while the GWS 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 depended content of the GWS and will therefore win the competition for global broadcasting (Shanahan & Baars, 2005). For example, driving in a car at night can lead to the consciously represented plan to watch out for any moving objects near the road. Any information stemming from a network that processes movement information will receive extra activational strength. In this case any moving object will gain access over a signal with higher bottom-up strength, for example a talking passenger. Yet, very strong bottom-up signal strength might still win over a signal that fits the current fitness function and “break through into consciousness” (Baars, 2005, p. 49), for example when the passenger starts yelling at you.

Taken together, the GWT provides a scientifically accessible conception of consciousness because it allows testable predictions about the observable functions that should separate conscious from unconscious processing. Therefore it also qualifies as an important background to think about the conditions for a transition from implicit to explicit (sequence) knowledge. Before considering the plausibility of different theories that are concerned with this more specific question in Chapter 3, one further important theory of consciousness should be presented here, namely the *Higher-Order Thought Theory*.

2.2 Higher-Order Thought Theories

The Higher-Order Thought Theory (HOTT) in its most popular form goes back to the work of Rosenthal (1997). Different from the neurologically oriented GWT, the HOTT originally was much more related to the philosophical side of consciousness theories. Nevertheless, it has developed to be a theory with empirically testable predictions which also is compatible with the current neurological and cognitive state of knowledge. The HOTT is concerned with the metacognitive aspects of consciousness. 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 GWT. Encapsulated information processing can be seen as a first-order state which per se 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 the HOTT, crucially depends on developing higher-order knowledge about this first-order knowledge. 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"). 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. Lau (2008a) suggests a Bayesian learning approach to describe the relation between perceptual first-order states and meta-cognitive higher-order judgements about these states. This approach builds on psychophysical signal-detection theory. Every performance depends on the sensitivity (*d'*) of a processing unit and its decision-criterion (*c*). Many studies on (un-) conscious perception use *d'* as a measure for awareness (*d'* = 0 is interpreted as unconscious perception). Lau (2008a) argues that operationalizing unconscious perception as *d'* = 0 severely underestimates the potential of unconscious processing as it only reflects poor perceptual performance capacity. Instead, he suggests that the decision criterion *c* is of much greater relevance for studies of consciousness. In a related study (Lau & Passingham, 2006), it has been demonstrated that manipulating attention towards a masked stimulus can result in equal sensitivity towards the stimulus but very different subjective reports of having seen or having not seen the stimulus. Likewise, the widespread cortical activity which the GWT relates to conscious processing is, according to Lau and Passingham, a confound due to differences in performance. When performance was matched, subjective reports of consciousness were only related to increased activity in the dorsolateral cortex.

What is important for the HOTT is how the cognitive system comes to a decision about its own internal states and their reference to external stimuli. Bayesian decision theory can help to understand how an optimal criterion is set. Via external feedback, the cognitive system can develop a first-order representation of the probability of a certain signal strength when a signal is present and when it is not. Additionally, the system also has to learn the base rate of the stimulus. Given these two probabilities, the cognitive system can estimate the opposite likeliness that a stimulus is present, given a certain signal strength, which is important for making optimal decisions. However, people often seem to fail to set an optimal criterion. A common and easily accessible example are blindsight patients, who, according to Lau (2008a), set very conservative criteria even though their performance is well above chance. Likewise, and more related to the content of the research presented in the following, experimental studies on unconscious learning and perception show the same above-chance performance while the participants claim that they are merely guessing (Cheesman & Merikle, 1984; Dienes & Berry, 1997). Lau (2008a) therefore argues that metacognitive higher-order representations need to be assumed in order to explain how the criterion for a subjective judgement about the accuracy of one`s own responses or, more general, perceptual consciousness is set. A higher-order mechanism is assumed which develops a representation of its own internal first-order signals which are related to the external world. For example, a person might learn via feedback that the mean fire rate for a signal being present is 25 Hz and a signal being absent is 10 Hz and sets an optimal criterion at 17.5 Hz. Now, due to a lesion, the internal signal decreases strongly and the first-order distributions now show a mean of 15 Hz for a signal and 5 Hz for noise. Failing to learn about these new internal distributions would result in keeping the criterion at 17.5 Hz which now leads to a very high amount of false-negatives, as can be observed in blindsight patients. Recently, Fleming and Daw (2017) gave a more detailed, computational description of the learning processes behind the development of metacognitive criterion setting. Thereby the authors also go into deeper detail between first-order criteria that reflect how a person might react to a stimulus (e.g. pressing or not pressing a button) and higher-order judgements of having consciously perceived something.

Taken together, HOTTs also provide a theoretical base for functionally discriminating unconscious from conscious processing. While the GWT is more focused on the behavioral enrichment that comes along with conscious processing on various levels (controllability, flexibility, verbalization, integration, combination, etc.), HOTTs rather pronounce the subjective, phenomenological side of conscious processing (Rosenthal, 2008). Without (unfalsifiably) assuming that there is a phenomenology to first-order processing, HOTTs propose that consciousness is given when a person can represent being (or not being) in a certain state.

3 The Transition from Implicit to Explicit Knowledge

Implicit learning became a popular research topic in cognitive psychology in the 1960s when George Miller (1958) established the paradigm of *Artificial Grammar Learning* (AGL). He demonstrated that participants learned to become sensitive to letter strings that followed hidden rule, even though they were not explicitly informed about these rules. Arthur Reber (1967) later explored the implicit aspect, namely the inability of the participants to verbally express any of their knowledge about the hidden grammar, more deeply. From then on, the interest in the ability to learn without being conscious of what has been learned or even that anything has been learned at all, spread to areas outside of linguistic research. This included paradigms on probability learning (A. Reber & Millward, 1965), dynamic system control (Berry & Broadbent, 1984) and, most important to the studies at hand, sequence learning within the *Serial Reaction Time Task* (SRTT), going back to Nissen & Bullemer (1987). Naturally, there has been an extensive, productive and still not univocally settled debate on whether the knowledge acquired in these tasks can be said to be unconscious. Over the many years of research on unconscious processing, convincing empirical evidence accumulated to justify the assumption of unconscious influences on behavior which functionally differ from consciously perceived input (Alamia et al., 2016; Kouider & Dehaene, 2007). Still there are scientists who hold the strong opinion that it has yet to be shown that there is such a thing as unconscious perception or decision making (Newell & Shanks, 2014; Peters & Lau, 2015; Shanks, 2016; Tran & Pashler, 2017). Surely, productive criticism on the methods being used to distinguish unconscious from conscious processing should not be dismissed easily. However, in the light of the above shortly introduced functional theories of consciousness, the assumption that there is no unconscious processing to be found, entails some theoretical problem which will shortly be discussed in Chapter 3.1.

Presuming that there are unconscious processes that influence behavior, implicit learning paradigms, especially the SRTT, provide immensely powerful tools not only for demonstrating the ubiquity and flexibility of implicit learning processes but moreover for studying the requirements for a transition from unconscious knowledge to conscious insight into this knowledge. Priming research, the probably most prominent paradigm for unconscious processing, has the difficulty that stimulus signal strength and task performance capacity are often not matched between the conditions (Lau, 2008a; 2008b). This means that comparing a dim, brief, masked or an “unseen” stimulus with a clearly visible or “seen” stimulus should result in different performances without the need to assume conscious and unconscious processing.

Implicit learning paradigms as the SRTT, by contrast, allow training people with a certain sequence over a long period of time leading to a strong implicit knowledge base. Thus, in sequence learning paradigms it is possible to match the associative strength, and therefore the signal strength, between conditions as well as the difficulty and therefore the task performance capacity. At the same time factors that are suspected to influence whether participants gain insight into the hidden sequence can be manipulated. Of course, the underlying theory about the nature of consciousness influences which measures can differentiate between unconscious and conscious processing and also strongly influence which manipulations are expected to influence the transition of unconscious to conscious sequence knowledge (see Chapter 4.1).

Moreover, the differences in the theoretical conception of conscious and unconscious processing also are reflected in implicit learning research. There is a long and not yet solved debate about the nature and the relation of implicit and explicit knowledge. The manifoldness of the different theoretical perspectives can best be simplified by dividing them into *single-* and *multiple-system accounts*. The following sections of this chapter present the most influential theories within both of these views. It is described how these accounts characterize implicit and explicit learning processes and how the transition from implicit to explicit knowledge is imagined. Furthermore, a short evaluation of the theoretical views on consciousness behind them is provided.

3.1 Single-System Views

The core assumption of single-system views within implicit learning research is that there is no need to assume any additional process which transforms unconscious into conscious knowledge. Yet, there are two classes of single-system views which differ strongly in their underlying assumption about the nature of consciousness. This difference lies in their stance on the verifiability of unconscious processing. Even though unconscious processing is supposed to be empirically demonstrated in a vast and long history of scientific research (see Kouider & Dehaene, 2007, for an overview), there still are scientists who reject that such successful demonstrations exist (Newell & Shanks, 2014; Perruchet & Vinter, 2002; Peters & Lau, 2015; Shanks & St John, 1994).

Some of these scientists state that unconscious processing has simply not been demonstrated convincingly even though it probably exists (Peters & Lau, 2015). Others (Perruchet & Vinter, 2002) instead put forward the strong hypothesis that there is an isomorphism between phenomenal experience and acquired knowledge about the world, hence that all mental representations are conscious. This latter assumption is incompatible with the functional views on consciousness, outlined in the former chapter. The idea of all representations being conscious cannot be falsified and is, depending on the understanding of the term "representation", also tautological. The authors define representations as "mental events" which are "involved in reasoning, inference, action planning, and other mental activities" (p. 299). They set an arbitrary cut-off where some neural activity, for example the change in connective weights, does not count as a representation yet. Therefore, according to this view, if a participant is trained in a SRTT and internally builds new associative weights, these weights are either below an arbitrary threshold and do not count as a representation or they exceed this threshold and can be tested via a direct knowledge test. So for these authors, a simple recognition task or a free generation task can demonstrate conscious sequence knowledge (Perruchet & Amorim, 1992). However, a large body of work has shown that above-chance performance in these tasks can differ profoundly from the persons' subjective experience of knowing (Cheesman & Merikle, 1984; Dienes & Berry, 1997) as well as it can differ from objective tests which require control of this knowledge (e.g. responding with the button which is the most *unlikely* to be next, Debner & Jacboy, 1994). Because of the theoretical pitfalls of this conception of conscious knowledge, this view is of no further importance to the studies at hand.

The methodological viewpoint that there has been no convincing demonstration of unconscious knowledge might be of somewhat more relevance to the following experiments. It surely is beyond the scope of this work to extensively debate on whether all studies have been adequately taken into account by the authors doubting that there has been a convincing demonstration of unconscious knowledge (Newell & Shanks, 2014; Peters & Lau, 2015). Nevertheless, the argument of the critiques is also interesting for the experiments presented here. As outlined in Chapter 2.2 HOTTs see the subjective criterion of perceiving or not perceiving something consciously as the most important or even the only relevant indicator of consciousness (Lau & Passingham, 2006; Lau & Rosenthal, 2011). Contrary to the GWT they do not necessarily assume that conscious and unconscious processing will differ in objective performance capacity. Whether this argument can hold for all types of elaborate, complex behavior is a different debate. Nevertheless, Lau's (2008a) point that operationalizing unconscious perception via d'= 0 and conscious perception as d'> 0 is defective, is important. Comparing a dim, brief, masked or an "unseen" stimulus with a clearly visible or "seen" stimulus should result in different performances, independent of whether this processing was unconscious or conscious. Lau (2008b) therefore calls for experiments where stimulus signal strength and task performance capacity are matched between the conditions but still show that a difference in subjective judgements is possible. This, according to the critiques, could show the need for a two-system view, separating between conscious and unconscious processing, much more convincingly. Peter's & Lau's (2015) argument is based on priming research, where indeed it has been very common practice to compare "unseen" with "seen" stimuli. In such a setting, it is a difficult task to match performances with a *d'* > 0 while at the same time ensuring that the perception of the stimuli can still potentially be unconscious.

Implicit learning paradigms like the SRTT instead, allow training people with a certain sequence over a long period of time and accordingly build a strong implicit knowledge base. Various studies in this field show evidence that participants can discriminate between sequential and non-sequential trials very well (*d'* > 0) while at the same time showing no subjective insight into this knowledge (*zero-correlation criterion*, Dienes & Berry, 1997; Ziori & Dienes, 2005). Within the SRTT paradigm, the single-system claim that all knowledge shown in different objective tests leads back to conscious knowledge of different representational strength can be falsified. Here, it is easier to match the associative strength and therefore the signal strength as well as the difficulty and therefore the task performance capacity. At the same time, factors that are suspected to influence whether participants gain insight into the hidden sequence can be manipulated.

Contrary to the above described viewpoint that all behavior is based on conscious representations, there are single-system views which might be more important to the issue of the following experiments. As mentioned initially, the crucial difference is that these further single-system views assume that both unconscious and conscious processing influences our behavior. Nevertheless, these theories are also single-system views because they assume that there is a gradual difference in the representational strength of unconscious and conscious processing without the need to assume any transformational process or a different representational format. Bottom-up signal strength is viewed as a sufficient condition for the criterion that divides unconscious from conscious processing. This view fits well with the theories described in Chapter 2 which assume that there is a difference between phenomenal and access consciousness (e.g. Block, 1995; Chalmers, 1995). There are many researchers who claim that different gradual qualities of conscious perception lead back to different qualities of signal strength (Block, 2007; Lamme, 2003, 2010; Overgaard, Rote, Mouridsen, & Ramsøy, 2006; Windey, Vermeiren, Atas, & Cleeremans, 2004; Zeki, 2003,). Most commonly, these assumptions can be found in research on visual priming (Atas, Vermeiren, & Cleeremans, 2013; Nieuwenhius & de Kleijn, 2011; Windey, Gevers, & Cleeremans, 2013).

In the field of implicit sequence learning, the most influential single-system view stems from Cleeremans and Jiménez (2002). Different to the mentalistic view of Perruchet & Vinter (2002) they take a computational stance and assume that there is no subthreshold processing that does not count as a representation and is not assumed to be relevant for behavior. Instead, any kind of neural processing is constantly causally effective because any processing is embedded in a causal chain. The cognitive system, in their view, is best characterized as hierarchically organized, interconnected modules in which information processing leads to dynamic, transient patterns of activation. Modulated by the strength of their connections and activations, these modules influence each other's processing. Their definition of a representation therefore does not refer to content in a sense of a "meaningful component of the represented world" (Perruchet & Vinter, p. 299). Rather, representations are graded, transient patterns of neural activation which can vary on different dimensions, influencing the "quality of [a] representation" (Cleeremans & Jiménez, 2002, p. 18). Most important to their theory, the quality of a representation is influenced by the following three factors: (1*) Stability*, i.e. the time a certain activational pattern can be maintained, (2) *strength*, i.e.

the number of modules involved and their respective activational strengths, and (3) *distinctiveness*, i.e. the extent of overlap between representations within a functional network (see Kinsbourne, 1996, for a similar position).

Cleeremans and Jiménez (2002) define learning as an adaptive process that necessarily accompanies information processing, resulting in changes of connection weights. These changes are termed indirect effects of learning because they do not necessarily lead to changes in subjective experience. Implicit learning first leads to very weak, poor quality representations. These, though already causally effective, might not result in noticeable subjective changes in one's experience. But while these representations gain quality with further learning, their potential influence on action, their availability to control processes and the subjective experience also gradually changes (Figure 2). First there might be slight changes in the subjective experience, leading, for example, to a feeling of the task becoming easier and finally, as the representations gain quality, lead to full conscious knowledge of the acquired knowledge. The function of consciousness, within this framework, is the ability to control this knowledge. Yet, the authors are not fully explicit on how independent the ability to control knowledge, the potency of a representation and subjective experience are. They only state all three aspects are "closely related" (p. 22). With reference to Block (1995) it is remarked that there is a differentiation between access- and phenomenal consciousness within their framework. However, it remains open how and why representational quality affects these different aspects of consciousness to a different extent and, even more important, how these aspects could potentially be captured within their connectionist approach.

Furthermore, in their essay, the authors state that quality of representation is crucial for differentiating unconscious from conscious knowledge, but also note that these are necessary but not sufficient conditions for representations to become conscious. It is acknowledged, with a reference to the global workspace theory, that attention and integration of information play an important role in determining whether any sufficiently strong representation will eventually enter a state of conscious processing. This supposed involvement of top-down mechanisms is not elaborated any further. It is therefore not ultimately clarified whether the quality of the representation is changed when accessed by top-down mechanisms or how these mechanisms are triggered in the first place. Yet, even with this open question about the role of top-down processes and the sufficient conditions for the emergence of a conscious representation, the authors seemed to imply a single-system view behind the graded characteristics of conscious processing. The consciously processed information is the formerly implicit information that can now be accessed, and which gains further strength and stability by this accessibility. There is no need for a second system supporting a different representational format which would be required for consciousness.

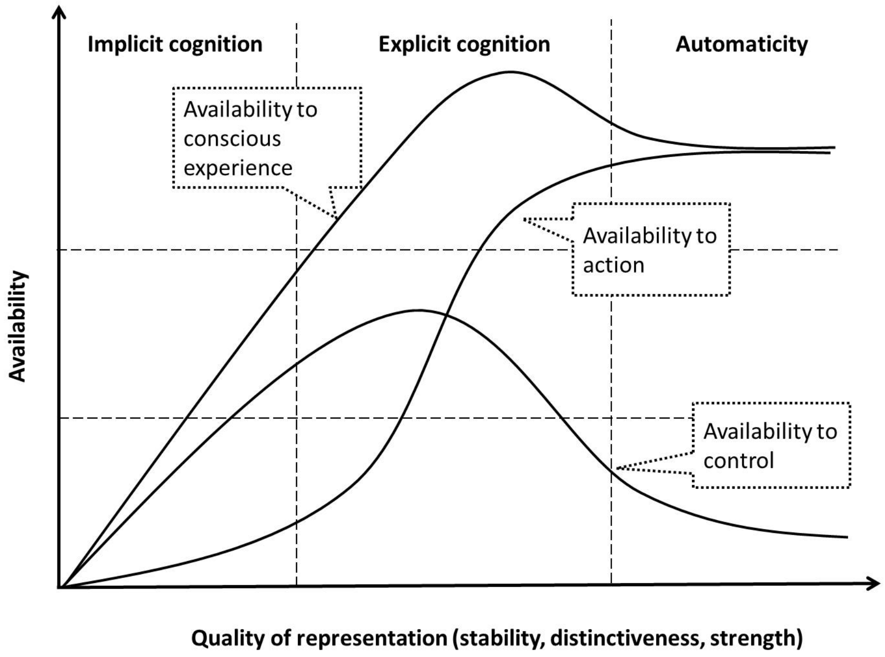

Figure 2. A graphical representation of the relation between quality of representation (X-axis) and availability to conscious experience, action, and control (Y-axis), according to Cleeremans and Jiménez (2002). It can be seen that all three aspects of availability gradually change from an implicit to an explicit state. Not further discussed in the current context is the area of automaticity, which is associated with a representation that became very strong due to extensive practice.

Interestingly, a few years later, Cleeremans (2008, 2011, 2014; Windey & Cleeremans, 2015; Windey et al., 2014) modified this theory of a graded consciousness in a way that, at first glance, might be interpreted as a two-system account. In fact, the new model could be seen as a hybrid between one- and two-system accounts. Instead of further elaborating the mechanisms behind the access to a global workspace, Cleeremans now takes the stance that global availability is not sufficient for consciousness. Rather, he turns towards the ideas of higher-order thought theories. Through interaction with the environment, a first-order representation is developed, gradually improving in quality, as originally assumed in the essay by Cleeremans and Jiménez (2002). The crucial addition to their former stance and the new adaption to HOTTs is that the acquired first-order information 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 metarepresentation; 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 a first-order system has changed, for example because something has been learned, and thereby develops a higher-order attitude towards the first-order knowledge (e.g. "I know that …", "I hope that …", "I see 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 systems. Obviously, this justifies labeling the model of Cleeremans as a two-system theory. Nevertheless, it also has to be considered how Cleeremans sketches the mechanism of the metacognitive higher-order system: It is stated that the representations of this system develop in the exact same way as first-order representations. The most important factor for a meta-cognitive, conscious representation to develop is that it gains stability, strength, and distinctiveness over the course of learning. More precisely, a first-order system is described as a simple feed-forward backpropagation network consisting of an input, a hidden, and an output unit (Figure 3). This network continuously develops increasing sensitivity to contingencies in the environment by developing associations between the input and the output and thereby improves so-called Type 1 responses (e.g. simple discriminative responses towards a stimulus). There is no intrinsic property within this network that is associated with consciousness. A similarly built second-order network is connected to this first-order network, receives input from there and learns about the internal states of the first-order network. It can learn how the internal state of the first-order network was, when a response was correct or when it was incorrect, and thereby develops higher-order judgement-knowledge about the first-order system possessing or not possessing knowledge in a given situation (Type 2 responses).

This general principle can, for example, be extended for distinguishing between imagination and factual perception, knowing or guessing, remembering or predicting. It can help in deciding whether a situation has changed or is the same, how consistent a situation is, how similar a situation is to associated situations and so on. Because the same strengthening mechanism lies behind first- and higher-order networks, consciousness is still a gradual phenomenon, changing on a trial-by-trial basis. This is what gives the proposed model characteristics of a single-system theory and justifies categorizing it as a hybrid between single- and multiple-system views.

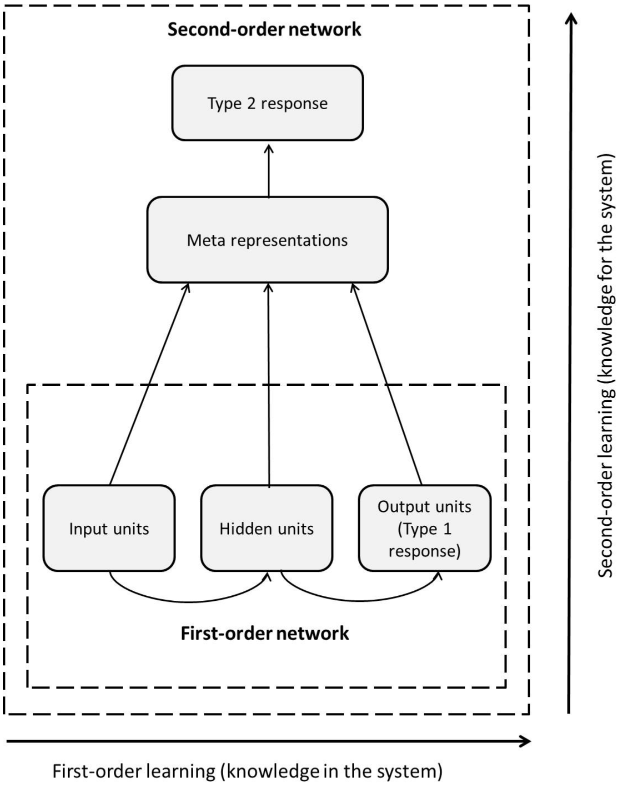
Second-order network
Type 2 response
Meta representations
Input units
Hidden units
Output units (Type 1 response)
First-order network
Second-order learning (knowledge for the system)
First-order learning (knowledge in the system)

Figure 3. A schematic representation of a metacognitive network according to Cleeremans (2011). The first-order network consists of a three-layer feedforward backpropagation network that gradually learns to produce a correct Type 1 response (e.g. a stimulus discrimination). The information from the first-order network is the input of the hidden units of a second-order network. The second-order network gradually learns to judge whether the first-order network has produced a wrong or correct response (Type 2 response).

3.2 Multiple-System Views

Opposed to single-system views, multiple-system views assume that implicit and explicit learning are supported by dissociable memory systems which build different forms of representations and are supported by different mechanisms. As the following section will show, multiple-system views are much more prominent in implicit learning research. This might be attributed to the fact that before implicit learning arose as a research topic around the 1960s (e.g. A. Reber, 1967), it was common to assume that human learning is usually guided by hypothesis-testing and leads to verbalizable, symbolic propositions (see e.g. models of information processing from Fodor, 1975; Newell & Simon, 1972). It might therefore seem intuitive that learning which happens in an incidental manner, without any intention or instruction and is not available for verbal report, is the product of a system that works differently from the system assumed in common information processing models.

The early works of Arthur Reber (1965, 1967, 1989) already imply that implicit learning should be differentiated from explicit learning processes in several ways. He argued that implicit learning should be imagined as a *differentiation* process, as opposed to an *enrichment* process. The former describes a primitive, rudimentary mechanism that develops sensitivity towards information in the environment and that is incapable of adding information to the provided material or using reflective strategies (Reber, 1967). While Reber (1965) first assumed that implicit knowledge can never be available to consciousness, he later (1989) assumed that additional conscious processes can interact with these implicitly learned representations and thereby result in conscious knowledge. Nevertheless, he did not specify this interaction any further.

A first theory that was more directly concerned with the interaction of separate implicit and explicit learning mechanisms was introduced by Willingham (1998) in form of the so called *COBALT* (control-based learning theory). In this model it is assumed that implicit and explicit learning processes work independent and in parallel. Primarily conceptualized as a neuropsychological theory of motor-skill learning, the COBALT is based on two separable neural processing pathways. The implicit learning system is supposed to be located in a dorsally located pathway. Learning in this network is less attention-demanding and leads to rather slowly developing but stable motor-skills. The representations developed by the implicit system are supposed to be coded in allocentric space. Explicit learning on the other side is supposed to be located in a ventral pathway. Learning within the explicit system is highly attention-demanding but also more accurate, strategic and can develop very quickly through, for example, instructions, observations or hypothesis testing and is more susceptible to forgetting. Representations in the explicit system are supposed to be coded in egocentric space. Both systems do not directly interact and develop knowledge completely independent of each other. The only interaction between both systems which is allowed within the COBALT is that the explicit system can override the implicit system but not the other way around.

A very similar account, more directly dedicated to implicit sequence learning than the COBALT, which is a more general theory for motor skill learning, has been developed by Keele, Ivry, Mayr, Hazeltine, and Heuer (2003). Their model adopts many of the ideas from the COBALT with some modifications. For example, also being a mainly neuropsychological model, the same mapping of an implicit learning system to a dorsal processing pathway and an explicit learning system to a ventral pathway is proposed. A main difference is that the model of Keele and colleagues is based on the assumption that implicit sequence learning is learning of contingencies within *dimensional modules*. Thus, the model is not exclusive to motor skill learning; dimensions can, among other things, refer to visual contingencies (e.g. colors or shapes; each being one dimension), auditive contingencies (e.g. pitch or timbre) or motor contingencies (e.g. hand or feet movements). It should be noted that Keele et al. acknowledge that the term dimension is not perfectly defined in their theory and a clearer suggestion can be found in Eberhardt, Esser, and Haider (2017).

Important to the theory of Keele et al. (2003) is that the dorsal, implicit learning system is exclusive to *uni-dimensional* learning. Information which is only correlated within one single dimension will be learned without the need for any attentional supervision. It follows that multiple sequences can be learned in parallel as long as none of the sequences compete for processing in the same module and none of the sequences depends on processing its covariation with another sequence in a different modality (e.g. a color predicted by a shape). Uni-dimensional learning is assumed to result in so-called *encapsulated* information, which means that it does not interact with any other module and therefore will always remain implicit. There is no mechanism that can transform this implicitly learned information into explicit knowledge. The explicit learning system on the contrary is responsible for *multi-dimensional* learning. Whenever two or more sequences covary over different dimensions, it is necessary that selective attention is drawn towards the relation between the involved dimensions. Because attentional supervision is needed, there is no parallel learning within the explicit system. Nevertheless, learning in the multi-dimensional system is still seen as an automatic process in a way that it does not need the intention to learn. All that is needed is that the correlating dimensions, and no uncorrelated dimensions, are specified as relevant by the current task set. The supposed automaticity of the explicit learning system is reflected in the assumption that any learned knowledge within this system also is implicit at first. The only difference to the implicit system is that this knowledge *can* become explicit, conscious knowledge. Keele et al. remain silent on what the sufficient and necessary conditions are for multi-dimensional learning to become conscious; they only state that "because such events are attended, they are accessible to processes underlying awareness and thus (…) can become explicit" (Keele et al., 2003, p. 317). Interestingly, even though the model of Keele et al. is commonly known as a dual-system theory (and is also included as such within this work) it is, much like the model of Cleeremans (2011), rather a hybrid between one- and multiple-system views. Different from the model of Cleeremans, Keele et al. do not assume any dependency of the explicit on the implicit learning system and rather propose two completely independent systems. Nevertheless, the model by Keele et al. also assumes that implicit and explicit learning can arise from one and the same (multi-dimensional) system and that both systems are based on the same learning mechanism.

All multiple-systems views, introduced so far, are very unspecific about the actual emergence of explicit sequence knowledge. They all are more oriented towards explaining how implicit learning can be realized and subsuming empirical findings under one unitary framework. They do not contain any references to theories of consciousness like the single-system view by Cleeremans and colleagues (Cleeremans, 2011; Cleeremans & Jiménez, 2002) did and rather presume that conscious knowledge can result from implicit learning but do not describe how it develops. There are, however, multiple-system frameworks consciousness which are specifically designed to distinguish implicit and explicit learning mechanisms and which take scientific theories of consciousness into account.

A first and very influential approach came from Dienes and Perner (1999) which is rooted in HOTTs. According to Dienes and Perner any representation can vary in their explicitness in three hierarchical structured constituents; these are *content*, *attitude*, and *holder*. A representation is defined as a propositional attitude which at least needs an explicit content. When a person learns a sequence in an SRTT, they will acquire a first-order representation of the sequential structure. The person can explicitly represent the content of that proposition. That means that the person is aware of the given stimuli and their responses and might even express experiencing some feeling of fluency or familiarity. However, any propositional attitude can furthermore comprise an attitude towards this content ("knowing", "imagining", "wanting", etc.) and lastly a holder ("I"). Dienes and Perner conceptualize implicit learning as being at least content explicit (more precisely as being *property* explicit as a further differentiation and most basic form of "content"; see Dienes & Perner, 1999) but attitude and holder implicit. However, for knowledge to become subjectively conscious (i.e. reportable sequence knowledge) the proposition also needs higher-order knowledge of the attitude (knowing that there is a sequence) and the holder ("it is I who knows that there is a sequence"). Important to the question how attitude and holder can become explicit, Dienes and Perner stated that observing one's own behavior and inner experiences (e.g. the feeling of fluency) can lead to inferences of one's own knowledge. These inferences constitute an explicit learning process that leads to a higher-order representation (i.e. attitude and holder explicitness) of one's own knowledge.

This conceptual framework of overserving one's own behavior by Dienes and Perner (1999) provides the basis for the *Unexpected Event Hypothesis* (UEH) which has originally been proposed by Frensch et al. (2003). The UEH aims to improve the description of the mechanism behind the initiation of explicit learning processes. Most importantly, it aims to improve the explanation how and when implicit learning can trigger an explicit inferential process. Concerning implicit learning processes, the UEH shares some assumptions with the models of Cleeremans and Jiménez (2002) as well as of Keele et al. (2003). Building on the computational stance on implicit learning, the UEH assumes that implicit learning is a byproduct of interacting with the environment. By repeatedly interacting with sequential information, associative weights will gain strength and implicit representations will develop. It is further assumed that this acquired knowledge is encapsulated within highly specialized subsystems. Unlike the model of Keele et al. implicit knowledge is not restricted to a dorsal processing path; instead, the modules which can acquire implicit knowledge are spread over different cortical and subcortical networks (see e.g. Conway & Pisoni, 2008; Gilbert, Sigman, & Crist, 2001, for overviews). For example implicit motor learning has often been associated with activity in the basal ganglia (see Karuza et al., 2013, for an overview), implicit visual learning has been found within the medial temporal lobe and the lateral and ventral occipital cortex (Rose, Haider, Salari, & Büchel, 2011; Rose, Haider, Weiller, & Büchel, 2002; Turk-Browne, Scholl, Chun, & Johnson, 2009) and auditory sequence learning within the primary auditive cortex (Kilgard & Merzenich, 2002). Similar to the model of Keele et al. (2003), it is postulated that implicitly acquired knowledge is unconscious due to its encapsulated nature and that there is no intrinsic conscious property to implicit knowledge, nor is there any additional mechanism that can transform it into explicit knowledge. Unlike the account from Cleeremans (2011), explicit learning does not develop through the same learning mechanism as implicit learning and does not require the slow strengthening of metacognitive knowledge. Explicit sequence learning instead constitutes an independent mechanism that is based on hypothesis testing.

The crucial idea of the UEH is that explicit sequence knowledge can only develop when a person unexpectedly notices a change in their own behavior. This can trigger an intentional search for the sequence. In an implicit learning situation, interaction with the task leads to 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 even an external event that directs the participant's attention towards noticing an underlying pattern and triggers following search processes. These search processes do not necessarily lead to a detection of the sequence if another explanation seems more likely to account for the unexpected event (Figure 4).

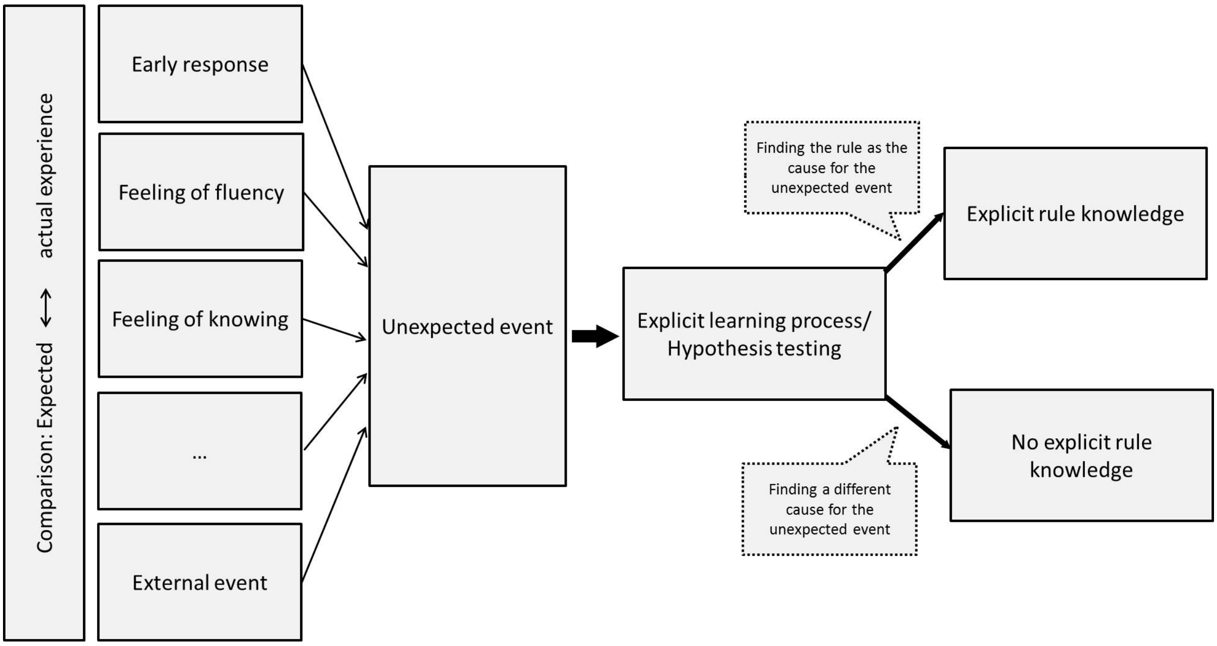
Comparison: Expected ⟷ actual experience
Early response
Feeling of fluency
Feeling of knowing
...
External event
Unexpected event
Explicit learning process/ Hypothesis testing
Finding the rule as the cause for the unexpected event
Explicit rule knowledge
Finding a different cause for the unexpected event
No explicit rule knowledge

Figure 4. A schematic representation of the Unexpected Event Hypothesis (Frensch et al., 2003). Any deviation between the expected and actual experience of a person constitutes an unexpected event. This comprises deviations from subjective, inner experiences (e.g. the feeling of fluency or knowing) as well as observable, external events (e.g. entering the correct response before target onset, constantly producing the same sequence of action-effects). The detection of an unexpected event leads to an explicit learning process which operates via hypothesis testing. Finding the hidden rule as the cause for the unexpected events leads to explicit rule knowledge, while finding another cause will hinder the development of explicit sequence knowledge.

Generally speaking, the UEH comprises a monitoring process which constantly compares expected and actual experiences. This comprises internal, experiential, as well as external, behavioral deviations from one’s expectation. This process allows detecting unexpected changes and initiates an attributional process for the detected conflict in order to adjust its predictions and reestablish coherence between the distant 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, Braver, Barch, Carter, & Cohen, 2001) metacognitive control (Kortiat, 2000, 2012, 2015), or memory (Whittlesea, 2002; Whittlesea & Williams, 2000).

Importantly, this leads to two differences to the model of Cleeremans (2011). First, although both theories assume an indirect relation between implicit and explicit knowledge, in the UEH the quality of the implicit knowledge is less important for the acquisition of explicit knowledge. In the model of Cleeremans, the first-order system has to produce mostly correct responses before the higher-order system can learn that something has been learned by the first-order system. In the UEH, even though a strong implicit representation is also more likely to lead to an unexpected event (e.g. by experiencing more automaticity in one’s own responses), an unexpected event can happen at any point during the learning process nevertheless. Any hint for an underlying sequence, even very early during the learning phase, for example noticing that no key has to be pressed successively, can serve as an unexpected event. In this example, the explicit learning process is triggered completely independent of the representational quality first-order knowledge. The second important difference is that explicit knowledge does not develop on a gradual trial–by-trial basis. Instead, once a person starts searching for a reason for the unexpected event, detecting and learning the sequence explicitly will happen via a rather sudden insight-process. It will only take a few trials from the search for the sequence to the fully developed explicit representation of it.

Concerning the compatibility with current views on consciousness, the UEH seems advantageous to the model of Keele et al. (2003). Multi-dimensional learning according to Keele et al. is a single process that starts with acquiring implicit representations which can develop into explicit representations. This implies that there is a certain network within the brain which is responsible for transforming unconscious into conscious representations. Even though Keele at al. add that attention is necessarily involved in multi-dimensional learning and that the ventral pathway processes categorized representations, this does not help to explain, in any way, why and how these representations can be conscious. Also, it is not defined if a difference between phenomenal and access consciousness is made. Without explaining which functions are related to explicit processing in their model, the ventral, multi-dimensional system creates a homunculus and it remains open at which point an implicit representation transforms into an explicit representation and how explicit representations differ from implicit ones.

The UEH instead is compatible with a functional framework of consciousness. There is no special mechanism or network that is supposed to be able to transform an implicitly learned representation into a conscious one. Neither is there an undefined gradual transition between unconscious and conscious states that leaves unspecified when exactly which functions of consciousness are enabled. Instead, the UEH can be positioned within a higher-order thought framework of consciousness as well as within a conceptualization of global availability of information. In its core, it assumes that consciousness requires learning that something has been learned. Concerning the HOTTs, the UEH proposes that sequence knowledge can only become explicit when a person has learned that they have learned something. Metacognition plays a major role in the UEH, as it assumes that there is no direct access to first-order knowledge; the only option to know that one knows (or not knows) is via observation of one's own behavior and its coherence with the held beliefs and assumptions about one's own first-order states. Furthermore, consciousness within the UEH is a dichotomous all-or-none state; once a representation is conscious, it is globally accessible to all functions of different networks.

The UEH is also compatible with global workspace theories of consciousness. Once an explicit representation of a sequence has developed, there is not only a subjective state of knowing, it will also be available for verbal report, flexible and strategic use (e.g. inhibition, transfer to a new task), integration with associated knowledge or stimulus-independent, intentional use. Importantly, this does not mean that consciousness cannot differ in the detailedness of the features that are represented explicitly. As Dienes and Scott (2005) suggested, explicit knowledge can be judgmental or structural. A person can perceive an unexpected event within an implicit learning situation and assume that an underlying structure of the task is the reason for that change. Nevertheless, the sequence might, for example, be very complicated or the participant is simply not interested in finding out the actual sequence. In this case, the participant will possess conscious *judgement knowledge*; they know that there is a sequence, but they don't know its exact structure. Another participant might detect the full sequence (or at least parts of it) and therefore will possess *structural knowledge*. Both participants can demonstrate all functions of consciousness about the exact features they are conscious about. The participants with judgmental knowledge can communicate that they know that there is a sequence or use this knowledge in a similar experiment in the future to start searching for a sequence from the beginning of the task. What they cannot do, for example, is intentionally producing the sequence in reverse order, when asked to do so. This would of course require explicit structural knowledge. Nevertheless, whether only judgmental or also structural explicit knowledge develops, is not dependent on the quality of the implicit representation but only on the explicit hypothesis testing process.

To summarize, the core assumptions of the UEH are that (a) implicit learning influences behavior, (b) this behavioral change is detected and does not match the expected performance, (c) this mismatch triggers a conscious attributional process, (d) this attributional process can lead to a detection of the sequence, and (e) the resulting explicit representation is a new representation, fully independent of the quality of the implicit representation and accessible to all functional networks.

4 Empirical Evidence for Single- and Multiple-System Views

Now that the theoretical outlines of single- and multiple-system views have been presented, it is important to take a look at the empirical evidence that supports each of these views. In order to understand the following experiments and their implications for the competing views, a short introduction into the methods of measuring implicit and explicit knowledge within the SRTT is necessary.

Factoring that there is little consensus on the existence of unconscious knowledge and the nature of conscious knowledge, there is also a broad variety of methods for measuring both kinds of knowledge. What a test does measure and what the demands for a test of conscious knowledge are, is highly dependent on the conceptualization of consciousness. The following section will therefore give a short overview of the most common tests and their interpretation under different conceptualizations of consciousness.

4.1 Measures for Implicit and Explicit Knowledge

Whenever a participant is trained within an SRTT, they will show an incremental decrease in their reaction times (RT) to sequential stimuli compared to non-sequential stimuli (even though this is somewhat more complicated to show for non-motor sequences; Haider, Eberhardt, Kunde, & Rose, 2012). This RT benefit is called an *indirect* test because it is merely a byproduct of task-performance. A *direct* test, which explicitly asks the participant to demonstrate what they have learned during training, is needed to identify whether their acquired knowledge is conscious or unconscious.

The most intuitive and also most used method of choice among the direct tests is to simply ask the participant to name the sequence after the training. If they cannot recall the sequence and possibly also claim that they have not even noticed that there was a sequence, it could be assumed that the knowledge is unconscious. Whether verbal report has any informative value about the implicit or explicit nature of the knowledge has been extensively criticized and discussed. In their very well-known critique, Shanks and St John (1994) stated that in order to be able to show that knowledge is truly implicit, a direct test must fulfill two criterions: The *information-* and the *sensitivity-criterion*. The information-criterion requires that the direct test has to measure the exact same information that was responsible for the performance in the indirect test. Verbal report does not fulfill this criterion because it asks for more than what might be needed for a decrease in RT. Instead of being able to name the whole sequence, the participant might have noticed that no stimulus position is used successively or might simply know about single, salient transitions in the task. Verbal report is therefore highly dependent on what the scientist considers to be important to ask for and asking for the wrong information (which might vary individually among the participants) might bias the participant and lead to an underestimation of explicit knowledge. The sensitivity criterion requires the direct test to be as sensitive as the indirect test. The test situation should ideally not differ from the training situation with the exception of the instructions and should motivate the participant to use any knowledge they have acquired during the training. Verbal report does not meet this criterion, either, because it creates a whole different recall situation and is less sensitive due to an unknown response-bias, motivation and expectation of the participant.

Preferred by Shanks & Johnstone (1999) and already suggested by Nissen and Bullemer (1987) is the so-called *generation task*. Here, the participants are confronted with a task that has the exact same surface as the training task. The only difference is that sometimes the stimuli sequence will be interrupted and the participants are asked to enter the next response they consider to be the most likely on their own. Knowledge is said to be conscious if the participant gave more correct responses than was to be expected by mere guessing. In a similar vein, in a *recognition test* (Shanks & Johnstone, 1999) participants are tested in a task highly similar to the training situation. Here, they are confronted with trials in which the training sequence is presented and trials in which a new sequence is presented. It is then their task to rate whether a just presented sequence was “old” or “new”. Again, classification performance above chance-level is considered to demonstrate explicit knowledge.

Besides various methodological criticism which is too broad to be discussed here (see, for example, Buchner, Steffens, Erdfelder, & Rothkegel, 1997; Jiménez, Méndez, & Cleeremans, 1996, on the generation- and the recognition task), there is one very profound methodological and one, maybe even more important, theoretical problem with these tasks. On the methodological side, the idealistic demand for a test that is process-pure for implicit or explicit knowledge has been criticized by Reingold and Merikle (1988). They stated that any behavior is always a product of conscious and unconscious processes. On the theoretical side, the search for the most sensitive measure has led to an operationalization of consciousness that is at odds with most functional accounts of consciousness. Making a correct predictive response might as well be the product of unconscious processes as it does not show that there is any subjective consciousness nor does it show that the knowledge can be used in any strategic way. These functions of consciousness cannot be assessed if the test is identical to the training situation.

To tackle the problem of process-pureness, Jacoby (1991) developed the *process-dissociation-procedure* (PDP) which has been adapted to the SRTT by Destrebecqz and Cleeremans (2001). It is the aim of the PDP to estimate the relative contributions of implicit and explicit knowledge on task performance. In the PDP, the participants face two intra-individual test conditions which both are highly similar to the training condition. Participants respond to the same sequence as during training, but sometimes, no stimulus is presented and they are asked to give the response they consider to be the most likely to occur next (*inclusion condition*; similar to the generation task). It is assumed that explicit and implicit knowledge contributes to the performance within the inclusion condition. On other trials, they are asked to give the response they consider to be most unlikely to occur next (*exclusion condition*). Here, it is assumed that implicit and explicit knowledge work antagonistically. Implicit knowledge will lead to sequence-conform responses (intrusion errors) while only explicit knowledge can be controlled in a way that sequence conform responses can be avoided. A high proportion of correct inclusion responses together with a low proportion of intrusion errors is supposed to be an indicator of explicit knowledge. The PDP can meet the information criterion to a high extent because the inclusion- and exclusion condition only differ from the training task in their instructions. At the same it is compatible with functional theories of consciousness like the GWT because it asks for strategic use of the knowledge during the exclusion condition. However, it should be noted that Barth, Stahl, and Haider (2016) demonstrated that a core assumption of the PDP, namely the invariance of implicit and explicit processes across inclusion and exclusion conditions, may be violated within the SRTT paradigm and therefore yield unreliable measures.

Moreover, the PDP still suffers from the shortcoming that it does not inform us about the subjective state of the participant. Dienes and Perner (1999) therefore asked for a test that combines objective and subjective performance while at the same time, it should try to meet the information and the sensitivity criterion to a high extent. In implicit learning research, an appropriate procedure has been suggested by Persaud, McLeod, and Cowey (2007) and adapted for the SRTT by Haider, Eichler, and Lange (2011). In the so-called *wager task*, the test task is constructed similar to the training task. Analogous to the generation task, participants are instructed to give the response they consider to be the next correct one whenever the task is interrupted and no target stimulus is shown. Right after their response, they are asked to rate their subjective confidence by wagering on their response. For example, they are instructed to bet 1 Cent or 50 Cent on the correctness of their response, reflecting their respective confidence. According to the zero-correlation criterion, implicit knowledge is assumed when a participant shows an amount of correct responses that is above chance-level while at the same time being unable to accurately rate the correctness of their responses. Overall, due to the high similarity between training and test, the wager task provides a highly sensitive task for measuring implicit and explicit knowledge. In addition, due to the wager component, it is also compatible with current functional theories of consciousness as the GWT as it tests for the ability to use one's knowledge strategically. Moreover, the wager task is also satisfying from a HOTT perspective because it also captures the subjective aspect of sequence knowledge by including the confidence rating. Today, the wager task is a widely accepted measure for explicit knowledge within implicit learning research and also has been the subject of various methodological improvements (see e.g. Fleming & Lau, 2014; Massoni, Gajdos, & Vergnaud, 2014; Pasquali, Timmermans, Cleeremans, 2010; Sandberg, Timmermans, Overgaard, & Cleeremans, 2010;).

Lastly, it should be noted that Rünger and Frensch (2010) have made a strong theoretical argument for also continuing to use verbal report as an important measure for explicit sequence knowledge with respect to the GWT. They argue that not sensitivity but exclusivity is the most important criterion for a test of explicit sequence knowledge. According to their theoretical conception, sensitive tests tend to overestimate explicit knowledge; what is needed is a test that is impregnated against implicit knowledge and strongly exclusive to explicit knowledge. Verbal report would be the best candidate for such a test. In order to deal with the critique about verbal report being very susceptible to response-biases of the participants, Rünger and Frensch argue that these problems can be overcome by careful construction of the questionnaire. It should begin with very open questions and become increasingly specific, thereby motivating the participant to mention any knowledge they have while reducing insecurity (Eriksen, 1960).

To summarize, the wager task as well as verbal report have been shown to be very valuable measures of explicit sequence knowledge and together are the most commonly used methods in implicit learning research. Moreover, Haider et al. (2011) have shown that both measures are correlated to a very high extent, making a final decision of one method over the other superfluous for the current research questions presented in the following studies.

4.2 Empirical Evidence for Single-System Views

As introduced in Chapter 3.1, single-system views can differ crucially on whether they assume that there are distinct types of acquired knowledge (i.e. implicit and explicit) within an SRTT. According to Shanks and colleagues (Shanks, 2005; Shanks & Perruchet, 2002), it has yet to be shown that performance in an SRTT is influenced by implicit knowledge. Their argument is mainly a methodological one. Any difference shown between an indirect and a direct test, which would be necessary to demonstrate dissociable learning systems, can be explained by the tests being unequal in their information- and sensitivity criteria. Different transformation processes are needed when a learned representation facilitates the response within an SRTT or when the participant is asked to, for example, verbally report the sequence. To support their single-system view, Shanks and Colleagues use recognition (Shanks & Johnstone, 1999) or generation tasks (Speekenbrink, Channon, & Shanks, 2008). Employing these highly sensitive tasks, the authors show that there is no performance difference between the direct test and the indirect test. It is therefore argued that knowledge acquired within an SRTT (and other tasks that are supposed to demonstrate unconscious processes) is accessible to the participants. However, as discussed in Chapter 4.1 these kind of direct tests do not fit the functional approaches to consciousness very well. There is neither any criterion showing that the participants have any kind of subjective experience of being conscious of that knowledge, nor is any higher, strategic cognitive function required by those tasks. For that reason, the discussion of these studies on single-system explanations is not further extended within this work.

Of higher interest are the studies by Cleeremans and colleagues (Cleeremans, Timmermans, & Pasquali, 2007; Pasquali et al., 2010). They assume that implicit and explicit knowledge can be dissociated but that both forms of knowledge arise from the same learning process, which is mainly dependent on the strengthening of associations throughout practice. Importantly, their work includes direct tests, like the wager-task, which are more congruent with the discussed functional theories of consciousness. In their studies, artificial neural networks following the metacognitive learning architecture suggested by Cleeremans (2008; 2011, see Chapter 3.1), are modeled. Cleeremans et al. (2007) trained a first-order feedforward network to learn a simple digit discrimination task (i.e. discriminating which number from 0-9 served as an input to the network). The input network consisted of input, hidden, and output units. The activation pattern of the hidden units was copied into a second input-layer of an independent higher-order feedforward network. These input units were again connected to a hidden layer of the higher-order network which, in turn, was connected to output units, representing the judgement of the higher-order network about the correctness of the output of the first-order network. This certainty could be expressed in a “high wager” if the higher-order network decided that the output was correct, or in a “low wager” if it decided that the output was incorrect. The first-order network gradually learned to improve the digit-classification. The higher-order network showed a different, u-shaped, learning curve. It started with a fairly good performance (i.e. high wagers on correct outputs of the first-order network and low wagers on wrong outputs), then dropped to chance-level and gradually improved again until it reached 100% correct responses. This can be explained by the performance of the first-order network. The first-order network starts with a very bad classification performance; the second-order network learns that the first-order network is almost always wrong and accordingly correctly puts low wagers on the correctness of the first-order network. When the first-order network surpasses chance-level, the second-order network’s performance will drop until it has learned that the state of the first-order network has changed and thereby begins to place high wagers on correct outputs.

In a subsequent study, Pasquali et al. (2010) aimed to compare the results produced by their network to data produced by participants in an implicit learning situation. Therefore, they used the data of Persaud et al. (2007) with which the wager-task has been introduced as a measure of subjective consciousness within an Iowa Gambling task (Bechara, Damasio, & Anderson, 1994), blindsight (Stoering, Zontanou, & Cowey, 2002) and Artificial Grammar Learning (A. Reber, 1967). The network built by Pasquali et al. (2010) replicated the participants’ data on all three tasks. Together these studies show, on a computational level, that metacognitive learning can be used to explain the dissociation between performance in indirect and direct tests in an implicit learning situation.

Besides the studies led by Shanks and colleagues, in which tests of ambiguous informative value have been used to assess explicit sequence knowledge and the simulation data of Cleeremans and colleagues, hardly any studies supporting a single-system view on the emergence of explicit knowledge in an implicit learning situation can be found.

4.3 Empirical Evidence for Multiple-System Views

Corresponding to the larger number of multiple-system views, there is also much more research dedicated to investigating them. This comprises modeling of neural networks which propose different learning mechanisms and representational formats for implicit and explicit knowledge (e.g. the CLARION model by Sun, Slusarz, and Terry, 2005; the TELECAST model by Hélie, Proulx, and Lefebvre, 2011). Concerning the question of potentially dissociable neural bases of implicit and explicit learning, there are neuropsychological studies on amnestic patients showing the involvement of different networks within an SRTT. These studies generally point to the circumstance that amnestic patients with heavily impaired declarative memory due to damage in the medial temporal lobe (MTL) show an inability to explicitly learn a sequence within an SRTT. Nevertheless, the implicit learning skills of these patients are comparable to healthy control groups (P. Reber & Squire, 1998; Vandenberghe, Schmidt, Fery, & Cleeremans, 2006; but see Speekenbrink et al., 2008, for a different view). There also are neurological studies with healthy participants pointing to a possible neurological dissociation between explicit and implicit learning systems. These studies usually demonstrate an association between increased activation within the MTL, generally characterized as an important network for declarative knowledge, and explicit sequence learning (Keele et al. 2003; Poldrack et al., 2001; P. Reber, 2013), even though MTL activity has also been demonstrated for implicit non-motor sequence learning (Rose et al., 2011; Rose et al., 2002; Turk-Browne, 2009). Furthermore, activity within the dorsolateral prefrontal cortex, the parieto-occipital, the premotor and inferior temporal cortex, and the anterior cingulate cortex, subsumed by Keele at al. (2003) as the ventral stream, has been associated with explicit sequence learning (Destrebecqz et al. 2003; Grafton, Hazeltine, & Ivry, 1995; Hazeltine, Grafton, & Ivry, 1997). More specifically, the ventrolateral prefrontal cortex and the ventral striatum showed increased coupling of their activity shortly before participants indicated awareness of the implicitly learned sequence (Rose, Haider, & Büchel, 2010; Wessel, Haider, & Rose, 2012). These latter studies provide interesting support for the role of the detection of unexpected events in the acquisition of explicit knowledge because the investigated network has been associated with feedback-based prediction learning and the implicit detection of violations from learned associations (Rose, Haider, & Büchel, 2005; Seger, 2008).

In congruence with the assumption that implicit learning can be found for various distinct response- and stimulus dimensions and that this learning happens within encapsulated modules (Eberhardt et al., 2017; Keele et al., 2003), implicit learning can be found in widespread subsystems within the brain. Thereby, the most robust finding is the involvement of the basal ganglia as long as motor sequences are involved (Karuza et al., 2013) whereas the much less investigated non-motor visual learning has been associated with activation in the lateral and ventral occipital cortex, the posterior caudate and the MTL (see P. Reber, 2013, for an overview; Rose et al., 2011).

Showing that different neural networks are involved in explicit and implicit learning might be seen as a hint towards different learning mechanisms being involved. However, given the multitude of multiple-system views as well as the huge variability in the actual experimental designs and the exact structure of the sequences involved, the neuroimaging data are far from being able to support any specific theoretical view on the mechanism behind the emergence from explicit knowledge in an implicit learning situation. More interesting are studies directly aimed at differentiating between single- and multiple-system views. There are a few experiments directly focused at testing parsimonious single-system accounts against the assumption that additional explicit learning mechanisms are needed to gain insight over encapsulated implicit knowledge.

The earliest study directly aimed at this question came from Haider and Frensch (2005). In this study the authors worked with a number reduction task (NRT; Frensch et al., 2003; Thurstone & Thurstone, 1941) instead of an SRTT. In the NRT, participants compare two digits and respond to them being identical or not identical. Like in the SRTT, there is an underlying hidden rule which is more abstract in the NRT. Despite these differences, both paradigms have in common a typical decrease in reaction times when the hidden rule is learned implicitly. In this study the authors tested a crucial prediction of the UEH: Unexpected events should trigger an attributional process that leads to explicit rule knowledge. More precisely, they programmed unexpected events into the task by having the program erroneously record a premature response, i.e. a response made before the target stimulus appeared. The participants were made to believe that they were responsible for the premature responses. It was manipulated whether the instructions gave the participants a sufficient explanation for these responses happening (i.e. attentional lapses), while the other group was given no explanation. The idea was that only participants without a given explanation would look for a reason for their erroneous behavior and therefore would have a higher likelihood to detect the hidden sequence. However, the results were not completely unequivocal. The group that was not offered a cause for their premature responses did show more explicit knowledge than a control condition without any premature responses inserted. At the same time, the group that did have a cause offered did not differ from that control group. Still, the groups with and without an offered cause for their premature responses only differed numerically but not significantly from each other. This might have been due to the problem that the offered cause (i.e. attentional lapses) did not necessarily prevent participants from wondering how they were able to give a correct response before the target stimulus was even shown. Therefore, there still might have been a substantial, though smaller, proportion of participants that were surprised by the unexpected event and searched for a cause outside of the simple explanation of inattention.

In another NRT study, Haider and Frensch (2009) manipulated the chance to experience unexpected events produced by the participants themselves. Here, the authors manipulated the response-stimulus interval (RSI); there either was a 500ms delay before the next stimulus was presented after the last response or it followed 250ms after the last response. Their assumption was that a long RSI would raise the likelihood for premature responses, while an RSI of 250 ms would make them less likely. The UEH therefore predicts an increase in the amount of explicit knowledge for the group with a 500 ms RSI. The results corresponded with this prediction. Because the amount of practice was equal between the conditions, it seems that the results are in favor of the UEH and against a simpler single-system view. However, there also was a significant performance difference between the 500 ms and the 250 ms RSI condition, with the 500 ms Condition showing shorter RTs. This might of course, in accordance with the UEH, be an effect of the increase in explicit knowledge due to the increased amount of unexpected events. But it might also be a sign that the 500 ms RSI condition learned more about the sequence, which, in accordance with a single-system view, gradually led to more explicit knowledge. In fact, Destrebecqz and Cleeremans (2001, 2003) argued that a longer RSI provides more opportunities to associate memory traces of high quality and thereby to develop higher quality representations. In a computational model they assume that there is a perception network which can produce motor responses and an additional memory network that can interact with the perception- and the motor network and learn about the sequence, but only if given enough time (Destrebecqz & Cleeremans, 2003). When the RSI is short, there is less chance that memory representations can develop. Even though this explanation has later been disputed by Rünger (2012), it is necessary to manipulate unexpected events in diverse other ways to strengthen the UEH explanation behind these data.

Such a different way of manipulating unexpected events, this time within the SRTT paradigm, has been explored by Rünger and Frensch (2008) as well as by Schwager, Rünger, Gaschler, and Frensch (2012). Rünger and Frensch (2008) manipulated whether the SRTT training contained only regular trials (control group) or was interrupted by either a new sequence or by random trials. Their assumption was that the interruption by a new sequence or by random trials would lead to an increase in RT, which the participants could notice as an unexpected event. Therefore, the groups who received additional random trials or trials with a new sequence should develop more explicit sequence knowledge. The amount of trials with the actual training sequence was kept constant across all groups. Therefore, the representational strength was supposed to be kept equal across the three conditions, implying no difference would be expected by single-system views. However, the authors did not find the expected advantage for any of the experimental groups. They only found less participants without any verbalizable knowledge in the condition which sometimes was interrupted by a new sequence.

Schwager et al. (2012) chose a similar approach. Here, participants were trained with either random material or with regular material. In a subsequent manipulation phase, they either received the same regular sequence as during training or a new regular sequence. Single-system views would predict that the group who had the same sequence during training and the manipulation phase should show the most explicit knowledge. If however, the change to a new sequence is perceived as an unexpected event, for example by recognizing an increase in the RT, this should, according to the UEH, lead to more explicit knowledge. Again, the results did not completely match the predictions. The group with the regular training and the new sequence in the manipulation phase did not show more explicit knowledge than the group who received the same sequence in both phases. Only the group with random training showed less explicit knowledge than the two groups who were trained with a sequence. Nevertheless, it remains an interesting point that from the two groups who received the new sequence in the manipulation phase, those participants who had already had a different sequence during training had more explicit knowledge of the new sequence than the control group who had had no former sequence. Both groups therefore had the same amount of practice with the new sequence in the manipulation phase. This result however is not unambiguously diagnostic about either a single-system approach or the UEH. It is not a clear prediction of the UEH that the group trained with random material before showed less knowledge than the group who had already received a different sequence. Eventually, the random group might as well be expected to perceive a salient unexpected event, as they changed from more difficult random material to an increasingly fluently-feeling sequential structure. A single-system account might explain the finding when strengthening of the training sequence led to explicit knowledge of the training sequence. It is trivial that a group who has learned that there is a sequence in the training would also search for the new sequence in the manipulation phase.

The studies discussed so far focus on the core assumption of the UEH, namely that there is a causal relation between experiencing an unexpected event and the emergence of explicit knowledge. There is another row of experiments which also should briefly be mentioned here because they focus on another important assumption that separates the UEH from single-system accounts. The following studies focus on the sudden transition from implicit to explicit knowledge by converging behavioral with neuroimaging data. A sudden insight emerging from hypothesis testing is opposed to single-system views which all assume a slow and gradual transition. A first investigation has been reported by Haider and Rose (2007). They showed that participants who were able to verbally report a sequence at the end of training showed differences in their RT performance compared to participants whose knowledge remained implicit. While the latter showed a typical gradual decrease in their RT, participants who had developed explicit knowledge showed a distinctive, sudden drop in their RTs which is likely to reflect the moment a participant was able to switch from stimulus- to plan-driven control. This finding was more thoroughly investigated in a study by Rose et al. (2010). Here the concurrence between an RT drop during training and the subsequently tested ability to name the rule was further established. Moreover, the authors could show that a clear change in neural activity in the ventrolateral prefrontal cortex and the ventral striatum preceded this RT drop. This cortico-subcortical network has previously been associated with expected value and the representation of prediction errors (see e.g. Chase, Kumar, Eickhoff, & Dombrovski, 2015, for a meta-analysis). In the context of the UEH an increase in activity in these regions might thus affirm the assumption that an unexpected violation of predictions is relevant for triggering explicit learning processes which soon after lead to a drop in the participant's RTs. Furthermore, both studies support the assumption of a sudden insight which leads to a qualitative change in information processing, rather than a gradual transition towards conscious knowledge.

Another interesting finding in that context came from Schuck et al. (2015). Here, the authors did not use an SRTT, but a similar task where participants were instructed to perform a simple discrimination task which, unbeknownst to the participants, contained a visual sequence that, if discovered, would allow a simpler task processing. Increased activation of the medial prefrontal cortex was associated with the exploration of color information that could be used for a strategy shift. This signal predicted which participants would show a strategy shift a few minutes later. This is one of the first studies to show that an active exploration of the task material, as it is proposed by the UEH, precedes the insight into an implicitly learned rule. Taken together, the results presented here all show very interesting, converging and diverse evidence that unexpected events are an important trigger for attributional processes which in turn can lead to explicit learning of an implicitly learned sequence. It is the aim of the following studies to build on these findings. The studies are designed to improve some of the issues discussed about the just presented studies. Furthermore, since the UEH is a very broad framework that allows many different opportunities to experience unexpected events and should account for different forms of implicit learning, the following experiments should also broaden the methodological horizon of manipulations within an implicit learning situation.

5 Overview of the Studies

Three studies were conducted with the aim of providing further insight into the mechanisms behind the emergence of explicit knowledge in an implicit learning situation. More specifically, all three studies were aimed at testing a parsimonious single-system strengthening account against a multi-system account represented by the UEH.

The first study was methodologically close to the studies of Rünger & Frensch (2008) and Schwager et al. (2012) presented in the former chapter. Its main aim was to manipulate the subjective feeling of fluency as an unexpected event between the participants, while also trying to find some methodological improvements to these studies.

Studies 2 and 3 took a different approach to investigating the assumptions of the UEH. In these studies we investigated the role of action-effect learning (Elsner & Hommel, 2001; Hommel, Müsseler, Aschersleben, & Prinz, 2001; see Shin, Proctor, & Capaldi, 2010, for a review) which, in previous studies has repeatedly been linked to learning within the SRTT (Stöcker & Hoffmann, 2004; Tubau, López-Moliner, & Hommel, 2007; Ziessler, 1998). Moreover, various studies showed increased explicit sequence learning when action-effect learning was involved (Hoffmann, Sebald, & Stöcker, 2001; Stöcker, Sebald, & Hoffmann, 2003; Ziessler & Nattkemper, 2001; Zirngibl & Koch, 2002). However, there has not been a direct investigation of the role of action-effect learning on explicit sequence knowledge so far. Therefore, studies 2 and 3 were designed to fortify this empirical evidence and also apply the UEH as an explanatory framework for these effects.

5.1 Study 1: The Emergence of Explicit Knowledge in a Serial Reaction Time Task: The Role of Experienced Fluency and Strength of Representation

The goal of Study 1 was to directly test the role of associative strength against the role of unexpected events on the emergence of explicit knowledge in an implicit learning situation. As discussed in Chapter 4.3 there only have been a few studies that directly tried to test these theories against each other. So far, these studies yielded some interesting insights by trying out various manipulations to keep the associative strength equal across the different conditions, while at the same time trying to manipulate the likelihood of unexpected events and the possible causal attributions for them. Taken together, these studies provide converging support for the UEH. However, individually they also showed that the manipulation of the relevant factors can become very tricky and that the paradigm as well as the complex research question at hand needs a multifaceted approach to deal with various unwanted side-effects and alternative explanations. The following study built on the logic of the studies from Haider and Frensch (2005, 2009), as well as the studies from Rünger and Frensch (2008) and Schwager et al. (2012) by trying to deal with the challenges these studies encountered.

The current study and all following ones used the SRTT paradigm, instead of the NRT, because it allows for a much broader variability of the sequences used and also has a big advantage through the great amount of research behind it. This in turn opens up a broader spectrum of accepted tests of knowledge. The essential studies by Haider and Frensch (2005, 2009), Rünger and Frensch (2008) and Schwager et al. (2012) all used verbal report of the sequence as the test for explicit knowledge. As discussed in Chapter 4.1 verbal report has its merits. Still, their results often did not show the expected differences and a more sensitive test might help to detect any differences in explicit knowledge more reliably. Therefore, in the current and all following studies we used the wager task (Haider et al., 2011) to estimate explicit knowledge.

Furthermore, because the UEH allows many different sources for unexpected events, we aimed to augment the empirical basis by using a new manipulation for the occurrence of unexpected events. So far, the studies by Haider and Frensch (2005, 2009) mainly used externally determined events (i.e. premature responses inserted by the program) which the participants were supposed to perceive as self-produced. This was a very good way to ensure that all participants encountered the exact same training phase. However, the UEH assumes that implicit learning should lead to changes in one's own behavior and that perceiving these changes leads to an attributional search process. For this reason, it would be helpful to find a manipulation that also covers the need to keep the training as comparable as possible while at the same time leading to different perceivable changes in one's own behavior. Rünger and Frensch (2008; Experiments 2a & 2b) tried to achieve this by inserting random blocks into the training phase. This should lead to a perceivable and unexpected slowing of the responses, in turn leading to attributional search processes, compared to a group with the same amount of regular trials but without any random trials. Rünger and Frensch however did not find the predicted differences in explicit sequence knowledge. This might be because the random blocks were too long (120 trials) and participants' search processes were not successful since they might have been triggered and started but also stopped again before the sequence had reappeared. Besides that, the results might also be in favor of a single-system strengthening account because participants, at least numerically, showed less sequence knowledge and slower RTs when trained with additionally inserted random blocks. It could be argued that inserting random blocks led to a weakening of the associative strength and therefore also to less sequence knowledge.

In the following study, we tried to implement an improvement to the design of Rünger and Frensch which should tackle the just mentioned difficulties of their study. We aimed for a design that kept the amount of regular and irregular trials equal across all groups, therefore not allowing any differences in the associative strength. The critical manipulation was the arrangement of the regular and irregular trials. For one condition the regular and irregular trials were mixed randomly, while for the other group both trial-types alternated every 22 (Experiment 1 & 2), respectively every 88 trials (Experiment 3). The expectation was that only the participants who encountered these alternating mini-blocks of 22, respectively 88 random and regular trials should be able to experience a difference in the experienced fluency between both types of material. Participants who were trained with both types of material mixed randomly should not be able to experience any differences in the fluency of the material because the material changes are too frequent. If the differences in the experienced fluency represent an unexpected event, the group who received the regular and irregular material in alternating mini-blocks should also show more explicit sequence knowledge.

Another improvement lay in the composition of the three experiments in Study 1. The UEH postulates that there is no direct relation between implicit and explicit knowledge. To increase the informative value of our study, we aimed to show that the difference in the arrangement of the regular and irregular trials has an influence on the explicit knowledge acquisition, but does not lead to a difference in the acquired knowledge in general. Therefore, in Experiment 1 we tried to make the emergence of explicit knowledge highly unlikely by presenting one group rather short alternating mini-blocks (22 trials) of regular and irregular material compared to a group who received randomly mixed material. Our expectation was that these mini-blocks were too short to allow the participants to explicitly learn the sequence. Experiment 1 should therefore provide an estimation of the difference in acquired sequence knowledge that does not go back to explicit learning. Our manipulation of the arrangement of the training material should generally not lead to any differences in the knowledge base of both groups. However, because we also expected the different arrangements to lead to differences in the experienced feeling of fluency, differences in the *performance* during the training task could also be expected. Participants who alternatingly experienced more and less fluency during the mini-blocks might show greater differences in their RTs than participants who received the randomly mixed material, who might generally be a little slower on both types of material. We used the wager task as a very handy option to estimate the acquired knowledge without resorting to RTs.

In a second step we directly aimed to test whether our manipulation led to the predicted subjective differences in experienced fluency of the training material (Experiment 2). This direct assessment of the subjective metacognitive attitudes towards the training material is a new, explorative approach to test the assumptions of the UEH. So far, the anteceding studies have only tried to ensure their manipulation affected the subjective perception of the participants' behavior by retrospectively asking them about what they perceived. Here, we tried to implement a more sensitive measure of the subjects' experiences by letting them rate their experienced differences in the fluency of the regular and irregular material within the SRTT itself. The group who received the regular and irregular material arranged in mini-blocks should rate the regular material as feeling more fluent than the irregular one, while the group who received both materials mixed randomly should not be able to rate one material as feeling more fluent than the other.

Finally, in a last step, we extended the length of the mini-blocks of regular and irregular material to 88 trials per mini-block. This should, opposed to the shorter 22 trial mini-blocks, allow the group that is trained with this blocked material to find a sequence, if in fact the difference in the experienced fluency triggers subsequent search processes. Therefore, Experiment 3 aimed to show that the group that wa trained with such longer alternating mini-blocks of regular and irregular material should acquire more explicit knowledge than the group that received both types of training material mixed randomly.

All three experiments taken together provided a clear, gradual deduction of the assumptions of the UEH. Experiment 1 showed that our manipulation of the subjective feelings of fluency did not lead to a different amount of acquired (implicit) knowledge, expressed in a subsequent wager task. Experiment 2 demonstrated that the manipulation of the trial-type arrangement did lead to differences in the experienced fluency. Lastly, Experiment 3 showed that the manipulation of the experienced fluency did lead to differences in explicit knowledge. More explicit knowledge was found for participants who could perceive a difference in the fluency between regular and irregular trials due to the blocked arrangement of these trial-types. Together the studies can show that the same underlying implicit knowledge base can result in different consciously perceivable changes of behavior and that it is this subjectively experienced, unexpected change that leads to an explicit search for the reason of this experience. This search ultimately can lead to explicit sequence knowledge.

5.2 Study 2: Implicit Visual Learning: How the Task Set Modulates Learning by Determining the Stimulus-Response Binding

The main goal of study 2 was to investigate the role of response-effect learning (R-E learning) on the emergence of explicit sequence knowledge. R-E learning is generally considered to be an important mechanism for intentional action control, as opposed to habitual or stimulus-dependent behavior (Hommel et al., 2001; Shin et al.; 2010; Balleine & O'Doherty, 2010). Hence, it might not be surprising that several studies have found a relation between R-E learning and the emergence of explicit sequence learning in an implicit learning situation (Hoffmann, Sebald, & Stöcker, 2001; Stöcker, Sebald, & Hoffmann, 2003; Ziessler & Nattkemper, 2001; Zirngibl & Koch, 2002). Yet, most of these studies rather considered R-E learning as a general mechanism guiding implicit sequence learning and which is comparable to R-R, S-S or S-R learning (e.g. Ziessler, 1998). Only a few studies have given the finding of an increase in explicit sequence knowledge special consideration. For example, Stöcker & Hoffmann (2004) suggested that auditive action effects lead to chunking processes that integrate these auditive effects into a melody and that it is these chunking mechanisms that promote the emergence of explicit knowledge. Similarly, Tubau et al. (2007) argued that auditive action-effects are integrated into a melody through the phonological loop. Because the phonological loop also plays an important role in plan-guided behavior, auditive action effects might promote plan-guided, as opposed to stimulus-guided behavior in an implicit learning task. The assumption that integrating action-effects into a higher-order structure (i.e. chunks) leads to explicit knowledge resembles a single-system account of explicit learning. However, the integration of a sequence into chunks might also be the effect of and not the cause for explicit sequence knowledge. The UEH provides an alternative explanation for the observation of an increase in explicit sequence knowledge. The essential link between the UEH and R-E learning is that both crucially depend on the perception of distal changes in the environment brought upon by one's own behavior. Different research points to the circumstance that R-E associations will only be acquired when the effects in the environment are interpreted as being caused by one's own actions (Herwig & Waszak, 2009, 2012). It has furthermore been shown that acquiring an action-effect association leads to the conscious anticipation of the effect before the action is initiated (Blakemore, Wolpert, & Frith, 2002; Haggard & Chambon, 2012; Moore & Haggard, 2008). It is this conscious anticipation of an action-effect that enables it to control intentional behavior (Hommel et al., 2001; 2017). Integrating the research in R-E learning into the UEH leads to the idea that learning R-E relations within an SRTT has a much greater likelihood to lead to perceivable changes in one's own behavior than learning about R-R, S-S or S-R relations. The latter three mostly lead to responses becoming faster and making fewer errors, which, if noticed at all, both can easily be attributed to mere practice effects. Different from that, experiencing that an action contingently leads to the same perceivable change in the environment, or even to a whole sequence of perceivable events, is a very salient unexpected event that points towards the existing of an underlying sequence – especially if these effects slowly start to be actively anticipated.

The following study was a first try to provide further evidence for the role of R-E learning in the emergence of explicit sequence knowledge. In their design, the experiments were built on a study by Haider et al. (2012). In this study, the authors used a very subtle manipulation to affect the stimulus-response mapping. Participants either responded to stimuli with the keyboard, as in the usual SRTT setting, or responded via mouse by clicking on the response stimuli displayed on the screen. Additionally, it was manipulated whether there was a pure visual sequence (i.e. in the colors of the target stimuli without any sequential motor responses) or a pure motor sequence (i.e. with random target colors) which the participants responded to. The interesting finding was that implicit learning seemed to be unaffected by the response device, but explicit learning critically depended on the combination of response device and sequence type. While the motor sequence was learned explicitly with the mouse as well as with the keyboard, only the participants who responded with the mouse were able to acquire explicit knowledge about the visual sequence. These results were tentatively explained with an R-E learning mechanism. The assumption was that if participants respond with the keyboard they encode mainly information about the location of the response. Responding with the mouse instead leads to encoding of both the color and the location of the clicked response stimulus on the screen. Therefore, when trained with a motor sequence, all participants can experience a contingent association between their response (coded by its location) and the outcome, i.e. the next stimulus location. This enhanced explicit learning for keyboard- and mouse-conditions. When the participants were trained with a visual sequence instead, only the ones who responded with the mouse, and thereby coded their response by the to-be-selected color, experienced a contingent response-outcome relation (i.e. the outcome being the next target's color). Hence, only the mouse-condition showed an increase in explicit knowledge about the visual sequence.

In the following three experiments, the first one represented a replication of the finding of Haider et al. (2012). In the second experiment, we aimed to fortify the assumption that participants who respond with a keyboard do not encode the visual characteristics of the target-, respectively response stimuli (or to a much lesser extent) but rather encode the relevant response locations. To test this, we only trained participants with a visual sequence this time. Again, the participants either responded with mouse or keyboard. Additionally, we introduced tones which were either bound to the colors of the response stimuli or to the response locations. Consequently, only when the tones were bound to the colors, the participants produced a melody with their responses, independent of whether they responded with the keyboard or the mouse. Should the fact that participants created a melody with their responses lead to more explicit sequence knowledge in both response device conditions, this could be taken as evidence for the assumptions put forward by Stöcker and Hoffmann (2004) or Tubau et al. (2007). If however, the contingent color-tone relation only leads to more explicit knowledge in the mouse condition, this could be taken as evidence that the mouse condition indeed encoded the colors of the response stimuli to a greater extent. Additionally, this could be taken as a further hint that experiencing a contingent relation between a response (here coded by colors) and the subsequent effects of this response (the next color and the contingent tone) are relevant for the emergence of explicit sequence knowledge. Finally, the third experiment should provide more evidence for our proposal that experiencing contingent R-E relations specifically impacts explicit sequence learning. In order to test this, participants again were trained with a visual sequence and responded either with the mouse or with the keyboard. This time, no additional tones were used. Additionally, 50% random trials were interspersed in the training sequence. This should hinder the acquisition of explicit knowledge. Two control groups received the same response device manipulation, but were trained with 100% regular trials. In line with the former experiments, we expected to find explicit knowledge in the mouse control group with 100% regular trials. The interesting comparison to distinguish between a single-system strengthening account and an R-E account was the comparison between the two 50% random trials experimental groups. Due to the random trials inhibiting the acquisition of explicit knowledge, the relevant comparison was in the acquired sequence knowledge per se (i.e. percent correct responses in the wager task).

If we found more sequence knowledge per se for the experimental group trained with 50% random material while using the mouse as a response device compared to the keyboard experimental condition, this would speak for a single-system strengthening account. It could be interpreted that the manipulation of the response device affects attention to the relevant dimension (i.e. stimulus color), so that the mouse-condition acquires stronger associations and therefore also can develop more explicit knowledge. If however, both conditions that were trained with 50% random material showed a comparable amount of knowledge, this would make a point against an interpretation based on the gradual increase of associative strength. We hypothesized that the difference in sequence knowledge between response devices would be specific to the acquisition of explicit knowledge, not associative strength per se. If R-E learning specifically affects the acquisition of explicit knowledge, there should not be a difference between the two experimental conditions because the 50% percent random trials impede these explicit learning processes.

The second study brought important further evidence for a special role of R-E learning in the emergence of explicit sequence knowledge. Experiment 1 replicated the findings of Haider et al. (2012) and Experiment 2 fortified the R-E learning explanation for these findings. More explicit knowledge was found when action-effect tones were bound contingently to the participant's responses coded in the relevant dimension. Experiment 3 showed that action-effects seemed to selectively enhance explicit but not implicit learning processes.

5.3 Study 3: Action-Effects Enhance Explicit Sequential Learning

Study 3 served to further explore the relationship between R-E learning and explicit sequence knowledge by ruling out some of the alternative explanations left open by study 2. Concerning the results of study 2, it might still be possible to explain them with a strengthening account even if Experiment 3 provided results which at least would require some additional assumptions. If participants trained with a visual sequence either coded their actions by location (in the keyboard condition) or by location and color (in the mouse condition), only the latter group experienced a contingent relation between the color they responded to and the next target color. This can be interpreted as a simple strengthening of S-S associations, without the need to assume that participants interpreted the next color or the tone as an effect of their actions. Additionally adding a tone to an S-S sequence adds another predictive dimension to that S-S relation so that accordingly, the quality of the representation can improve even further. In Experiment 3 of Study 2, the mouse condition which received 50% random trials did not show more significant explicit knowledge. It could be argued that the added noise from the random trials made learning so difficult that both the keyboard condition and the mouse condition only learned very little about the visual sequence at all and that it would have taken many more trials until the knowledge advantage for the mouse condition could have developed. This view might be supported by the fact that while none of the two experimental conditions showed an advantage in explicit knowledge, there was a numerical trend for the mouse condition to give more percent correct responses in the wager task. If, with more training, this trend became a significant difference, it would oppose the assumption that experiencing contingent R-E relations specifically impacts explicit sequence learning but not sequence learning per se.

To rule out these explanations with Study 3, we developed an experiment with a related but different design which manipulated (a) the contingency of the responses and the effects and (b) the interpretation of the additional effect-tones as either being an additional but response-independent stimulus event or as an effect of one's own actions. The most obvious difference was that the comparison between two response devices was no longer the manipulation of choice.

Concerning the contingency of the responses and their effects, the experiment in Study 3 created a situation where all participants responded with the keyboard to a motor sequence. There was an additional and uncorrelated color sequence in the stimuli which was irrelevant for all groups. Hence, differently to Study 2, there was no difference in the dimensions the participants were induced to pay attention to. In one condition of Study 3, participants had a tone contingently bound to each response location, so that they produced a melody by responding to the sequence (Contingent-Tone Group; CT-Group). In another condition, participants had the exact same training but for them, the tones were bound to the uncorrelated and irrelevant color sequence (Non-Contingent-Tone Group, NCT-Group). This led to the effect that both conditions had the exact same response sequence and heard the same melody but only in the CT-Group, the effect-tones were contingent to the key-presses. This is basically a replication of Experiment 2 in Study 2, with a slightly different manipulation of the action-effect contingency. In Study 2 the manipulation of the response device affected the internal representation of the task and thus, depending on the induced task-set, the tones were either bound to the sequential dimension (color) or to another, irrelevant dimension (location). This might have led to a somewhat greater variance since participants in Study 2 could vary in their internal weighting of the different response dimensions. The manipulation in Study 3 provided a clearer manipulation of the action-effect contingency because this time, the contingency did not depend on the internal coding of the task but instead was determined by the task itself. This difference between Study 2 and 3, in combination with the switch from a visual to a motor sequence, might help to provide an even clearer data picture.

An even more important novelty of Study 3 is that we manipulated the interpretation of the action effects as being response-independent stimulus events or as action-dependent effects. In a third condition participants had the exact same training as the other two groups. Also, this third condition experienced the same response-tone contingency as the CT-Group. The only difference was that this time, the tone did not immediately follow the response, as it was the case for the CT-Group. Instead, it followed with a 400 ms delay to the response (Stimulus-Tone-Group, ST-Group). The temporal delay between an action and a following stimulus has repeatedly shown to be an important factor in perceiving an event as an action-effect or as an action-independent stimulus (Blakemore et al., 2002; Elsner & Hommel, 2004). Comparing the ST- to the CT-Group helps ruling out the possible alternative explanation for the results of Study 2; namely that the manipulation of the response device led to more attention on the predictive dimensions which in turn led to greater associative strength, explaining the differences in explicit knowledge. Here, both conditions contained the exact same response-tone contingencies between the predictive elements of the sequence and there was no attentional difference between both conditions. The only difference was that the CT-Group was more likely to interpret the tones as consequences of their own actions, while the ST-Group was more likely to perceive these tones as action-independent stimulus events. Hence, showing that only the CT-Group expresses more explicit knowledge in a subsequent wager task, while the ST-Group does not, would be difficult to reconcile with a simple single-system strengthening account. The results might still be explained with a hybrid between single- and multiple-system accounts. For example, according to the model of Keele et al. (2003), the results could be explained by constituting that R-E learning is a form of multi-dimensional learning with explicit knowledge gradually developing. According to this explanation, the tones would enhance learning in the CT-Condition by constituting a correlated sequence to the visual sequence. To deal with this explanation we also included a forth control group, who had the same training task but without any tones. According to the model of Keele et al., multidimensional learning is perturbed when a sequence is interspersed with unpredictive elements. If indeed a multi-dimensional learning process was responsible for the CT-Condition showing an increase in explicit knowledge, then the same multi-dimensional learning mechanism should be perturbed in the NCT-Condition by the non-contingent tones. In this case the NCT-Condition should show less knowledge than the control condition which did not hear any tones.

Study 3 revealed that only participants who perceived the tones as contingent effects of their own actions (CT-Group) showed an increase in explicit knowledge. Bayesian statistics showed that all other conditions (NCT-, ST-, and No-tone Control-Group) did not show any difference in their explicit knowledge. Together, the results of Study 2 and 3 are able to provide converging evidence that R-E learning plays an important role for the emergence of explicit sequence knowledge. Even though these studies do not directly contain any test on whether learning about the R-E relations indeed leads to unexpected events, the UEH seems to be a good explanatory framework for the data.

Together the studies are not conforming to any straight-forward predictions any single-system theory would make. Especially the comparison of the CT- and the ST-Group in Study 3, who were trained with the same sequential material, make an argumentation based on associative strength difficult. Instead, they suggest that contingently producing the same effect leads to different salient cues which can indicate that there is a meaningful relation between one’s actions and stimuli in the environment, which in turn can trigger search processes for the underlying contingencies resulting in the acquisition of explicit knowledge.

6 Conclusion

Our brains are able to pick up an enormous amount of information in parallel and adapt to its structure. In many cases, these learned statistical relations are either too complex or support fundamental cognitive functions (e.g. increasing neuronal sensitivity towards certain objects we encounter often) and therefore do not need to be represented consciously. Following a functional view on consciousness, we need conscious processing to control information in a highly flexible, strategically adaptive way. Once information is available to the global workspace network, it can be accessed by any specialized network so that any mental operation is enabled for this globally broadcasted information. A new, consciously represented rule can be applied without, or with hardly any, previous practice. It can be transferred to a new situation or be inhibited. It can be evaluated whether we have learned similar rules before and whether it would be reasonable to replace old knowledge with the new rule and never use the old rule again. We can share our new knowledge with any person verbally if we want to. So obviously, becoming consciously aware of the fact that something has been learned comes with various behavioral advantages. The important question is how can we know that we know something, when we have never consciously represented this knowledge before?

The GWT and the HOTT both provide a starting point for the question how an unconscious representation can become a conscious one. Moreover, both theories together might compensate for each other's weaknesses. The GWT provides a very strong framework for conceptualizing implicit and explicit representations via their encapsulation, respectively their accessibility. Its asset compared to the HOTT is that it is much clearer about the mechanism how and why an unconscious representation is selected over any other competing unconscious representation. The selection for access to the global workspace is realized via a variation-selection mechanism. At the bottom-up level, each unconscious representation can be characterized by a certain activational strength based on factors like salience or associative strength. All of these representations constantly compete for access to the global workspace and the strongest representation will win this competition. However, the activational strength of each representation is not based on these bottom-up factors alone. On the top-down level there always is a certain conscious state that represents the goal states in any given situation. These goal states provide a fitness function for the competing bottom-up signals. An unconscious representation that has a rather weak bottom-up signal strength can gain significantly in strength if it fits well into the current situational requirements.

The HOTT instead provides no indication of how the problem of the selection of multiple possibly relevant and competing representations gets solved. It remains completely open when which representation will be conscious in a given situation. Rather, it is concerned with the question how a learning mechanism might be realized which is able to build meta-knowledge. In the GWT, these meta-cognitive processes are one of many different functions the information in the global workspace has access to and are not payed special attention to. In any HOTT, these processes are the core functions of consciousness. With respect to the question how knowledge of which we do not know that we possess it can become conscious, HOTTs play a significant role. It seems very unlikely that an unconsciously acquired representation could gain so much bottom-up strength on its own that a breakthrough into the global workspace, regardless of any current conscious content, could happen. Even the first and more simplistic draft of Cleeremans' and Jiménez' (2002) single-system theory pronounces that some top-down process has to be involved which puts the unconscious representation into the focus of attention. All important theories on the transition from implicit to explicit knowledge assume that metacognition plays a crucial role in evoking a conscious state that allows unconscious knowledge to be detected (Cleeremans, 2008, 2011; Dienes & Perner, 1999; Frensch et al., 2003).Three different conceptions can be characterized which aim to explain how conscious sequence knowledge results from implicit learning via meta-cognitive processes.

The first fits the suggestions from Cleeremans and Jiménez (2002). What is needed is a fitness-function provided by the current conscious workspace, which can enhance the activational strength of a certain unconscious representation enough to grant it access to the global workspace. Even though Cleeremans and Jiménez did not specify how this top-down process could be initiated, metacognition like, for example, a feeling of fluency could play an important part here. To put it simple, a person who has acquired unconscious knowledge somehow needs a conscious state that makes them ask themselves “Why do I feel like that?”. This seems similar to the assumptions of the UEH. However, opposed to the UEH, the idea here is not that a new explicit learning process acquired a new explicit representation but instead, the formerly encapsulated knowledge is “loaded” into the workspace. Now the implicit representation is connected to and accessed by a broad amount of subsystems, like, most important to the paradigm, speech modules. There is however one fundamental problem with this account. Different to unconsciously processed stimuli in the environment which can be attended to be processed consciously, encapsulated implicit knowledge is conceptualized as internal prediction-weights influencing the expectation of the next stimulus or next response. There is nothing about these associative weights per se that can become conscious; there is no corresponding distal event to that knowledge. As Cleeremans (2011) later put it himself, the implicitly acquired sequence knowledge is knowledge in the system, but not for the system.

Therefore, the second assumption, which incorporates meta-cognitive processes explicitly also stems from Cleeremans (2008, 2011, 2014). This time the problem of what could be the conscious content of a formerly implicit representation that now is accessed by various subsystems, is circumvented. Instead, a new meta-cognitive representation is created on its own. The result of this learning process is knowledge *for* the system. The cognitive system learns about its own internal states and how these relate to events in the environment. This conception has proven to be viable in replicating human data of metacognitive judgements via computational models. Yet, it remains that the assumption here is a gradual transition from unconscious to conscious knowledge. Further, it remains open how the moment a person reports insight into the sequence is determined. Moreover it remains unclear at which point other subsystems gain access to this meta-knowledge. Nevertheless, the account from Cleeremans and colleagues surely is an interesting, important and testable assumption.

The third and last conception is the UEH. It is proposed that higher-order learning processes lead to expectations of the subjective experience in a particular situation, based on experiences in similar situations in the past. A mismatch between the expectation of one's own experience and the actual experience will trigger the need for an explanation. When someone participates in an SRTT, they might become very fast over the course of training, however this might still match the subjectively perceivable expectation of mere practice. Yet, suddenly encountering a random block with a higher amount of errors made and moreover the sudden feeling of becoming slower, or the responses feeling less fluent, is a strong unexpected event in one's own experience. Contrary to the first of the three suggestions here, this unexpected event does not allow the first-order implicit representation to become conscious. Instead, and more similar to the second account, a new explicit representation has to be created. What the unexpected event does is to direct attention to an erroneous model of the expected experiences and its apparent need for an update that re-establishes congruency between expectations and actual experiences. A decisive difference to the model of Cleeremans (2008) is that the explicit learning process, which is based on hypothesis testing and not on associative strength, leads to a sudden insight rather than to a gradual development. Moreover, the UEH is able to explain why the sequence knowledge will remain implicit whenever another more plausible explanation for the violation of expectancy is found (Haider & Frensch, 2005).

It was therefore the aim of the current studies to test the relevance of representational strength, as it is pronounced by the account from Cleeremans (2008), against the assumptions of the UEH (Frensch et al., 2003). All three studies presented here were aimed at finding a method to balance the representational quality via keeping the associative strength equal, while at the same time manipulating the likelihood of experiencing an unexpected event.

Taken together, all three studies provided new and converging evidence for the assumptions of the UEH. First and foremost it has been demonstrated that the emergence of explicit sequence knowledge is related to the opportunity to perceive unexpected metacognitive feelings (i.e. an unexpected difference in the feeling of fluency, Experiment 3 of Study 1) or unexpected effects produced by one's own actions (Experiment 2 of Study 2, and Study 3). The use of different manipulations for the likelihood to experience an unexpected event is crucial for investigating the UEH. In the UEH, any event with the potential to make one ask why a mismatch between one's expectation and actual behavior occurred can trigger explicit learning processes. Moreover, a broad variety of manipulations for the occurrence of unexpected events of course also helps ruling out alternative explanations of possible side effects each single manipulation brings.

A second major point made by all three studies is that the manipulations of the internal change of metacognitive experience as well as the externally observable change in produced action-effects did not seem to affect implicit learning processes but instead affected an additional explicit learning process. This was most directly tested in Experiment 3 of Study 2, where the 50% randomly inserted irregular trials led to a disappearance of the effect of the response-device manipulation which has been shown to lead to more explicit learning in Experiment 1 and 2. Moreover, implicit learning was not affected by the insertion of irregular trials and there was no advantage in implicit learning for the group that still experienced a contingent action-effect relation in at least 50% percent of the trials. Further supporting evidence for a selective effect on explicit learning by the different manipulations also can be found in Experiment 1 of Study 1 and in Study 3. In Experiment 1 of Study 1, the arrangement of regular and irregular trials did not affect the amount of learning as measured by the wager task. In Study 3 we exploratively removed all participants with entirely explicit knowledge and also no longer found a gradual advantage in knowledge for the group that had experienced contingent action-effect relations.

Lastly, a third very important point of all presented studies is that the experiments provided different designs which aimed to carefully match the associative strength across the different conditions. This equalized associative strength can help with two problems in implicit learning research. First, from a methodological view, our different methods for balancing associative strength are important for meeting the justified objections from researchers like Lau (20008b). Lau demanded manipulations of the subjective judgements while matching signal strength in order to investigate unconscious processes without performance biases. It can be shown that the implicitly built knowledge base (a) is already a strong representation on its own (d' > 0) and (b) is comparable across the different groups. Especially Experiment 1 of Study 1 is suited for demonstrating that the manipulation of the subjective experience of the task did not affect the implicit knowledge base. Experiment 2 of Study 1 showed that the subjective experience can differ despite the comparable implicit knowledge base. Second, and even more important in the context of the goals of all three studies, is the theoretical advance this matching of associative strength brings. Due to the differences in explicit knowledge, despite equalized representational strength between conditions, a simple strengthening account (Cleeremans & Jiménez, 2002) or a more complex higher-order strengthening account (Cleeremans, 2008) has difficulties to explain these results. There is no explicit assumption in any of these single-system accounts that can explain why there are differences in explicit knowledge between the groups, or why implicit learning did not seem to be affected by any of our manipulations.

Rather, an account like the UEH which proposes an additional explicit learning process that is triggered by observable changes in one's own behavior seems favorable.

However, the results presented here are not able to completely rule out an explanation based on strengthening mechanisms, respectively not all assumptions of the UEH have been tested. A next step would be to show that the manipulations of subjective experiences like the experienced differences in fluency or distal action-effects actually constitute an unexpected event for the participants. This of course could be done by detailed interviews of the participants in their subjective perception of the training. While such interviews might surely be an important first step for investigating the subjective feeling of surprise, it is not the most direct test. It only provides data about the retrospective reconstructions of feelings of fluency during the training. An online measure of surprise during training might be even more insightful. This could be achieved by using EEG or fMRT data. Butterfield and Mangles (2003) for example have shown that the fronto-central P3a amplitude shows increased positivity when a metacognitive mismatch, i.e. a conflict between metacognitive expectations (for example the expectation to produce very few errors) and the actual outcome, occurred. Furthermore, the anterior cingulate cortex, the medial frontal gyrus and the cingulate cortex have been associated with metacognitive mismatch in an fMRT study by Metcalfe, Butterfield, Habeck, and Stern (2012). It would thus be interesting to investigate whether the different arrangement of regular and irregular stimuli leads to a difference in the activity in these areas. It would probably be best to use the specific design of Experiment 1 and 2 of Study 1. Here, the rather short mini-blocks of regular and irregular trials did not lead to explicit knowledge despite affecting the subjectively perceivable changes of fluency. This way it could be circumvented that explicit sequence knowledge contaminates the data. Once it could be shown that the manipulation of the fluency can be associated with a measure for metacognitive mismatches, the design could be extended in order to show that the moment a mismatch is detected stands in close relation to detecting the sequence. This could be achieved by, for example, analyzing the data for sudden RT-drops as suggested by Haider and Rose (2007; Rose et al., 2010). Moreover, neuroimaging data like the sudden coupling of gamma-band activity, respectively increases of the BOLD-signal in the ventrolateral prefrontal cortex, the medial prefrontal cortex and the ventral striatum could be analyzed (Rose et al., 2010; Schuck et al., 2015; Wessel et al., 2012). This would provide a necessary test of the prediction of the UEH that unexpected events lead to a sudden onset of hypothesis testing and subsequent detection of the sequence, instead of a slowly, gradually developing explicit representation as assumed by Cleeremans (2008).

In order to further exclude a single-system explanation, it would be important to show in a more direct way that unexpected events lead to the development of a new representation via hypothesis testing, instead of somehow transforming the implicit representation into an explicit one. To achieve this, again the design of Experiment 1 and 2 of Study 1 could be used to train participants until they start to perceive differences in their feeling of fluency when trained with the blocked material and then presenting them a new, unknown sequence. Showing more explicit knowledge for the new sequence when participants were trained with the blocked arrangement of the material, than when trained with the randomly arranged material, would make a strong point that the sequence is explicitly searched and learned instead of being derived from the implicit knowledge base.

Taken together, the studies provide further, converging evidence that the experience of unexpected events lead to the emergence of explicit knowledge in an implicit learning situation. Such unexpected events can be characterized as a metacognitive mismatch between the expectancy of internal states as well as the expected observable effects of one's own behavior and the actually experienced internal state, respectively behavioral effects. The current studies should encourage further investigations which focus on (a) the close relation between the occurrence of an unexpected event and a sudden insight into the hidden regularity and (b) the assumption that explicit learning processes lead to a new representation which is independent of the implicitly learned one.

The introduced designs of all three studies provide three different, practicable and powerful manipulations of the occurrence of unexpected events while matching associative strength. They can also be flexibly adjusted to different research questions investigating the relation and the differences between unconscious and conscious processing. Designs like the presented ones can, for example, be of value for further development in the debate about the source signals for metacognitive learning processes. Increasing the options to manipulate or balance first-order signal strength while independently varying metacognitive judgements can help to distinguish between models that are based on a single channel bottom-up relation (e.g. Galvin, Podd, Drga, & Whitmore, 2003), hierarchical models (e.g. Maniscalco & Lau, 2016) where different processes underlie objective and subjective judgements, but the latter always evaluates the judgmental quality of the former, or dual channel models, where objective and subjective judgements are based on fully independent cognitive processes (e.g. Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009). Understanding the fundament of metacognitive learning processes is highly important to several practical areas. This includes educational contexts, where the ability to accurately judge one's own knowledge base is crucial to self-regulated learning (Karpicke, 2009), research on artificial intelligence, where metacognitive expectancy deviations are a key variable to perturbation-tolerant behavior (Anderson & Perils, 2005), or mental illnesses like pathological gambling (Brevers et al., 2014), schizophrenia (Moritz et al., 2016) or autism (Williams, Bergström, & Grainger, 2016).

Literature

Abrahamse, E. L., Jiménez, L., Verwey, W. B., & Clegg, B. A. (2010). Representing serial action and perception. *Psychonomical Bulletin and Review, 15*(5), 603-623. doi: 10.3758/PBR.17.5.603

Alamia, A., Orban de Xivry, J. J., San Anton, E., Olivier, E., Cleeremans, A., & Zenon, A. (2016). Unconscious associative learning with conscious cues. *Neuroscience of Consciousness, 1*(1). doi:10.1093/nc/niw016

Anderson, M. L., & Perils, D. (2005). Logic, self-awareness, and self-improvement: The metacognitive loop and the problem of brittleness. *Journal of Logic and Computation, 15*(1), 21-40. doi:10.1093/logcom/ex034

Atas, A., Vermeiren, A., & Cleeremans, A. (2013). Repeating a strongly masked stimulus increases priming and awareness. *Consciousness & Cognition, 22*(4), 1422-1430. doi:10.1016/j.concog.2013.09.011

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. (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

Balleine, B. W., & O'Doherty, J. (2010). Human and rodents homologies in action-control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology Reviews*, *35*(1)*,* 48-69. doi:10.1038/npp.2009.131

Barth, M., Stahl, C., & Haider, H. (2016). *Assumptions of the process-dissociation procedure are violated in sequence learning*. Manuscript submitted for publication.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7-15. doi:10.1016/0010-0277(94)90018-3

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalisable knowledge. *Quarterly Journal of Experimental Psychology, 36*(2), 209-231. doi:10.1080/14640748408402156

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalties in the awareness of action. *Trends in Cognitive Science, 6*(6), 237-242. doi:10.1098/rstb.2000.0734

Block, N. (1995). On a confusion about the function of consciousness*. Behavioral and Brain Sciences, 18*(2), 227-287. doi: 10.1017/S0140525X00038188

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences, 9*(2), 46-52. doi:10.1016/j.tics.2004.12.006

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*(5-6), 481-499. doi:10.1017/S0140525X07002786

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience, 7*(4), 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*(3), 642-652. doi:10.1037//0033-295X.108.3.624

Brevers, D., Cleeremans, A., Bechara, A., Greisen, M., Kornreich, C., Verbanck, P., & Noël, X. (2014). Impaired metacognitive capacities in individuals with problem gambling. *Journal of Gambling Studies, 30*(1), 141-152. doi:10.1007/s10899-012-9348-3

Bucher, A., Steffens, M. C., Erdfelder, E., & Rothkegel, R. (1997). A multinominal model to assess fluency and recollection in a sequence learning task. *Quarterly Journal of Experimental Psychology, 50*(3), 631-663. doi:10.1080/713755723

Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research, 17*(3), 793-817. doi:10.1016/S0926-6410(03)00203-9

Chalmers, D. J. (1995). Facing up to the problems of consciousness. *Journal of Consciousness Studies, 2*(3), 200-219. doi:10.1093/acprof:oso/9780195311105.003.0001

Changeux, J. P., & Dehaene, S. (1989). Neuronal models of cognitive functions. *Cognition, 33*(1-2), 63-109.

Chase, H. W., Kumar, P., Eickhoff, S. B., & Dombrovski, A. Y. (2015). Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cognitive, Affective & Behavioral Neuroscience, 15*(2), 435-459. doi:10.3758/s13415-015-0338-7

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics, 36*(4), 387-395.

Clark, A. & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language, 8*(4), 487-519. doi:10.1111/j.1468-0017.1993.tb00299.x

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*. doi:10.3389/fpsyg.2011.00086

Cleeremans, A. (2014). Connecting conscious and unconscious processing. *Cognitive Science, 38*(6), 1286-1315. doi:10.1111/cogs.12149

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness: An empirical, computational and philosophical consensus in the making?* (pp. 1–40). Hove: Psychology Press.

Cleereman, A., Timmermans, B., & Pasquali, A. (2007). Consciousness and metarepresentation: A computational sketch. *Neural Networks, 20*(9), 1032-1039. doi:10.1016/j.neunet.2007.09.011

Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences, 15*(6), 358-364. doi:10.1016/j.tics.2011.06.008

Conway, M. C., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of New York Academy of Sciences, 1145*, 113-131. doi:10.1196/annals.1416.009

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2*, 263-275.

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 304-317. doi:10.1037/0278-7393.20.2.304

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200-227. doi:10.1016/j.neuron.2011.03.018

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*(5), 204-211. doi:10.1016/j.tics.2006.03.007

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

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*(9), 2531-2540. doi:10.1093/brain/awp111

Dennett, D. C. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), *Minds, Machines and Evolution* (pp. 129-151). Cambridge: Cambridge University Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little, Brown & Co.

Dennett, D. C. (2015). Why and how does consciousness seem the way it seems? *Open MIND, 10*. doi:10.15502/9783958570245

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review, 8*(2), 343-350. doi:10.3758/BF03196171

Destrebecqz, A., & Cleeremans, A. (2003). Temporal effects in sequence learning. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 181–213). Amsterdam: John Benjamins Publishing Company.

Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Del Fiore, G., Aerts, J., Luxen, A., van der Linden, M., Cleeremans, A., & Maquet, P. (2003). Cerebral correlates of explicit sequence learning. *Cognitive Brain Research, 16*(3), 391-398. doi:10.1016/S0926-6410(03)00053-3

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4*(1), 3-23. doi:10.3758/BF03210769.

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

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

Eberhardt, K., Esser, S., & Haider, H. (2017). Abstract feature codes: The building blocks of the implicit learning system. *Journal of Experimental Psychology: Human Perception and Performance, 43*(7), 1275-1290. doi:10.1037/xhp0000380

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 229-240. doi:10.1037//0096-1523.27.1.229

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research, 68*(2-3), 138-154. doi:10.1007/s00426-003-0151-8

Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review, 67*(5), 279-300, doi:10.1037/h0041622

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review, 124*(1), 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

Fodor, J. A. (1975). *The language of thought*. New York, NY: Harper & Row.

Frensch, P. A., Haider, H., Rünger, D., Neugebauer, U., Voigt, S., & Werg, D. (2003). The route from implicit learning to awareness of what has been learned. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 335-366). New York, NY: John Benjamins Publishing Company.

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*(4), 843-876.

Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience, 7*(4), 497-510. doi:10.1162/jocn.1995.7.4.497

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron, 31*(5), 681-697. doi:10.1016/S0896-6273(01)00424-X

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology, 22*(10), 390-392. doi:10.1016/j.cub.2012.02.040

Haider, H., Eberhardt, K., Kunde, A., & Rose, M. (2012). Implicit visual learning and the expression of learning. *Consciousness and Cognition, 22*(1), 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(3), 658-672. doi:10.1016/j.concog.2010.10.021

Haider, H., & Frensch, P. A. (2005). The generation of conscious awareness in an incidental learning situation. Psychological Research, 69(5-6), 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*(6), 817–834. doi:10.1007/s00426-008-0199-6

Haider, H., & Rose, M. (2007). How to investigate insight: A proposal. *Methods, 42*(1), 49-75. doi:doi.org/10.1016/j.ymeth.2006.12.004

Hazeltine, E., Grafton, S. T., & Ivry, R. (1997). Attention and stimulus characteristics determine the locus of motor-sequence encoding. A PET study. *Brain, 120*(1), 123-140.

Hélie, S., Proulx, R., & Lefebvre, B. (2011). Bottom-up learning of explicit knowledge using a Bayesian algorithm and a new Hebbian learning rule. *Neural Networks, 24*(3), 219-232. doi:10.1016/j.neunet.2010.12.002

Herwig, A., & Waszak, F. (2009). Intention and attention in ideomotor learning. *The Quarterly Journal of Experimental Psychology, 62*(2), 219-227. doi:10.1080/17470210802373290

Herwig, A., & Waszak, F. (2012). Action-effect bindings and ideomotor learning in intention- and stimulus-based actions. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00444

Hoffmann, J., Sebald, A., & Stöcker, C. (2001). Irrelevant response effects improve serial learning in serial reaction time tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(2)*, 470-482. doi:10.1037/0278-7393.27.2.470

Hommel, B. (2017). Conscious and unconscious control of spatial action*. Reference Module in Neuroscience and Biobehavioral Psychology*. doi:10.1016./B978-0-12-809324-5.05929-0

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral & Brain Sciences, 24*(5), 849-878. doi:10.1007/s00426-009-0234-2

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513-541. doi:10.1016/0749-596X(91)90025-F

Jiménez, L., Mendez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(4), 948-969. doi:10.1037/0278-7393.22.4.948

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*(4), 469-486. doi:10.1037/a0017341

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain & Language, 127*(1), 46-54. doi:10.1016/j.bandl.2012.11.007

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110*(2), 352-339. doi:10.1037/0033-295X.110.2.316

Kilgard, M. P., & Merzenich, M. M. (2002). Order-sensitive plasticity in adult primary auditory cortex. *PNAS, 99*(5), 3205-3209. doi:10.1073/pnas.261705198

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.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*(2), 149-171. doi:10.1006/ccog.2000.0433

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119*(1), 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). New York, NY: Psychology Press.

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 857-875. doi:10.1098/rstb.2007.2093

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*(1), 12-18. doi:10.1016/S1364-6613(02)00013-X

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences, 10*(11), 494-501. doi:10.1016/j.tics.2006.09.001

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience, 1*(3), 204-240. doi:10.1080/17588921003731586

Lau, H. C. (2008a). 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. (2008b). Are we studying consciousness yet? In L. Weiskrantz & M. Davies (Eds.*), Frontiers of Consciousness* (pp. 2008-2245). Oxford: Oxford University Press.

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *PNAS, 103*(49), 18763-18768. doi:10.1073/pnas.0607716103

Lau, H. C., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences, 15*(8), 365-373. doi:10.1016/j.tics.2011.05.009

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

Massoni, S., Gajdos, T., & Vergnaud, J. C. (2014). Confidence measurement in the light of signal detection theory. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.01455

Metcalfe, J., Butterfield, B., Habeck, C., & Stern, Y. (2012). Neural correlates of people‘s hypercorrectionof their false beliefs. *Journal of Cognitive Neuroscience, 24*(7), 1571-1583. doi:10.1162/jocn_a_00228

Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology, 56*(6), 485-491. doi:10.1037/h0044933

Moore, J. W., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and Cognition, 17*(1), 136-144. doi:10.1016/j.concog.2006.12.004

Moritz, S., Balzan, R. P., Bohn, F., Veckenstedt, R., Kolbeck, K., Bierbrodt, J., & Dietrichkeit, M. (2016). Subjective versus objective cognition: Evidence for poor metacognitive monitoring in schizophrenia. *Schizophrenia Research, 178*(1-3), 74-79. doi:10.1016/j.schres.2016.08.021

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall

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

Nieuwenhuis, S., & de Kleijn, R. (2011). Consciousness of targets during the attentional blink: a gradual or all-or-none dimension? *Attention, Perception, & Psychophysics, 73*(2), 364-373. doi:10.3758/s13414-010-0026-1

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

Overgaard, M., Rote, J., Mouridsen, K., & Ramsøy, T. Z. (2006). Is conscious perception gradual or dichotomous? A comparision of repot methodologies during a visual task. *Consciousness and Cognition, 15*(4), 700-708. doi:10.1016/j.concog.2006.04.002

Pasquali, A., Timmermans, B., & Cleeremans, A. (2010). Know thyself: Metacognitive networks and measures of consciousness. *Cognition, 117*(2), 182-190. doi:10.1016/j.cognition.2010.08.010

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.

Perruchet, P. & Amorim, A. (1992). Conscious knowledge and changes in performance in sequence learning: evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 785-800. doi:10.1037//0278-7393.18.4.785

Perruchet, P. & Vinter, A. (2002). The self-organizing consciousness. *Behavioral and Brain Sciences, 25*(3), 297-388. doi:10.1017/S0140525X02000067

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

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414*(6863), 546-550. doi:10.1038/35107080

Reber, A. S. (1965). Probability learning and memory for event sequences. *Psychonomic Science, 3*, 431-432.

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

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219-235.

Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia, 51*(10), 2026-2042. doi:10.1016/j.neuropsychologia.2013.06.019

Reber, P. J., & Squire, L. R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience, 10*(2), 248-263. doi:10.1162/089892998562681

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 44*(6), 563-757. doi:10.3758/BF03207490

Rose, M., Haider, H., & Büchel, C. (2005). Unconscious detection of implicit expectancies. *Journal of Cognitive Neuroscience, 17*(6), 918-927. doi:10.1162/0898929054021193

Rose, M., Haider, H., & Büchel, C. (2010). The emergence of explicit memory during learning. *Cerebral Cortex, 20*(12), 2787-2797. doi:10.1093/cercor/bhq02

Rose, M., Haider, H., Salari, N., & Büchel, C. (2011). Functional dissociation of hippocampal mechanism during implicit learning based on the domain of associations. *Journal of Neuroscience, 31* (39), 13739-13745. doi:10.1523/jneurosci.3020-11.2011

Rose, M., Haider, H., Weiller, C., & Büchel, C. (2002). The role of medial temporal lobe structures in implicit learning: An event-related fMRI study. *Neuron, 36* (6), 1221-1231. doi:10.1016/S0896-6273(02)01105-4

Rosenthal, D. (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. (2008). Consciousness and its function. *Neuropsychologia, 46*(3), 829-840. doi:10.1016/j.neuropsychologia.2007.11.012

Rünger, D. (2012). How sequence learning creates explicit knowledge: The role of response-stimulus interval. *Psychological Research, 76*(5), 579-590. doi:10.1007/s00426-011-0367-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*(5), 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*(2), 121-137. doi:10.1007/s00426-008-0225-8

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition, 19*(4), 1069-1078. doi:10.1016/j.concog.2009.12.013.

Schuck, N. W., Gaschler, R., Wenke, D., Heinzle, J., Frensch, A., Haynes, J.-D., & Reverberi, C. (2015). Medial prefrontal cortex predicts internally driven strategy shifts*. Neuron, 86*(1), 331-340. doi:10.1016/j.neuron.2015.03.015

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

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews, 32*(2), 265-278. doi:10.1016/j.neubiorev.2007.07.010

Shanahan, M., & Baars, B. (2005). Applying the global workspace theory to the frame problem. *Cognition, 98*(2), 157-176. doi:10.1016/j.cognition.2004.11.007

Shanks, D. R. (2005). Implicit Learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202-220). London: Sage.

Shanks, D. R. (2016). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomical Bulletin & Review, 24*(3), 752-775. doi:10.3758/s13423-016-1170-y

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*(6), 1435-14551. doi:10.1037//0278-7393.25.6.1435tomaschke ideomotor

Shanks, D. R., & Perruchet, P. (2002). Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomical Bulletin Review, 9*(2), 362-367.

Shanks, D. R., & St John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*(3), 367-447. doi:10.1017/S0140525X00035032

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin, 136*(6), 943-974. doi:10.1037/a0020541

Speekenbrink, M., Channon, S., & Shanks, D. R. (2008). Learning strategies in amnesia. *Neuroscience & Biobehavioral Reviews, 32*(2), 292-310. doi:10.1016/j.neubiorev.2007.07.005

Stöcker, C., & Hoffmann, J. (2004). The ideomotor principle and motor sequence acquisition: Tone effects facilitate movement chunking. *Psychological Research, 68*(2-3), 126-137. doi:10.1007/s00426-003-0150-9

Stöcker, C., Sebald, A., & Hoffmann, J. (2003). The influence of response-effect compatibility in a serial reaction time task. *The Quarterly Journal of Experimental Psychology, 56*(4), 685-703. doi:10.1080/02724980244000585

Stoering, P., Zontanou, A., & Cowey, A. (2002). Aware or unaware: Assessment of cortical blindness in four men and a monkey. *Cerebral Cortex, 12*(6), 565-574. doi:10.1093/cercor/12.6.565

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach*. Psychological Review, 112*(1), 159-192. doi:10.1037/0033-295X.112.1.159

Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence.* Chicago, IL: University of Chicago Press.

Tran, R., & Pashler, H. (2017). Learning to exploit a hidden predictor in skill acquisition: Tight linkage to conscious awareness. *PLoS ONE, 12*(6). doi:10.1371/journal.pon.0179386

Tubau, E., López-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*(1), 43-63. doi:1037/0096-3445.136.1.43

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*(10), 1934-1945. doi:10.1162/jocn.2009.21131

Vandenberghe, M., Schmidt, N., Frey, P., & Cleeremans, A. (2006). Can amnestic patients learn without awareness? New evidence comparing deterministic and probabilistic sequence learning. *Neuropsychologia, 44*(10), 1629-1644. doi:10.1016/j.neuropsychologia.2006.03.022

Wessel, J., 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. W. A. (2002). Two routes to remembering (and another to remembering not*). Journal of Experimental Psychology: General, 131*(3), 325-348. doi:10.1037//0096-3445.131.3.325

Whittlesea, B. W. A. (2000). The source of feelings of familiarity: the discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 547-565. doi:10.1037/0278-7393.26.3.547

Williams, D. M., Bergström, Z., & Grainger, C. (2016). Metacognitive monitoring and the hypercorrection effect in autism and the general population: Relation to autism(-like) traits and mindreading. *Autism: International Journal of Research and Practice*. doi:10.1177/136231316680178

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105*(3), 558-584.

Windey, B., Gevers, W., & Cleeremans, A. (2013). Subjective visibility depends on level of processing. *Cognition, 129*, 404-409. Doi: 10.1016/j.cognition.2013.07.012

Windey, B., Vermeiren, A., Atas, A., & Cleeremans, A. (2014). The graded and dichotomous nature of visual awareness. *Philosophical Transactions of the Royal Society B, 369*(1641), 1-11. doi:10.1098/rstb.2013.0282

Windey, B., & Cleeremans, A. (2015). Consciousness as a graded and an all-or-none phenomenon: A conceptual analysis. *Consciousness and Cognition, 35*, 185-191. doi:10.1016/j.concog.2015.03.002

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences, 7*(5), 214-218. doi:10.1016/S1364-6613(03)00081-0

Ziessler, M. (1998). Response-effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 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 time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 595-613.

Ziori, E., & Dienes, Z. (2005). Subjective measures of unconscious knowledge of concepts. *Mind & Society, 5*(1), 105-122. doi:10.1007/s11299-006-0012-4

Zirngibl, C., & Koch, I. (2002). The impact of response mode on implicit and explicit sequence learning. *Experimental Psychology, 49*(2), 153-162. doi:10.1027//1618-3169.49.2.15

Appendix: Published Articles and Contributions of the Authors

Publication 1

Esser, S., & Haider, H. (2017). The emergence of explicit knowledge in a serial reaction time task: The role of experienced fluency and strength of representation. *Frontiers in Psychology, 8*, doi: 10.3389/fpsyg.2017.00502

Sarah Esser
Developed the hypotheses for Experiment 1, 2, and 3, designed Experiment 1, 2 and 3, programed Experiment 1, 2 and 3, analyzed the data for Experiment 1, 2, and 3, and wrote the manuscript

Hilde Haider
Helped developing the hypotheses for Experiment 1, 2, and 3, helped with the design of Experiment 1, 2 and 3, and revised the manuscript

Publication 2

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

Hilde Haider
Developed the hypothesis for Experiment 1, 2, and 3, designed Experiment 1, 2 and 3, programed Experiment 1, 2, and 3, analyzed the data for Experiment 1, 2, and 3, wrote the manuscript

Katharina Eberhardt
Helped developing the hypothesis for Experiment 1, 2, and 3, helped with the design of Experiment 1, 2 and 3, revised the manuscript

Sarah Esser
Helped developing the hypothesis for Experiment 1, 2, and 3, helped with the design of Experiment 1, 2 and 3, and revised the manuscript

Michael Rose

Helped developing the hypothesis for Experiment 1, 2, and 3 and revised the manuscript

Publication 3

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

Sarah Esser
Developed the hypotheses for the experiment, designed the experiment, programed the experiment, analyzed the data for the experiment, and wrote the manuscript

Hilde Haider

Helped developing the hypotheses for the experiment, helped with the design of the experiment, and revised the manuscript