Budapest University of Technology and Economics

Department of Cognitive Science
Psychology PhD School

Ferenc Kemény

Stimulus dependence and developmental differences in Probabilistic Categorization and Sequence Learning

PhD Thesis

Supervisor:

Ágnes Lukács, PhD

Budapest, 2012

Table of contents

List of abbreviations

- AGL: Artificial Grammar Learning task.
- FOP: First order predicative sequences. In these sequences, the elements can be predicted by the previous item.
- IC: Ice-Cream task.
- IL: Implicit Learning.
- MTL: Mediotemporal lobe.
- PD: Parkinson's disease.
- PDH: Procedural Deficit Hypothesis.
- PDP: Process Dissociation Procedure.
- RSI: Response-Stimulus-Interval.
- RT: Reaction-time.
- SL: Statistical Learning.
- SLI: Specific Language Impairment.
- SOP: Second order predicative sequences. Sequences where elements may be predicted from the previous two items.
- SRT: Serial Reaction-Time task.
- TD: Typically developing.
- TP: Transitional probability. The probability with which one element follows another in a sequence.
- WP: Weather Prediction task.

Introduction

Some of our memories are factual, well defined and easy to recall consciously, but we also all have memories that are different. Each of us has different bits of knowledge that are difficult to verbalize. We may not even be aware of some of this knowledge. The former, factual type belongs to our explicit or declarative type of knowledge; the latter bits of acquired information are implicit or procedural. These are mostly complex, process-like (non-factual) representations, and are usually categorized as skills in the psychological literature. These skills form the background for a great range of our everyday behaviour in the motor (Willingham, 1998), cognitive (Knowlton, Mangels, & Squire, 1996) and social (Lieberman, 2000) domains. Skills do not only share a number of characteristics in the underlying representations, but also in the way they are acquired. This way of acquisition is generally referred to as implicit learning. Implicit learning (IL) is the incidental acquisition of complex information, where neither the process of learning, nor the acquired information can be consciously accessed. Since skills are related to a number of events in our everyday life, the way they are acquired is central from the perspective of behaviour acquisition, behaviour planning and behaviour modification.

This type of learning is usually contrasted with explicit learning. Explicit learning refers to the conscious access of the acquired information, that is, learning may still be incidental, the representation may be recollected (Graf & Schacter, 1985). This distinction is usually treated along with the declarative-procedural distinction, where declarative representations are factual and easily verbalizable, while procedural representations are process-like and difficult to define (Squire, Knowlton, & Musen, 1993). In relation to the learning process, the former ones are rapidly acquired, while the latter are learned slowly, gradually (Henke, 2010). The declarative-procedural distinction is based on the nature of representation, and the representational differences lead to differences in conscious access. That is, declarative representations are factual, hence they are easily verbalizable, hence they can be accessed, whereas procedural representations are process-like, which is difficult to define, leading to the inability to be consciously accessed. In this case, recollection is just a side-effect of representational differences. Despite the differences (see e.g. Berry & Dienes, 1993), a major line of research, including studies of the WP, SRT and AGL tasks, identifies the two distinctions as mapping directly onto each other. Although we are not entirely convinced in the identification of the two distinctions, due to the tradition in these studies, and to the fact that the dissertation does not focus on the nature of these distinctions, we treat the notions of 'implicit' and 'procedural' as synonyms, just like the expressions 'explicit' and 'declarative'. By the end of Study 4, however, it became clear that the distinctions should be subject to research for the better understanding of representational versus access differences.

The topic of the dissertation is implicit/procedural learning. We inquire into the nature of the learning process by looking at how different stimulus types, configurations and presentation designs (stimulus dependence) affect learning and whether there are differences of learning across different populations (children versus adults, language impaired versus typically developing) in Probabilistic Categorization and Sequence Learning. Four studies are presented in the dissertation. The first study is a neuropsychological study testing children with Specific Language Impairment (Kemény & Lukács, 2010). The focus of the study is whether children with delayed linguistic abilities show decreased performance in implicit learning, using a probabilistic category learning task, the Weather Prediction (WP) task. The second and third experiments study the role of stimulus information. The central question of these studies is whether different domains of stimuli and different structures in stimuli affect learning performance on implicit learning tasks. The studies employ different implicit learning paradigms. The second study uses the WP task for testing the effect of two variables on learning: Combination and Transparency. Combination refers to the mode of stimulus presentation, whether different predictive elements appear as features of a single image, or as separate images. Transparency refers to the link between cues and outcomes: if a link between the stimulus and outcome is transparent, the association is motivated (e.g. they are in a part-whole or a causal relationship in the natural world), if it is non-transparent, there is only a statistical association established by presentation in the experiment (Kemény & Lukács, 2009). The third study uses the Serial Reaction-Time (SRT) task. The focus of the study is whether learning on the SRT task is based on perceptual (stimulus) or motor (response) information, and to identify how each component may contribute to learning (Kemény & Lukács, 2011). The fourth study focuses on the implicit versus explicit nature of the WP task: whether self-reports by healthy adults are in line with any one of the previous hypotheses considering the WP task as partially implicit or fully explicit (Kemény & Lukács, 2012).

The following introduction covers four general issues. First, we introduce and compare three traditional implicit learning tasks based on previous neuropsychological and neuroimaging results. The three traditional implicit learning tasks are the Weather Prediction task, the Serial Reaction-Time task and the Artificial Grammar Learning task (AGL). As the studies of the dissertation employ both the WP and SRT tasks, it is important to see the similarities and differences between the tasks. Neuropsychological and neuroimaging results are complemented with results and theories relating implicit/procedural learning and language. These issues are important as Study 1 tests implicit/procedural learning in Specific Language Impairment using the WP task.

The second major section of the introduction compares the existing literature on the effects caused by different stimulus sets on the traditional implicit learning paradigms. These issues involve modality and domain dependence, as well as differences between stimulus structures. As the past literature of the three tasks is concerned with diverse topics, these topics will be covered in the “Domain and modality independence in Probabilistic Categorization, Artificial Grammar Learning and the Serial Reaction-time task” section (Section 3). Studies 2 and 3 are connected to this section. In Study 2, we tested four different versions of the WP task. The question in focus was whether different stimulus-stimulus (the way stimuli are presented in stimulus combinations) and stimulus-outcome (the way a stimulus is connected to an outcome) relations affect learning differently. In Study 3 we employed the SRT task, and our question was whether learning on the SRT task takes place in the perceptual (learning sequence of stimuli) or response (learning sequences of responses) domains, and whether the different domains affect each other.

The third major section covers previous data on the development of implicit learning. Despite the fact that there are three different theories in connection with the development of implicit learning, data on the issue is sparse. These studies are usually comparing the learning performance of adults and children or younger and older adults but lack systematicity. As Studies 1 and 2 reported children’s data, these data are compared in order to have a clearer picture on how probabilistic categorization performance – measured on the WP task – changes with time in primary school age.

The focus of the fourth major section of the introduction is the implicit versus explicit nature of the three tasks. All three tasks were designed to show that clinical populations with a severe declarative deficit are able to demonstrate unconscious learning. Behavioural results show, however, that healthy participants are in fact at least partially aware of what they learn during these traditionally implicit paradigms. There are two sources of data: participants are either prompted directly about their conscious access to their knowledge, or modifications are used to selectively impair or clash conscious and unconscious functioning. Behavioural studies on the implicit versus explicit nature of the three tasks will be reviewed in Section 4.

As self-insight measures are missing from the WP literature, Study 4 adapts a subjective self-report method from AGL studies to test explicit access to the acquired knowledge in probabilistic category learning (Dienes & Scott, 2005).

1. Implicit/Procedural learning in the light of neuropsychology and neuroscience

1.1. Paradigms of implicit learning

As introduced in the above outline, the dissertation focuses on testing learning effects in two tasks: the Weather Prediction task and the Serial Reaction-Time task. The WP task is a forced-choice, non-sequential probabilistic category learning task (Knowlton, Squire, & Gluck, 1994). In the task participants face one, two or three out of four different cues, and have to predict whether there would be rain or sunshine. Participants receive no other instructions, but feedback is provided after each decision. Participants are not aware of the fact that each cue has a pre-set value. This pre-set value is the probability with which the given cue is associated with the outputs (sunshine and rain). In the early phases of the task participants' performance is comparable to chance level, while they are expected to reach as high as 80% performance by the fourth block of 50 items (Gluck, Shohamy, & Myers, 2002).

Participants may rely on different strategies in solving the task; three different types of these were identified: the Singleton strategy, the One-cue strategies and the Multi-cue strategy (Gluck et al., 2002). Participants using the Singleton strategy respond consistently if there is a single cue appearing alone. In all other cases – when there are two or three cues appearing simultaneously – their response is random. This strategy is sub-optimal: it leads to a very low, but above chance performance. One-cue strategy users focus on one of the cues, and if that specific cue is present, they respond consistently, but in the absence of that specific cue their response is random. As there are four different cues, there may be four different One-cue strategies. All four of these strategies are sub-optimal. In both the Singleton and One-cue strategies, participants only focus on one cue at the time. This similarity clusters these response patterns into the same group. The last strategy is the optimal strategy. Multi-cue strategy users use all cues for making prediction. The predictive values are combined, and the decision is based on the mean value. Note that these response patterns were identified as implicit strategies: that is they are identified using a mathematical model, and were found not to correlate with self-reports (Gluck et al., 2002)

The SRT task is a computerized sequence learning task (Nissen & Bullemer, 1987) where participants see four different locations, and a target stimulus appears in one of the locations. The goal of the participants is to press the button that corresponds to the given location as quickly as possible. No other instructions are given. Unknown to the participants, the appearance of the target stimulus is not random, but follows a usually 12 element long sequence. As time passes, participants' reaction times decrease. After a long training period, the appearance of the target stimulus becomes random; this causes an increase in reaction times. The RT decrease observed in the training blocks is usually considered a complex training effect that is caused by sequence learning as well as a general motor learning and getting used to the task in general (Howard & Howard, 1989). On the other hand, the RT increase from the sequence to the random block is most likely to be due to sequence specific knowledge (Nissen & Bullemer, 1987). The sequences used in the SRT task may be deterministic (Meulemans, Van der Linden, & Perruchet, 1998) or probabilistic (Deroost, Zeeuws, & Soetens, 2006). The distribution of pattern elements may also differ: there are studies in which each button-press is a member of the sequence (Destrebecqz & Cleeremans, 2001), or pattern elements may be separated by one (Howard & Howard, 1997) or two randomly appearing stimuli (Bennett, Howard, & Howard, 2007).

Though our studies presented below do not employ the Artificial Grammar Learning task, it is useful to introduce this paradigm too, as it is also a very frequently used IL task. And also due to task similarities: a number of hypotheses and methods developed in the AGL literature are relevant to the topic of the dissertation. Based on these similarities, in Study 4, we rely on a method originating in the AGL literature and use it in the WP task.

The AGL task is a task for learning mini grammars. It is mostly considered a typical non-motor probabilistic sequence learning task, and is mostly referred to as a statistical learning task. For a short training period, participants face different strings of letters (Reber, 1967), syllables (Saffran, Aslin, & Newport, 1996), words (Saffran, 2002), tones (Saffran, Johnson, Aslin, & Newport, 1999) or spatial configurations (Fiser & Aslin, 2002). Unknown to the participants, the different elements do not appear randomly. The sequences observed follow the rules of a finite-state grammar (Reber, 1967), or a small set of rules (Saffran, 2002). After the training task, participants are informed that there was an underlying structure in the task, and that their role in the following forced-choice test-phase would be to compare pairs of strings and decide which one of the two strings is more similar to the earlier sequences. While participants are not aware of the set of rules, their performance – due to the training – is significantly above chance.

From certain aspects, the three tasks seem to be very different: the AGL and SRT tasks focus on the acquisition of sequential information, while the WP task lacks a sequential structure. The AGL and the WP tasks focus on the extraction of statistical information, whereas in the case of the SRT task, some versions use deterministic, others use probabilistic sequences. Also, the WP and SRT tasks are similar in that both test performance online, throughout learning, while in the AGL task, performance is measured with an offline test-phase. According to our knowledge, there is only one study examining the relationship between the three tasks (Aczél & Gönci, 2005). This specific study found no correlation between performances on the three tasks. We are also currently working on a lifespan developmental study using all three of the task on a sample of 480 participants in 10 different age-groups. Yet unpublished results of our study also show no correlation between the three tasks (Lukács & Kemény, 2012).

There are, however, also a number of similarities between the three tasks. These similarities include vulnerability of performance in similar clinical populations, reliance on similar anatomical structures as evidenced by data from imaging studies, and reliance on implicit versus explicit processes shown by behavioural studies. These similarities as well as the differences will be explored in detail in the following sections: first, we discuss neuropsychological and imaging results using the three tasks in sections 1.2 and 1.3. As there are several lines of research in the literature that propose a close connection between IL and language (Friederici, Steinhauer, & Pfeifer, 2002; Gomez & Gerken, 2000; Ullman & Pierpont, 2005), we are also going to look into the nature of that relationship by reviewing existing results in 1.4, and devoting a study to testing whether the impairment of language is associated with impairment of IL.

In Section 3, we focus on how stimuli of different modalities, domains or stimulus structures affect learning on the three tasks. This section reports numerous differences in the main stream of research among the tasks. Then in Section 4, looking into the implicit versus explicit nature of the three tasks, we will again see more similarities than differences between them.

For all three tasks, a number of modifications and different versions have been developed during the years of research to test more specific questions. As here we are interested in the general structure and properties of the tasks, in these introductory chapters we will mainly focus on the original versions of the tasks (this means that in the case of the SRT task we will focus on learning deterministic sequences, while in the case of the AGL task, learning a finite-state grammar). We will only introduce data from other versions where the argument requires us to do so.

1.2. Neuropsychological results

The three different traditional implicit paradigms were designed to show that patients with a severe declarative deficit show a learning trajectory comparable to healthy controls. Patients with amnesia can be characterised with a loss of memory functions where conscious recollection is required. These patients are impaired on explicit memory functions, while implicit memory functions are intact (Knowlton, Mangels et al., 1996), which leads to the assumption that if amnesic patients show decreased performance on a given task, then that specific task requires explicit memory functions. At the same time, if amnesic patients show normal performance on the given task, it is a sign that only implicit functions are required.

Whereas amnesic patients show impaired explicit learning, patients with Parkinson's disease (PD) can be characterised with impaired implicit learning (Knowlton, Mangels et al., 1996). If Parkinson's patients are impaired on a specific task, that task is assumed to rely on implicit processes, while if patients show intact performance, the task is purely explicit. Note that studies testing implicit versus explicit learning in amnesia and Parkinson's syndrome usually assume dual memory systems: one for the implicit and one for the explicit memory functions (see below and Frensch & Rünger, 2003). In the following section we review evidence from neuropsychological studies. These studies are introduced task by task, first focusing on patients with a declarative deficit (amnesia or Alzheimer's disease), then on patients having procedural impairments (Parkinson's patients).

As most experiments of the current dissertation deal with the Weather Prediction task, more emphasis will be given to this particular paradigm. Knowlton, Squire and Gluck (1994) used three tasks with identical structure but different stimulus sets to compare patients with amnesia and age- and education-matched healthy participants. Results showed that on the early stages of the task the two groups' performance did not differ from each other, and both groups performed above chance. Later in the task healthy participants showed higher performance, while both groups still performed above chance.

As mentioned above, patients with amnesia are often contrasted with Parkinson's patients. PD patients show a deficit on the WP task already in the early phases, and their performance hardly rises above chance level even later (Knowlton, Mangels et al., 1996). Also, patients with amnesia are unable to answer debriefing questions like 'How many cues could have appeared on the screen simultaneously?', while PD patients can answer these questions easily (Knowlton, Mangels et al., 1996). Since patients with a procedural deficit (PD patients) show impairment in the early phases of the task, and patients with declarative deficit (amnesic patients) show difficulties in the later phases of the task, solving the WP task is concluded to rely on the procedural system in the early stages, and the declarative system in the later phases (Knowlton, Mangels et al., 1996).

As described above, another line of research tapped into qualitative differences in the process of learning or categorization throughout the WP task by analysing strategy use (Gluck et al., 2002). Papers cited above only used performance measures to show that patients with amnesia and PD show different types of impairments on the WP task. These studies were replicated later in order to see whether different clinical groups differ from healthy control participants in the way they solve the task. Results showed that while Parkinson's patients were unable to develop a strategy in which they consider all cues (Shohamy, Myers, Onlaor, & Gluck, 2004), whereas hypoxic patients (with hippocampal malfunctions) were unable to develop strategies at all (Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004). These results again suggest a mixed model of implicit and explicit functioning, though in this case the difference is not time based, but strategy-based. These studies confirm the hypothesis of Gluck et al. (2002), who suggested that strategies where participants are required to focus on only one cue (One-cue and Singleton strategies) are easy to verbalize, hence they require explicit processes. On the other hand, multi-cue strategy, that requires the combination of cues is difficult to verbalize, and is expected to be implicit.

In sum, neuropsychological results of the WP task are contradictory. This is not unique though. The same applies to the Serial Reaction-Time task. Early results showed that people with Alzheimer's disease (Knopman & Nissen, 1987), patients with amnesia of several origins (Reber & Squire, 1994), and patients with amnesia due to Korsakoff's Syndrome (Nissen, Willingham, & Hartman, 1989) perform normally on the SRT task. Results also showed that Korsakoff's patients were even able to retain the sequence for a week. However, neither clinical groups showed any explicit awareness (Hartman, Knopman, & Nissen, 1989).

These results have been questioned later. The hypothesis of Curran (1997) was that the previously observed intact learning is due to sequence characteristics: clinical patients with severe declarative malfunction are able to learn a sequence if the adjacent sequential elements can be predicted (first order predicative sequence, FOP), but not in the case of second order predicative sequences (SOP). A FOP sequence used by Curran (1997) was A-B-A-D-B-C-D-C-A-D-B-C, where each letter signifies one of the locations. In this sequence, all element-types have a frequent antecedent (adjacent in 67%), an infrequent antecedent (adjacent in 33%), and a non-antecedent (never adjacent). In the case of second order predicative sequences (SOP), sequential elements are predicted from the preceding two elements. In a sequence like A-B-A-D-B-C-D-A-C-B-D-C (Curran, 1997), all location combinations appear with the same frequency. However, a given combination is always followed by the same location: A-B combination is always followed by another A element. Testing amnesic patients showed that the RT difference between random and sequence blocks was comparable between the clinical and control groups if FOP sequences were used, while there was a significant sequence learning advantage of the control group in the SOP condition (Curran, 1997).

The same controversy applies to Parkinson's disease. A number of studies revealed a severe impairment of learning on the SRT task (Siegert, Taylor, Weatherall, & Abernethy, 2006). The observed impairment may be different though from publication to publication. Jackson and colleagues (1995) found that PD patients show no learning at all: their reaction times for the random block does not differ from the surrounding sequence blocks. Other studies found a different pattern though. Ferraro, Balota and Connor (1993) compared non-demented PD patients to healthy elderly adults (as well as two other clinical groups), and found that while the reaction times in general hardly differed in the two groups, the PD group showed a significantly lower RT increase on the appearance of the random block. This suggest that while the practice effect did appear, sequence-specific learning was smaller in PD. Similar results were observed by Pascual-Leone et al (1993). Westwater et al (1998) tested whether decay in Parkinson's patients may be due to peripheral motor problems. Instead of a button press, participants were expected to verbally respond to the location of the stimulus. Results showed that PD participants had higher average reaction times as well as a reduced sequence learning effect. Note that similarly to Ferraro et al (1993) and Pascual-Leone et al (1993), also this study found a sequence-specific RT increase from the sequential to the random block. On the other hand, Smith, Siegert and McDowall (2001) also used a verbal SRT task. They found that although non-demented Parkinsons's patients are significantly slower than age-matched controls, their sequence-specific RT increase was comparable to healthy controls. In general, results are mixed. PD patients seem to be impaired on the SRT task, but results are not convincing whether this impairment is response specific (showing higher motor or verbal RTs) or sequence specific.

In Artificial Grammar Learning, the first paper to study AGL performance in amnesia (Knowlton, Ramus, & Squire, 1992) found that patients with amnesia performed similarly to healthy control participants on the AGL task, while when the task was to base classification on explicit knowledge of exemplars, performance significantly differed between the two groups – in favour of the control group. Results also showed that the recognition of the previously used sequences was also impaired in amnesia. These results were also confirmed by a later study (Knowlton & Squire, 1996). Three experiments are reported in Knowlton and Squire (1996). In Experiment 1, patients with amnesia were found to perform comparable to healthy control participants, while Experiment 2 showed a deficit of the clinical group for declarative learning of chunks. Experiment 3 tested whether either group was able to transfer knowledge from learning with one set of stimuli to a test phase with another set of letters. Results showed that both groups were able to transfer abstract knowledge, and the level of transfer was comparable between the two groups. It is also important though, that performance on the transfer condition was significantly lower than performance of the same letter set. Studies of Alzheimer's disease (Reber, Martinez, & Weintraub, 2003) found the same phenomenon: patients showed intact artificial grammar learning and impaired recognition performance compared to age and education matched healthy controls. These results show that the acquisition of abstract information takes place in both amnesia and healthy functioning.

Channon et al (2002) also examined the performance of patients with amnesia as well as the role of abstraction in AGL. In their design, there were high and low familiarity sequences. High familiarity sequences were composed of bi- and trigrams that appeared in a number of times during the training phase, while bi- and trigrams that made up low familiarity sequences were infrequent. Results revealed that both groups showed a significantly higher performance on the high familiarity elements. This result suggests that neither clinical nor control participants were able to abstract the knowledge during the training. Also, in general, patients with amnesia showed a significantly lower performance on the AGL task. The authors suggest that these may be due to previous studies being too weak to detect group differences, i.e. control for familiarity was the element that made a difference.

As reviewed above, studies using the WP and SRT tasks mostly found an implicit learning impairment in Parkinson's disease. This is not the case however in Artificial Grammar Learning. Meulemans, Peigneux & Van der Linden (1998) compared the performance of PD patients and age-, sex- and education-matched healthy controls. Results showed that the performance of the two groups did not differ from each other significantly for the first presentation. Another study, that was already introduced at the SRT section, showed that PD patients show intact performance on both the SRT and the AGL tasks (Smith et al., 2001). These results were further confirmed by Witt, Nühsman and Deuschl (2002). In general, papers showing that PD patients are not impaired on the AGL task, also suggest that artificial grammar learning does not rely on the striatum (Peigneux, Meulemans, Van der Linden, Salmon, & Petit, 1999).

There is also neuropsychological evidence showing that PD patients do in fact show a reduced performance on the AGL task. Smith and McDowall (2006) employed a feedback-based AGL task. In the task there was no training session: participants had to decide whether presented sequences were grammatical or ungrammatical. They received feedback after each decision. Results showed that PD patients showed a decreased performance, especially in the early phases of the task. As the as authors suggest, the decreased learning of PD patients may be due to the fact that learning relied on feedback. These results are in concert with a previous study suggesting that PD patients are impaired on feedback based learning (Shohamy, 2004). In sum, neuropsychological results show that both patients with amnesia and PD patients show preserved artificial grammar learning.

Neuropsychological studies on the three traditional implicit tasks show that patients with amnesia or Alzheimer’s disease mostly show preserved performance. This intactness however seems to be reliant only on specific settings. Results from patients with amnesia and Alzheimer’s disease suggests that these tasks do not rely on hippocampus dependent explicit/declarative processes (Knowlton et al., 1994). PD patients on the other hand show diverging results. A severe deficit is seen in the case of the WP and SRT tasks, PD patients’ performance on the AGL task however seems to be intact. These studies suggest that while all tasks are non-declarative, they might rely on different brain circuitry (Skosnik et al., 2002). This will be further examined in the following section, covering neuroimaging studies.

1.3. Imaging results

Positing the existence of two distinct memory systems, and having different clinical populations being selectively impaired on a given memory system suggest that these systems should not only dissociate behaviourally, but their neural bases may also be different. In general, patients with amnesia are mostly associated with hippocampal malfunctions (Knowlton & Squire, 1993). The implicit/procedural memory system on the other hand seems to rely on a more widespread network of brain areas. These areas include the basal ganglia, the cerebellum, Broca’s area, the fronto-striatal pathways, and areas of movement, action planning and motor execution (Ullman & Pierpont, 2005). In general, tasks eliciting hippocampal activation are usually considered an explicit/declarative task, while the lack of hippocampus activation suggests implicitness/proceduralness (Poldrack et al., 2001; Poldrack & Foerde, 2008; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999).

The WP task being partially implicit has been underpinned by imaging results. In a set of fMRI experiment Poldrack et al (2001; 1999) found that the mediotemporal lobe (MTL) is active in the beginning of the task, but rapidly becomes deactivated. The striatum on the other hand is not active in the beginning, but rapidly becomes activated. Combining these data with neuropsychological findings, Poldrack et al (1999) suggest that early in the task participants acquire task-related declarative information, but as solving the task does not require declarative functioning, striatal activation emerges. This hypothesis is in line with the data that patients with amnesia learn the associations between cues and outcomes, but do not acquire non-predictive task specific information, as revealed by the debriefing questions (Knowlton, Squire et al., 1996). Gluck et al (2002) on the other hand interprets these results in a way that the initial MTL activation signifies declarative functioning in connection with the use of the singleton or one-cue strategies, while the later MTL deactivation and striatal activation is the sign of procedural functioning underlying multi-cue usage.

In another experiment Foerde, Knowlton and Poldrack (2006) found that performance on the WP task was correlated with MTL activity, but if a demanding declarative secondary task was introduced, the level of striatal activation predicted WP categorization performance. These results support the competitive memory systems approach, suggesting that the declarative system is only employed if resources are available, but if resources are not available, the procedural system is in use. These are further confirmed by the fact that dual task setting did not interfere with performance, however declarative knowledge measured on debriefing questions were significantly lower than in the single task setting (Foerde et al., 2006). Later behavioural results confirmed these results, showing that learning takes place in dual task conditions, but performance is significantly lower (Foerde, Poldrack, & Knowlton, 2007).

Previous SRT studies have also confirmed the dual-memory systems approach (Squire et al., 1993). PET results showed that an implicit sequence learning condition, where participants were naïve to the appearing sequence elicited right premotor, striatal and thalamic activation. Different areas were activated by the explicit condition, where participants were informed of the presence of the repeating sequence. The explicit condition elicited activation in the visual and language areas. The results are interpreted as signs of conscious learning strategies (Rauch et al., 1995). Later fMRI studies confirmed these results: both younger and older adults showed bilateral parietal, frontal, supplementary motor, cerebellar and basal gangliar activation for implicit sequence learning in the SRT task (Daselaar, Rombouts, Veltman, Raaijmakers, & Jonker, 2003). Neuroimaging studies were integrated in the competitive dual memory systems framework (Poldrack et al., 2001) by showing striatal activation for implicit and MTL activation for explicit sequence learning. Results also showed that there was a negative correlation between striatal and MTL activation, especially for younger adult participants (Dennis & Cabeza, 2011).

Previous imaging results showed that striatal and MTL activation contrasts implicit and explicit learning. There are other results though suggesting a more complex account. Following the line of research introduced by Curran (1997), Schendan and colleagues (Schendan, Searl, Melrose, & Stern, 2003) compared implicit and explicit sequence learning on second-order predicative sequences. Results showed that the instructions concerning the presence of the repeating sequence made no difference: MTL activation was recorded in both cases. This is in line with Curran’s (1997) results, suggesting that MTL activity may be required for higher order predictions. Results also showed that the MTL activation is present regardless of the conscious awareness of sequence knowledge (Schendan et al., 2003).

So far, imaging data of the WP and SRT tasks showed that these tasks at least partially rely on both the striatal system and the MTL. Striatal and MTL activations are in negative correlation, which is interpreted as competition between the two distinct memory systems (Poldrack et al., 2001). As reviewed above, there were neuropsychological differences observed between the AGL and the other two tasks. These differences appear in neuroimaging studies too.

Using the AGL task, Seger et al (Seger, Prabhakaran, Poldrack, & Gabrieli, 2000) found that grammaticality judgements were accompanied by the suppression of the precuneus, and activation of the left frontal, left occipital and left parietal areas, whereas recognition activated right frontal and medial occipital areas and the precuneus. These results again are in concert with the multiple memory systems approach (Poldrack et al., 2001). Others results also confirmed the competitive nature of declarative and non-declarative memory systems. Similarly to Channon et al (2002), Lieberman and colleagues (Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004) incorporated chunk strength into the AGL task. Neuroimaging confirmed that activation with rule-adherence was associated with activation in the right caudate, whereas the processing of chunk strengths elicited MTL activation. There was also a strong negative correlation between MTL and right caudate activation (Lieberman et al., 2004).

Others on the other hand suggested that widespread increases in occipital, posterior temporal, parietal and prefrontal cortices during grammaticality judgement reflect cognitive demands (Skosnik et al., 2002). Contrasting grammatical and non-grammatical sequences showed that greater left superior occipital and right fusiform gyrus activation was found for grammatical sentences. Also, correct categorization elicited greater activity in the left superior occipital cortex and the left angular gyrus. The comparison of grammaticality versus recognition showed increased left angular gyrus activation in the former. These results suggest that processing grammatical sentences show more similarities with word-form processing and mental calculation than with other non-declarative tasks. The authors suggest that these results might indicate a separate non-declarative mechanism for AGL, relying on left superior occipital and inferior parietal cortices (Skosnik et al., 2002).

In sum, neuroimaging results seem to be in line with the dual memory systems approach. For all three tasks there is a negative correlation in MTL and striatal activation.

However, the comparison of results using the WP and the SRT tasks seem to show that the AGL relies on somewhat different neural structures. These data are in concert with neuropsychological studies: while the intact performance of patients with amnesia indicates independence from MTL activity, differences in PD performance shows that the striatal system is differentially activated by the different tasks. Neuroimaging results confirmed this hypothesis.

1.4. Implicit/procedural learning and language

As mentioned in the introduction, there is a great body of evidence indicating that language and implicit learning are related to each other. Evidence in favour of this approach comes from two sources. One is the Procedural/Declarative theory of language (Ullman et al., 1997) suggesting that different aspects of language rely on different memory systems. The other is the literature on AGL and statistical learning arguing that this type of learning is one of the basic learning mechanisms in the acquisition of several aspects of language. As it is discussed below, the former issue is closely related to the above reviewed neuropsychological and imaging studies suggesting the dual nature of human memory.

Ullman's approach takes the traditional view of language depending on the two major functions of the mental lexicon and the mental grammar (Ullman, 2004). The mental lexicon is a place for storing words, their meaning, their phonological forms, the way these forms can be modified, and all kinds of idiosyncratic information. Elements of the lexicon are suggested to be "memorized", one by one, through learning. Grammar on the other hand is composed of rules and regularities that define how different lexical elements may be combined, how hierarchical relations can be coded within language.

The fact that elements of the mental lexicon are factual bits of information, while elements of the grammar are procedures dragged attention to the similarities between the lexicon-grammar and declarative-procedural memory distinctions. This was further confirmed by neuropsychological studies. Patients with Alzheimer's disease, Parkinson's disease, anterior aphasia (agrammatism) and posterior aphasia (word finding deficits) faced an inflection task of regular and irregular verbs (Ullman et al., 1997). Regular past tense forms are suggested to be generated by grammatical regulations, whereas irregular past tense forms are thought to be retrieved from the lexicon. Results showed that patients with declarative deficits (Alzheimer's patients and patients with posterior aphasia) produced more errors in finding the irregular forms, whereas grammatical (anterior aphasia) and procedural deficits (PD patients) lead to difficulties in generating regular past tense form.

This theory in general suggests that grammar relies on the procedural memory, while the lexicon depends on the declarative system. This theory was also generalized to Specific Language Impairment (SLI), a developmental deficit that can be characterized as a specific impairment of linguistic abilities with intact memory, intelligence and lack of auditory or neurological impairments (Bishop, 1992; Leonard, 1997). The procedural-declarative framework of language suggest that the impairment in SLI is not confined to the linguistic domain: according to Ullman and Pierpont's (2005) Procedural Deficit Hypothesis, language impairment is a result of deficits in the procedural system.

As indicated above, numerous researchers consider language acquisition a form of implicit statistical learning, and, accordingly, implicit learning paradigms have extensively been used to model the acquisition of mother tongue. In all cases, a small, non-existing language is invented and taught to participants. Gomez and Gerken (2000) categorized these studies into four different categories. These categories differ from each other in the amount of abstraction required for the task.

The first level requiring the least abstraction is the segmentation of word-like elements (Aslin, Saffran, & Newport, 1999; Saffran et al., 1996). This is a preferential looking paradigm designed for testing infants. Similarly to the original artificial grammar learning task, participants (infants in the present case) face a training and a test session. In the training session, there is a continuous stream of auditorily presented syllables. There is no pause between the stimuli. The only structure in the task is the probability with which one element may follow another (Transitional Probability – TP). In the original experiment there were two TPs: 0.33 and 1. That is, a given syllable could either be followed by one specific syllable, or by three different syllables. This way, there were trisyllabic pseudowords made up by CV syllables, and the coherence of syllables was only indicated by TPs. In the test phase participants faced a pair of trisyllabic pseudowords: both TPs were 1 in one of the words, while one of the TPs was 0.33 in the other. Looking times towards the source of stimuli were measured. The question was whether infants were able to differentiate trisyllabic words (with both TPs = 1) and word fragments (one of the TPs = 0.33). Results showed that children already at 8 month of age were able to differentiate between stimuli based on TPs. This paradigm is mostly referred to as Statistical Learning (SL). Studies of SL are further discussed in the light of modality and domain issues in Section .

The second level in abstraction is the acquisition of word-sequence information (Gomez & Gerken, 2000). This includes learning long distance dependencies (Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006) as well as the acquisition of finite-state grammars (Reber, 1967). Long distance dependencies are mostly examined using context free grammars. Context free grammars are models of centre embedding where a pair of stimuli is embedded between elements of another pair. This leads to a structure like $A1A2B_2B_1$, where the pairing is shown by the indices. The question is whether participants are able to abstract and learn the underlying structure of the grammars. Some of these data have already been introduced above.

The third level is the generalisation of relations. Studies of this aspect usually employ simple rules like ABA (Marcus, Vijayan, Rao, & Vishton, 1999). Similarly to the SL paradigm, infants are familiarized to the rule. Later, infants are tested whether their rule-knowledge is transferrable to other sets of stimuli. That is, infants face ABA and ABB type sequences, and looking times reveal whether they are able to differentiate the two structures. Results show that infants at the age of 7 months are able to categorize the different sequences (Marcus et al., 1999). The difference from the second level is that infants do not only have to abstract the information, but also transfer it to stimulus sets they have not met before.

The fourth level of abstraction taps on the abstract usage of syntactic categories (Gomez & Gerken, 2000). Studies using this approach usually employ more complex grammars than the finite state grammar used in the AGL task. Friederici, Steinhauer and Pfeifer (2002) trained participants on a game that required the mastery of complex language that included syntactical rules as well as semantic coding of objects. Results showed that after an up-to 6 hour long training, syntactic anomalies elicited similar event-related potentials for the newly learned artificial language as in their mother tongue.

This section showed that research on implicit learning and language are linked in two different ways. On the one hand, there is evidence that the grammatical domain of linguistic abilities uses the same networks as the procedural system, and there is also an overlap in the mental lexicon and the declarative memory. On the other hand, different artificial grammar learning methods are used to model language acquisition, with the presumption that artificial languages are acquired in the same way as natural languages. We cited electrophysiological studies supporting this hypothesis. Evidence supports the link between the linguistic and procedural networks. The link between the procedural system (and implicit learning) and language predicts that the decrement in grammatical abilities is associated with impairments in procedural learning. Study 1 taps into this issue. Focusing on the Procedural Deficit Hypothesis, we tested children with SLI and age-matched controls on the Weather Prediction task, and found a severe procedural learning deficit in the former.

2. Developmental changes in IL

Despite the fact that previous systematic developmental studies are sparse, three theories emerged in connection with the development of implicit learning. The first theory originates from Reber (1993), and puts the development of implicit learning into an evolutionary perspective. Reber's view relies on the assumption that the procedural memory system is both phylogenetically and ontogenetically older than the declarative memory system (Squire, 1986). In this framework, the fact that children acquire a large amount of information easily within the first few years of life can only be explained in terms of the invariability and age-independence of procedural learning functions. That is, children learn procedurally because that is the only way they can acquire information, and since all children gather a great amount of knowledge, procedural learning is suggested to be invariant (Reber, 1993). This approach predicts a constant learning trajectory in implicit learning.

There are empirical studies confirming this hypothesis. Meulemans, Van der Linden and Perruchet (1998) compared the performance of six- and ten-year-old children to that of young adults (between 18-27 years of age). Results showed that despite the differences in age-related reaction-time baseline, sequence-dependent RT differences were comparable among the three groups. The lack of age-related sequence learning effect was also confirmed by an analysis in which a transformed [(RandomRT - SequenceRT)/(RandomRT + SequenceRT)] ratio was used to minimize the impact of RT baseline differences. Results of accuracy also indicated the lack of age-effects in sequence learning.

Not all studies suggest age independence. Thomas and Nelson (2001) compared the performance of four-, seven- and ten-year-old children, and found that age-differences only affected the baseline RTs and explicit knowledge of the task. Thomas and Nelson (2001) however also note that there might be some age-related differences, as 4-year-olds showed a somewhat lower sequence learning effect than the other two groups. A similar phenomenon was observed in a study with Artificial Grammar Learning by Witt and Vinter (2012). In the study, 120 children between 5 and 8 years of age were compared in four age-groups. Results showed that performance was similar between ages 5 to 7, whereas 8-year-old children showed a more complex profile, with more sensitivity to grammatical rules. This observation again suggests that implicit learning is not age-dependent. In a third study, Thomas et al (2004) compared the performance of adults (23-33 years) and children (7-11 years) on the SRT task. Using Z-transformed RT data, a significantly lower sequence-specific RT difference was observed for children than for adults, while both adult and children participants were reported to have no explicit awareness of the sequence. Another study using artificial grammar learning found the same pattern of age-related differences (van den Bos & Poletiek, 2006). These four studies suggest that implicit learning is indeed age-dependent, and the older children get, the better they perform in implicit learning tasks. Note that the reported studies only report performance of children and adults, and not that of older adults. The reason for the selective introduction is that the studies presented in the dissertation only report data from children and young adults.

So far we reviewed two alternatives: age-independence and age-related increase in learning performance. A third possibility is the critical period hypothesis. According to common knowledge, there are skills that should be learned in early ages otherwise there is no chance for the proper control of the given skills. There is ample evidence that in the case of second language acquisition (Johnson & Newport, 1989) or playing a musical instruments (Watanabe, Savion-Lemieux, & Penhune, 2007) the earlier one starts learning, the better performance can reach. These results suggest that the procedural system is more sensitive in early ages than later. Janacsek, Fiser and Németh (2012) made a systematic analysis of data by 9 different age-groups from 4 to 85 years of age. They used the Alternating SRT task, which is a second order predicative version of the original task where every second element of the sequence is deterministic, while the elements in between appear randomly (Howard & Howard, 1997). Using raw RTs showed that the gap between sequence and random RTs are highest in children, somewhat lower in younger adults, and lowest in older adults (Janacsek et al., 2012). Interestingly, a different pattern emerged using z-transformed RT data. These results seemed to indicate that when baseline RTs are controlled, younger adults show significantly higher learning performance than both children and older adults. In sum, results are not clear, but Janacsek and colleagues conclude that children show better sequence learning than younger adults do, who in turn do better than older adults.

Three different theories emerged, all predicting different learning trajectories during childhood. The constancy hypothesis (Reber, 1993) suggests that implicit learning is age-independent, predicting the same level of performance for adults and children. The improvement hypothesis (Fletcher, Maybery, & Bennett, 2000) predicts better learning for adults than for children, while the reverse is predicted by the critical period hypothesis (Janacsek et al., 2012). Note that only studies using the SRT and AGL tasks were reviewed in this section. The reason for this is that apart from our studies (Study 1 and 2), only adults or clinical population were tested using the WP task. The only exception is by Kéri, Szlobodnyik, Benedek, Janka and Gádoros (2002), who used the WP task for testing children with Tourette syndrome and age-matched TD children, but they did not include an adult group for comparison. In Study 1, which focuses on the comparison of children with SLI and typically developing children, the data of the two groups is also compared to adults. Study 2 on the other hand reports and compares data from typically developing children and adults in four different conditions.

3. Domain and modality independence in Probabilistic Categorization, Artificial Grammar Learning and the Serial Reaction-time task

Strong claims of modality- and domain independence in implicit memory mostly revolve around the acquired knowledge, i.e. abstractness vs. modality specificity of the representations that result from learning, not the learning process itself. A number of papers suggest that domain- and modality independence could only be investigated with experiments using transfer (Altmann, Dienes, & Goode, 1995), and claim that the acquired information may be abstract enough to be transferred from one set of stimuli or responses to another set. At the same time implicit and procedural information are sometimes considered to be inflexible and superficial enough not to be transferrable (Cleeremans, Destrebecqz, & Boyer, 1998). Even in these cases, the lack of transfer does not entail that the learning mechanism itself is specific to a given domain or modality. Another way of defining modality and domain independence is to assume the same effectiveness of the given learning mechanism with stimuli across different modalities and domains. The debate over the abstractness vs. modality and stimulus specificity of learning is most clearly present in the artificial grammar learning literature, and papers cited below mostly suggest or reject modality or domain independence based on experiments that compare structurally equivalent tasks that employ different sets of stimuli.

Previous studies of domain and modality dependence mostly employed the original Artificial Grammar Learning or the Statistical Learning task. Studies using the SL task showed that infants are able to learn transitional probabilities in a different domain, i.e. if nonverbal auditory stimuli are used in the same task (Aslin et al., 1999; Saffran et al., 1999).

Modality independence was also tested with a similar paradigm, and infants were able to acquire spatial dependencies based on statistical information in a visual task as well (Fiser & Aslin, 2002). To conclude, statistical learning proved to be a domain general and modality independent learning mechanism.

Others suggest that when the task is to acquire deterministic rules, learning might not be domain-independent. As explained above (Section 1.4), 7-months infants were shown to be able to transfer simple rules, like ABA from one set of stimuli to another (Marcus et al., 1999). Later results suggested that this kind of rule abstraction may be restricted to the linguistic domain, since infants showed no learning if visually presented geometric shapes were used instead of auditory verbal stimuli (Marcus, Johnson, Fernandes, & Slemmer, 2004). However as critiques have pointed out (Saffran, Pollak, Seibel, & Shkolnik, 2007), this effect is not necessarily due to the domain dependence of rule learning, but is probably just an artefact that reflects infants' categorical perception abilities, i.e. infants are unable to categorize geometric shapes as belonging to the same category. If stimuli are designed in the way that infants categorize each stimulus as a member of the same category, rule learning appears. Results are in concert with this hypothesis: infants are able to learn and generalize rules if stimuli consisted of visually presented pictures of different dogs (see Quinn, Eimas, & Rosenkrantz, 1993 for details of categorization; see Saffran et al., 2007 for more details on visual rule learning). All in all, results show that rule learning – just like statistical learning – is a domain and modality independent learning mechanism.

Altmann, Dienes and Goode (1995) examined intermodal knowledge transfer in the original Artificial Grammar Learning task. In the training phase, participants faced sequences of tones with different pitch or different letters (Experiments 1 and 2), sequences of spoken syllables (Experiment 3) or arbitrary graphical symbols (Experiment 4). Testing of grammar knowledge was done using the same grammar, but stimuli from another modality: letter or tone sequences in Experiments 1 and 2, arbitrary graphical symbols in Experiment 3 and written syllables in Experiment 4. Results show that in all four experiments, participants were able to transfer their grammar knowledge from training to test modality. This provides strong evidence that Artificial Grammar Learning is not bound to any modalities.

Conway and Christiansen (2005; 2006) on the other hand, review and extend the evidence for modality constraints on the mechanism of artificial grammar learning. They found both qualitative and quantitative differences in learning in the tactile versus visual versus auditory domains, showing that for sequential information, auditory learning is more effective than learning in either the visual or the tactile domain. Also, auditory learning was most sensitive to chunk information at the end of items, while initial chunk information was more important in tactile learning. The authors argue that modality-constrained statistical learning reflects "general processing differences that exist among the various statistical learning subsystems, with the auditory system excelling at encoding statistical relations among temporal elements and the visual system specializing primarily in computing spatial relationships" (Conway and Christiansen 2005, p. 37.).

The focus of studies examining domain and modality dependence in the Serial Reaction Time task is somewhat different. The SRT task has mostly been used with visual stimuli: an asterisk or any other target stimulus appearing on different target locations. A number of studies however found implicit sequence learning effects with different sound (Buchner, Steffens, Erdfelder, & Rothkegel, 1997; Buchner, Steffens, & Rothkegel, 1998; Schmidtke & Heuer, 1997; Zhuang et al., 1998), but there are also other studies showing no sequence specific learning with sounds (Perruchet, Bigand, & Benoit-Gonin, 1997). In Riedel and Burton (2006) participants faced a Serial Reaction-Time task where the four target stimuli were four auditorily presented colour names pronounced by four different speakers. For half of the participants, the task was to press different buttons according to the speaker's identity, whereas the target dimension was the colour names for the other group of participants. Results showed that sequence specific decrease was only observed in both groups if the target dimension became random. That is, participants' RTs in the colour condition only increased if the colour sequence became random, the random organization of the speakers did not affect reaction times. These results show a special case domain dependence. In AGL and SL studies, the notion of domain specificity explores whether the extraction of transitional probabilities is strictly confined to verbal stimuli as it is a form of language acquisition. In the current case, domain dependence means that there are learning differences based on whether the specific domain in which the sequence appears is the target domain (the domain that defines the response) or an unattended domain. While Riedel and Burton (2006) found that learning in the SRT task is only possible in the target domain, there are other studies showing the reverse. Mayr (1996) showed that in the visual modality, participants are able to learn both attended and unattended domains. In his experiments, participants faced a visual SRT task where the two domains were stimulus location (where did the target appear) and stimulus identity (what exactly appeared). Results showed that randomizing either the target or the unattended domain resulted in an increase in RTs.

Differentiating target and unattended domains leads to the question of what exactly is being learnt during the Serial Reaction-Time task. There are three competing theories. On the one hand, learning might be based on effector-specific motor movements (Deroost et al., 2006). In such a case specific effectors, specific "organs" study the sequence. This may be interpreted as a sequence of specific muscle movements or the sequence of movements by larger units like fingers. The second possibility is that participants learn an effector-free movement sequence: the sequence of responses (Willingham, Wells, Farrell, & Stemwedel, 2000). In this case, participants learn to predict the next expected response. The third possibility is learning a perceptual sequence (Remillard, 2003). This would suggest that participants learn to predict the next percept they will receive. As it will be further discussed in Study 3, researchers are likely to suggest the exclusivity of one of the learning domains (Remillard, 2003; Willingham, 1999; Willingham et al., 2000), but there are results suggesting that the different learning domains may interact with each other (Deroost & Soetens, 2006a, , 2006b; Deroost et al., 2006; Nattkemper & Prinz, 1997; Nemeth, Hallgato, Janacsek, Sandor, & Londe, 2009). In Study 3 we will examine the link between perceptual and response learning, and whether the two learning domains are in fact dissociated.

The question of modality and domain specificity has not been raised in the literature concerning the Weather Prediction task, and even hypothesis driven modifications of the task are hard to find. While the first published experiment (Knowlton et al., 1994) already used three different tasks with the same structure, stimuli of the different tasks were presented visually, one of them being verbal while the other two non-verbal. Unfortunately, data from the different tasks were not statistically compared, however, some differences appeared. These differences are further discussed in Study 2.

There is only one previously documented hypothesis-driven modification to the WP task. As Parkinson's patients are impaired on shifting attention between different stimuli (Owen et al., 1993) a possible interpretation of impaired categorization performance measured by the WP task was that this impairment is due to the fact that in the case of cue-combinations, PD patients are unable to shift their attention between the simultaneously appearing cues. This may explain the decay in categorization. In a study, Shohamy, Onlaor, Myers and Gluck (2001) employed a modification to the WP task in which the predictive cues did not appear as separate images, but as features of a single item. That is, instead of showing 1, 2 or 3 out of four different cards simultaneously, participants were shown a toy that may have a moustache, a bow-tie, a pair of glasses or a hat. These four features were the different cues that were in a probabilistic relationship with the possible outcomes. The possible outcomes were not sunshine and rain, but participants had to decide whether these toys ask for chocolate or vanilla flavoured ice-creams. In this Ice-Cream (IC) task, PD participants showed a chance-level performance, suggesting that they really are impaired on probabilistic categorization, and the reduced performance is not due to problems with attention shifting (Shohamy et al., 2001). While PD patients are impaired on both the WP and the IC task, healthy participants show an uneven performance on the two tasks, with better learning on the WP task. Our Study 2 raises the question whether this decreased performance is due to the combination of cues into a single image, or due to differences in the story-line: in the WP task it is quite clear already from the beginning that different geometric shapes have nothing to do with weather, on the other hand it seems to be plausible that different people prefer different ice-creams (IC task). Even if we acknowledge that ice-cream preference is probably not correlated with the outfit of the customers, there is still a seemingly natural link between cues and outcomes in the IC task. An even more natural link is observed in the Medical Diagnosis task (Gluck & Bower, 1988; Knowlton et al., 1994), another version of the WP. In this task, participants face one, two or three out of four different symptoms, and their role is to categorize people with the given symptoms with one out of two diagnoses. In this case, the link between cues and outcomes (that is: symptoms and syndromes) is completely transparent. Unfortunately, neither these results were statistically compared with WP or IC results. This stimulus-outcome transparency – along with the holistic versus cue-based presentation – is also examined by Study 2.

Although the majority of neuropsychological and imaging results seem to show converging trends in the three traditional implicit learning tasks, this is not the case with research on modality-, domain- and stimulus dependence. In AGL and SL, studies of modality and domain specificity focus on whether learning appears in one or the other modality or domain, and whether a structure that has been acquired in one modality or domain can be transferred to another. In the case of the SRT, the notion of domain mostly focused on the domain of learning, that is whether participants learn a sequence of stimuli, a sequence of responses or a sequence of individual movements. The same question arises in target versus unattended domains: can pure perceptual sequence learning appear if another response domain is present (Mayr, 1996; Riedel & Burton, 2006). The WP literature lacks experimental psychological studies focusing on the effect of different stimulus sets on performance, except for the comparison of the Ice-Cream and Weather Prediction tasks (Shohamy et al., 2001). This theoretical line parallels that of the SRT task. In the SRT task, learnability is defined as the sequence being present in one or the other domain in the SRT task. The major question of the WP task is whether cues and outcomes appear in the same stimulus and same domain, or in different stimuli and domains.

As the notion of domain differs in the three tasks, it elicits different lines of research. In the case of the WP task the major question is whether cue-cue and cue-outcome domains affect learning. While previous literature did not study this question, collapsing previously reported data suggests that it may have an important role in probabilistic categorization. Study 2 tests whether the mode of stimulus presentation and the transparency between cues and outcomes affects learning performance on the WP task. On the other hand, the question of domains in the SRT task has been subject to intensive research. It is quite clear that there are different domains (stimulus, response, effector) in which learning can appear, it is unclear whether these domains are treated as independent of each other, or whether different structures may interact. These questions are in the focus of Study 3.

4. Implicit versus explicit knowledge in the IL paradigms

The traditional approach to implicit/procedural learning states that a task can be considered a procedural task if patients with severe declarative impairment show intact learning on that task. Two questions arise on this ground that might cast doubt on the validity of this approach. The first is whether healthy participants use the same learning strategies as patients with amnesia. It might be the case that a task is solved by patients with amnesia, but healthy participants use explicit/declarative strategies, as these strategies are available for them. This phenomenon is well explained by the competing memory systems approach (Poldrack et al., 2001), and has been confirmed by earlier imaging results (Foerde et al., 2006). Difference in the reliance on either of the memory systems may as well lead to better learning performance by healthy participants. The second question is a general question of experimental psychological studies in connection with the traditional implicit learning tasks: are they solved in an implicit or explicit way by healthy participants. Two major approaches are employed throughout the literature. The implicit versus explicit nature of the task can be directly examined by the collection of subjective self-insight measures. The other possibility is to employ modifications to the tasks that are expected to selectively impair implicit or explicit learning. The rationale for both methods is provided, and then highlights of research using the WP, SRT and AGL tasks are summarized. Previous literature is reviewed task by task.

Subjective data may come from two different sources: knowledge of the structure and knowledge of the decision. Structural knowledge is the (present or absent) knowledge of the underlying structure of the task the participant is exposed to (Dienes & Scott, 2005). Structural knowledge may be implicit or explicit. Implicit structural information is when one is not aware of the knowledge; still the presence of the knowledge is apparent from performance. A piece of explicit structural knowledge is when a participant reports some knowledge on how a cue and an outcome is associated (WP task), or how fragments of button presses follow each other (SRT task).

Neither implicit nor explicit structural knowledge is necessarily correct. In a number of cases, decisions are affected by biases. Koehler et al (1995) reported an experiment in which participants had to decide whether the person accused of murder was guilty or innocent. The evidence included a match of DNA sampled from blood found under the nail of the victim, and the blood of the defendant. Two conditions were used. One of the groups received information that there is a chance for random DNA match, and the chance is 1 in 1 000 000 000. At the same time, the probability of a human error to appear during the laboratory tests is 1 in 1000 cases. The other group was only informed that the aggregate probability of error is 1 in 1000. Note that in both cases the probabilities match, as the higher value defines the chance for an error. Results though showed that participants are more likely to present a guilty verdict in the first case (where the random match probability is provided). The guilty verdicts in this group were almost twice as frequent as in the second group. Presenting probabilities with higher numbers seem to lead to stronger decisions.

Explicit structural knowledge may also be incorrect. The most extreme case is having a strategy when there is no structure in the task to be solved. Such a case is the lottery. While winning numbers are drawn randomly, people still have strategies or favourite numbers that they use in trying to guess them. Similarly, still in lottery, when a number is not drawn for a long time, people may think it is becoming more and more likely for that number to be redrawn. The chance for the given number to be drawn is the same though as for any other numbers. This phenomenon is known as the *gambler's fallacy* (Ayton & Fischer, 2004).

These examples show that structural knowledge is a piece of information a decision may be made on. This information may be implicit or explicit, and may be correct or incorrect. The same possibilities may apply to self-insight too. Self-insight may characterise the decision process, whether the decision process is accessed consciously (Lagnado, Newell, Kahan, & Shanks, 2006). Similarly to structural knowledge, people may have or not have self-insight into their decisions. Also this self-insight may be correct or incorrect. An incorrect explicit decision is when one states to make a decision based on previous experience; while lacking previous experiences. This may include events like false memories: after listening to a list of words in a given category people are more likely to falsely report that some within-category words were on the list they have previously heard (Roediger & Mcdermott, 1995). An incorrect implicit decision is when participants report guessing while they in fact decide based on some other knowledge. This is what allegedly happens in implicit learning: the presence of structural knowledge with the lack of self-insight.

Self-insight measures are difficult to collect though, as introspection is not generally considered a reliable research method. The other line of research relies on the general assumption that specific events selectively impair implicit versus explicit processes. According to categorization literature, executive functioning has a positive relationship with explicit categorization, while there is no link with implicit categorization (Maddox & Filoteo, 2001). It has also been documented that an attention demanding secondary task impairs explicit but not implicit category (Waldron & Ashby, 2001; Zeithamova & Maddox, 2007) as well as sequence learning (Jimenez & Vazquez, 2005). Experimental psychological results are introduced below along with studies using self-insight. This is especially important in the case of the WP task, where direct measurements of self-insight have not yet been reported. Study 4 is an attempt at filling that gap.

4.1. The WP task

The experimental results published in connection with the implicit versus explicit nature of the WP task are sparse. These papers questioned the implicitness of the task based on two grounds: Lagnado et al. (2006) tested structural knowledge both blockwise and itemwise, while Newell et al. (2007) and Price (2009) used dual-task paradigms and also applied different modifications to the WP task.

Lagnado et al (2006) report two experiments. In their first experiment, participants provided cue-strengths after each block of 50 trials. Results showed that for the strong cues, the reported predictive values were near the pre-set value already at the beginning of the task, while the probabilities of the weaker cards were learned only later (probability ratings). Participants also reported that as time passed, they relied more on the stronger cues than on the weaker ones (cue-usage ratings). These results were replicated by trial-by-trial data collection in Experiment 2 (Lagnado et al., 2006).

As explained earlier, dual-task paradigms are employed to selectively impair declarative functioning. Along with the previously introduced studies (Foerde et al., 2006; Foerde et al., 2007, see Section 1.3), this method was applied by Newell et al. (2007). Results showed that a concurrent numerical stroop task decreased WP categorization performance (Experiment 1 of Newell et al, 2007). In Experiment 2, Newell et al used the same trial-by-trial reports employed by Lagnado et al. (2006) for both a declarative (observational) and a procedural (feedback) version of the task. The procedural version of the task was the original WP task, while in the declarative version, participants faced cues and outcomes simultaneously, and were instructed to learn stimulus-outcome associations.

Results showed that participants of the observer and feedback groups showed a similar performance, and also their cue-ratings were very similar. In term of cue-reliance, both groups reported that they rely on strong cues more than weak cues very early (Block 6 in the observation and Block 9 in the feedback task, with 5 trials in each block). Newell et al. suggest that this small difference should not be interpreted as evidence for separate systems underlying the different tasks, instead that both tasks rely on the same system, with the observation task being somewhat easier. In sum, neither Lagnado et al. (2006), nor Newell et al. (2007) suggest that learning on any form of the WP task is implicit.

Note that both the Lagnado et al. (2006) and the Newell et al. (2007) studies focused on the perception and recollection of cue-outcome contingencies, a task that was in part correctly solved by amnesic patients (Reber, Knowlton, & Squire, 1996, see above). On the other hand these papers also criticized analyses of strategy use of the WP literature (Gluck et al., 2002), and suggested that the low number of multi-cue strategy-use in the earlier studies was due to the fact that multi-cue strategy is defined as a maximization of the probability of correct responses. In the Gluck et al. (2002) study, participants were considered multi-cue strategy users if they responded according to the mean predictive values of the presented cues. That is, if a cue combination leads to sunshine in 75% of all cases, multi-cue users are expected to answer sunshine in all cases. Instead of the maximization, Lagnado et al. (2006) suggested a probability match, i.e. if a cue-combination leads to 75% sunshine, participants are expected to answer sunshine in 75% of the cases, and rain in the remaining 25%. Lagnado et al. (2006) found that 65% of participants used this multi-match strategy already in the first block, i.e. they were already able to integrate probability information from different cues. At the same time, Lagnado et al. reports that “The observed interaction between card type and block suggests that participants learned about strong cards (and discriminated between them) sooner than they did for weak cards” (page 166, Lagnado et al, 2006). In the same experiment, in a blocked cue-usage rating measure, participants of the same study reported that from Block 2, they found the strong cues being more important than the weak cues. Summing up the results by Lagnado et al. (2006), we find that in the early blocks, participants learn the probabilities of strong cues, rely similarly on both cue-strengths, and use a multi-cue strategy. In the later blocks, they show knowledge of cue-strengths for all cues, they report to rely more on the stronger cues, and still use a multi-cue strategy. This discrepancy between strategy use and self-report might suggest that, contrary to the interpretation of Lagnado et al. (2006) there may be a dissociation between explicit evaluation and performance itself (Jacoby, 1991).

Price (2009) employed two experiments to selectively disrupt implicit (Experiment 1) or explicit (Experiment 2) learning. In Experiment 1, a delay was introduced between cues and feedbacks. This was expected to decrease implicit categorization performance, but leave explicit categorization intact (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005). In Experiment 2, a short, one-item memory-scanning task was administered either immediately after feedback, or somewhat later, leaving time for feedback processing. This manipulation was expected to selectively impair explicit hypothesis testing, but not implicit categorization (Maddox, Ashby, Ing, & Pickering, 2004). Results showed that delayed feedback had no effect on learning in the WP task, while the feedback-processing disruption caused a serious decrease in categorization performance. In line with Lagnado et al. (2006) and Newell et al. (2007) this result suggests that the WP task relies on explicit processes. In sum, these results contradict neuropsychological and neuroimaging results, and suggest that the WP task is purely explicit, at least when it is solved by healthy young adult participants.

4.2. The SRT task

Similarly to the controversy concerning the WP task, there is a long standing debate whether learning on the Serial Reaction Time task can be considered fully implicit. In one of the early SRT experiments, Willingham et al (1989) explored the role of conscious awareness on learning. In Experiment 1, participants faced four blocks of 100 trials. After the fourth block, participants were asked 1) whether they noticed something, and 2) whether they noticed a repeating pattern throughout the experiment, and if so, then they were asked to replicate the pattern. Results showed that, out of the 60, 12 participants reported no knowledge of the sequence, 7 participants reported knowledge on the presence of the sequence, but were unable to provide any fragments of it. 29 participants reported the presence of the sequence, and were able to provide 4-9 element long fragments, while the remaining 12 participants were able to replicate the 10 element long sequence. Results showed that overall RTs of the no-knowledge and some-knowledge groups did not differ from each other significantly, while RTs of participants with full explicit knowledge were significantly lower than the other two groups on blocks 3 and 4 (with 100 trials in each block). This experiment suggests that two thirds of the participants of an SRT study acquire at least partial explicit sequence specific knowledge.

Others point out that these values may even be higher. Willingham and colleagues (1989) used a free recall task for measuring conscious awareness: participants had to provide the sequence by their own. However, other results show that the free recall method is insensitive (Shanks & St. John, 1994). Another method is cued recall, in which participants are shown a fragment of the sequence, and their task is to predict the next button press. Results show that even participants who are categorized as unaware based on the free recall method may perform well above chance on the cued recall task (cf. Jackson & Jackson, 1995). This suggests that even more than two thirds of SRT task solvers may acquire some explicit sequence knowledge. A problem with the cued recall task though that in a number of cases, participants may respond above chance based on gut feeling (Destrebecqz & Cleeremans, 2001). In such a case their response is contaminated by implicit knowledge (Neal & Hesketh, 1997).

According to Destrebecqz and Cleeremans (2001), the major problem is the assumption of process-purity: tasks are generally considered to rely purely on one system or the other. The process dissociation procedure (Jacoby, 1991) takes into account both the automatic and intentional processes during memory retrieval. The PDP suggests the usage of two experimental paradigms in memory retrieval studies: inclusion and exclusion. The inclusion paradigm is the retrieval of the elements to be remembered, while the exclusion paradigm is generation with the avoidance of the target elements. Representations of the inclusion paradigm are facilitated by both automatic and intentional processes. In such a case, both automatic and intentional processes point to the same direction: that is, the activation of the given representation. On the other hand, in the case of the exclusion paradigm, the two processes interfere. Automatic processes facilitate, intentional processes inhibit the retrieval of the given representation. Application of the PDP to sequence learning (Destrebecqz & Cleeremans, 2001) would suggest that both automatic and intentional processes work together in both free and cued-recall. On the other hand, if participants are required to provide a sequence of button presses that do not overlap the previous repeating pattern, implicit knowledge may still facilitate the reproduction of the pattern, while intentional processes inhibit it. The greater the overlap between inclusion and exclusion paradigms, the less likely that explicit processes have much contribution to retrieval. Great overlap means that the activation of implicit knowledge is present, but participants have difficulties with the intentional inhibition of response structure.

In the SRT study of Destrebecqz and Cleeremans (2001), two conditions were employed: a condition where there was no Response-Stimulus-Interval (RSI), and a condition where the RSI was 250 milliseconds. Results showed that both groups performed above chance on the inclusion test, while the groups only marginally differed from each other. Earlier studies on this data would suggest that both groups show explicit sequence knowledge. The exclusion paradigm showed a different picture though: participants of the RSI group were able to exclude sequence fragments during the generation task, whereas no-RSI participants continued to generate chunks of the previously learned sequence, indicating no control over the expression of sequential knowledge.

In a recognition task, participants faced triplets that were either part of the previously learned sequence, or not (Destrebecqz & Cleeremans, 2001). Note that as second order predicative sequences were used, all target-pairs were part of the sequence, that is, only the third element defined the fragment as part of the repeating pattern or not. Results indicated that participants of the RSI condition could differentiate well between the two types of fragments, whereas the confidence reports of the no-RSI participants were comparable for the two types. This confirms results of the exclusion criterion, that is, RSI participants have more conscious control over sequential knowledge, while mostly implicit processes underlie learning in the no-RSI condition.

Results of previous studies concerning the explicit versus implicit nature of the SRT task are mixed. It seems to be clear however that participants are very likely to acquire some kind of explicit sequence knowledge. This however may differ by procedure (Destrebecqz & Cleeremans, 2001), as well as motivation and task-difficulty (Fu, Fu, & Dienes, 2008). The most important aspect of the above introduced studies is that the SRT task cannot be discussed in the framework of process purity. Similarly to the WP task, the SRT task cannot be considered a purely implicit task. This assumption is in concert with the reviewed neuropsychological and imaging results.

4.3. The AGL task

The implicit nature of the AGL task has also been questioned by a number of studies. In three experiments Perruchet and Pacteau (1990) showed that declarative training on bigrams results in the same classification performance as procedural training on strings generated by an artificial grammar. They also reported that participants showed a very low, but above chance hit rate if valid bigrams appeared on a wrong position. Experiment 3 of

Perruchet and Pacteau (1990) showed that performance on the AGL task is sufficiently explained by explicit knowledge of bigrams. During the test phase, participants were asked to decide whether they recognize the fragments being shown. Participants of the intentional and incidental conditions gave similar ratings: recognition scores were significantly higher for grammatical than for ungrammatical letter pairs in both groups. The recognition scores by type hardly differed between the two groups.

A study by Higham, Vokey and Pritchard (2000) adapted the PDP (Jacoby, 1991) to Artificial Grammar Learning. In two experiments, participants had to memorize strings generated by two different grammars. In Experiment 1, in the exclusion task, participants were asked to classify strings conforming to only one of the grammars as correct, while in the inclusion condition, they had to categorize a string as correct if it conformed either of the grammars. Results showed that strings conforming to the to-be-excluded grammar were more likely to be accepted by inclusion than exclusion participants. This effect was somewhat decreased if there was a response time limitation. These results show that grammaticality judgements can be controlled if there is no time pressure. Another analysis showed that – in the exclusion condition – the acceptance of the to-be-excluded grammar strings was significantly higher than that of the non-grammatical strings. This effect argues for the presence of automatic processing. These results were reproduced in Experiment 2, which also showed that differences due to controlled processes disappeared after a 12-day delay period, while automatic processes were not affected. That is, grammatical knowledge of both grammars was still present after 12 days, whereas the earlier exclusion of one of the grammars did not affect performance (Higham et al., 2000).

Dienes and Scott (2005) combined the AGL task with self-reports of conscious awareness. After a training phase, participants were asked to classify new strings as grammatical or ungrammatical. After each decision, participants were asked to report the basis of their judgement. They could report that they had relied on guessing, intuition, pre-existing knowledge, rules or memory. Due to lack of such responses, the category of pre-existing knowledge was dropped. At the same time, guess and intuition, and rule and memory knowledge were clustered into implicit (former two) and explicit (latter two) categories due to theoretical and empirical (similarity in the number of responses) considerations. Results showed that there was a greater proportion of implicit than explicit answers. Results also showed that participants were more likely to give a correct classification when attributing explicit processes to the judgement. At the same time, implicit attribution also led to above chance performance. In Experiment 2, Dienes and Scott (2005) compared participants on instruction and attention. One group of participants were asked to memorize sequences during the training phase, while the others were asked to search for rules. Participants were also divided based on attentional load: half of the participants could devote full attention to the task, while the other half had to generate random numbers in time with an electronic metronome. Results revealed that neither of the manipulations had any effect. Again, these results suggest that both implicit and explicit processes take place during artificial grammar learning. In sum, behavioural studies showed that the AGL task may not be considered a completely implicit task. These results are again in line with the presumptions of the Process Dissociation Procedure (Destrebecqz & Cleeremans, 2001; Jacoby, 1991), suggesting that tasks should not be considered process-pure.

To conclude, behavioural data are mixed. Results of the SRT (Destrebecqz & Cleeremans, 2001) and AGL (Dienes & Scott, 2005) tasks show that both tasks rely on both memory systems. The WP literature lacks direct examination of self-insight as well as the application of PDP. Experimental psychological results however suggest that the task mainly relies on explicit processes (Lagnado et al., 2006; Newell et al., 2007; Price, 2009).The dissimilarity of the results of the three paradigms might be twofold. One possibility is that the

WP task is in fact different from the other two, and healthy adults in fact use explicit/declarative strategies while solving the task. The other possibility is that this inequality is simply due to the differences in methodology: the advanced methods of examining the implicit nature of the SRT and AGL tasks have not yet been applied to the WP task. As a first step to bridge the methodological gap between the WP and the other traditional implicit learning paradigms in this regard, in Study 4 we applied a subjective self-report method of the AGL task, adapted from Dienes and Scott (2005).

5. Synopsis of the presented studies and theses

As explained above, four studies are included in the dissertation. Three out of the four use the WP task, while the fourth study employs the SRT task. Study 1 is a neuropsychological study that applies the WP task to children with Specific Language Impairment. The rationale of the study is to test whether SLI is also associated with procedural learning deficits, as suggested by the Procedural Deficit Hypothesis (Ullman & Pierpont, 2005). Performance of children with SLI is compared to typically developing children and adults, this way we also acquired data on the development of probabilistic categorization measured by the WP task. Study 2 tests whether different inner structures of stimuli and stimulus-outcome relations affect learning performance, or whether learning is identical regardless of the set of stimuli. Transparent and Arbitrary stimulus-outcome associations are used. Also cue-combinations are provided either as standalone images presented simultaneously, or as different features of the same image. Here we also present data from children and adults. Study 3 also tests the effect of different stimulus sets from different domains, but the employed implicit learning paradigm is the Serial Reaction-Time task. The central focus of the study is (1) whether response sequences are acquired in the absence of one-to-one stimulus-response correlation, (2) whether unattended stimulus sequences are acquired, and (3) whether learning of response sequences are affected by unattended systematicity in stimulus appearance. Study 4 tests explicit awareness on the WP task using subjective self-insight measures, and comparing those to measures provided by a control group with no predictive structure in the task.

5.1. Thesis 1: Probabilistic categorization measured by the Weather Prediction task is impaired in Specific Language Impairment

Study 1 tests the Procedural Deficit Hypothesis of Specific Language Impairment (Ullman & Pierpont, 2005). If children with SLI have a procedural learning deficit, then it should be manifest in their performance on a non-linguistic procedural learning task like the WP task. According to the PDH, children with SLI are expected to perform similarly to adults with Parkinson's syndrome: they are expected to show chance-level performance (Knowlton, Mangels et al., 1996). 16 children with SLI (mean age: 11;3, Sd: 1;3), 16 age-matched typically developing children and 16 adults (mean age: 20;5, Sd: 1;7) completed a version of the WP task that was modified to be suitable for testing children. Results showed that children with SLI performed significantly worse than typically developing children. Also, children with SLI only performed above chance level in Block 3. Strategy use was also compared. Results showed that only one third of the clinical group was able to develop any strategies, and the developed strategy was one of the singe strategies. At the same time, the majority of TD children used one of the single strategies, and one third managed to develop the multi-cue strategy. There were only two TD children not showing any signs of strategy use throughout the task. These results show that there is a procedural deficit in Specific Language Impairment. This is in line with the theories suggesting that language is related to procedural learning, and the underlying procedural deficit might explain a number of symptoms observed in SLI – in concert with the Procedural Deficit Hypothesis (Ullman & Pierpont, 2005).

5.2. Thesis 2: Performance on the WP task improves with age

In both Study 1 and Study 2, we tested children on the Weather Prediction task. Our hypothesis was that probabilistic categorization is age-dependent, and performance on the WP task improves with age. Results showed that while in Study 1, a typically developing control group of 11;3 years mean age showed a numerically lower performance than young adults, the performance of the two groups did not differ from each other statistically. In Study 2 however, four groups of 8;6-8;9 year old children (SDs: 0;3-0;6) showed significantly lower performance than young adults of 20;9-21;6 years of age (SDs: 1;3-1;11). Also, children showed more single strategy usage, while the majority of adults relied on multi-cue strategy. These studies suggest a slow development with age, which is in concert with the improvement hypothesis of developmental changes in IL (e.g. Fletcher et al.. 2000).

5.3. Thesis 3: Cue-based cue presentation enhances categorization performance in the WP task

Study 2 tested whether the structure of cue presentation accounts for the differences observed between the Weather Prediction task and the Ice-Cream task (Hopkins et al., 2004; Shohamy et al., 2004; Shohamy et al., 2001). We compared versions of the WP task in which cues either appeared as standalone images (Cue-based conditions), or as features of one combined image (Holistic conditions). Based on previous studies we hypothesized that Cue-based presentation would lead to better performance than if cues are presented as features of the same image. Also, based on previous literature, cue presentation was hypothesized to lower the chance of multi-cue strategy usage. Results showed that performance on the Cue-based conditions were significantly higher than in the case of the Holistic conditions. Results on strategy use revealed that Holistic versus Cue-based stimulus presentation does not affect strategy use. These results suggest that in fact, the holistic presentation of cues lead to a significantly lower level in performance, but not in strategy use.

5.4. Thesis 4: Transparency between cues and outcomes enhances categorization in the early stages of the task

Our major question in connection with the IC task was whether the WP and IC tasks are in fact only differ in the holistic versus cue-based nature of cue-presentation. In Study 2 we also tested whether a transparent link between cues and outcomes affects performance. In the Transparent conditions cues were fragments of a line-drawing, whereas feedback revealed the complete line-drawing; showing why exactly the cues led to the specific outcome. We hypothesized that cue-outcome transparency would enhance early learning performance, as it serves as an anchor point. However, we also expected that this advantage turns into a disadvantage, as observed by earlier studies of Shohamy et al (2001; 2004). Results showed that in the early blocks transparency between cues and outcomes enhances categorization performance; while in the later phases performance was higher for the Arbitrary conditions, where the only link between cues and outcomes was the statistical association provided by feedbacks. Analysing strategy use revealed that participants of the Arbitrary conditions are more likely to develop a multi-cue strategy during the task, whereas participants of the Transparent conditions use the single strategies. These results suggest that cue-outcome transparency enhances participants to learn single-cue-outcome associations, however these associations seem to be stronger, and more difficult to modify. This way, participants become unable to combine cues.

5.5. Thesis 5: Response learning is present in the SRT task even in the absence of one-to-one stimulus-response mapping

Study 3 employed the SRT task, and focused on whether response and perceptual sequential structures are acquired in the absence of the other. In the Response condition of Study 3, participants faced a six block SRT task with no one-to-one mapping of stimuli and responses. Participants faced pictures of four categories. The pictures were those of mammals, furniture, fruits or tools. Participants had to respond with a button press in accordance with the category of the stimulus. Within each category, the specific pictures appeared randomly. Our hypothesis was that response sequence learning takes place even in the absence of a perceptual sequence. Results confirmed our hypothesis, as participants showed a significant sequence specific increase in reaction times.

5.6. Thesis 6: Response-independent perceptual learning does not appear on a short, 6 block SRT task

We showed that response sequence learning may take place in the absence of a correlating stimulus sequence. The question of the Stimulus location condition of Study 3 was whether an unattended stimulus-location sequence can be acquired. In the Stimulus location condition of Study 3, participants faced a 6-block SRT task in which they had to respond to categories of pictures. There were ten pictures within each category, and both appearances of within and between category elements were random. Stimuli could appear in the four corners of the screen. There was a 12-element-long sequence in the location of the stimuli, while participants had to respond to the identity of the stimuli. Our hypothesis was that an unattended perceptual sequence will not affect reaction times (Riedel & Burton, 2006).

Results showed that the removal of the stimulus location sequence caused no RT increase. Results confirm earlier studies (Riedel & Burton, 2006) showing that pure perceptual learning does not take place in the SRT task if the perceptual sequence is not in the response domain.

5.7. Thesis 7: Probabilistic perceptual information impairs response sequence learning

The Response and Stimulus location conditions of Study 3 confirmed that response sequences may be learnt without a one-to-one mapping between stimuli and responses, and that the unattended stimulus location sequence is not subject to learning – at least in such a six-block setting of the SRT task. Some earlier studies suggested that effector-, response- and perceptual types of learning may not necessarily be independent of each other, but might interact (Deroost et al., 2006; Nattkemper & Prinz, 1997). In the Extra condition of Study 3 we tested whether the stimulus domain has any effect on learning the response sequence. Participants faced a response sequence identical to that of the Response condition. The appearance of the stimuli however varied in a probabilistic way: each category had a high frequency location (appearing at this location in 55% of all cases), and three low frequency locations (appearing at these locations in 15-15% of all cases). Results showed that learning both the response sequence and the location frequency was reduced in this simultaneous condition. This suggests that while pure perceptual learning did not appear in the unattended domain (Stimulus location condition), a probabilistic structure in the perceptual domain had a detrimental effect on response sequence learning and vice versa. This is in concert with earlier results suggesting that the different learning domains are in interaction (Deroost et al., 2006; Nattkemper & Prinz, 1997).

5.8. Thesis 8: Explicit knowledge plays an important role in learning the WP task

In Study 4, we investigated whether participants consider their own decisions explicit or implicit. Item-by-item subjective self-insight measures were collected: participants were prompted after each decision to characterize the basis of their response: Guessing, Intuition, 'I think I knew the answer', Memory or Rule-knowledge. As introspection is not necessarily a reliable measure, especially in implicit learning (Frensch & Rünger, 2003), responses of participants in the Experimental condition (facing the original WP task) was compared to those of a Control condition. The Control condition lacked a predictive structure: cues and outcomes were paired randomly. Results showed that Experimental participants responded with significantly more explicit answers than the Control condition. Also, for both groups, implicit performance was at chance level, and the two groups only differed in explicit performance: chance level for Control participants, while a performance almost reaching 90% in the Experimental condition. Examining strategy use, participants using single strategies gave less explicit answers than multi-cue users, but they showed the same level of performance on both implicit and explicit decision. These results contradict both the implicit-first (Knowlton, Mangels et al., 1996) and strategy hypotheses (Gluck et al., 2002), and are in line with the explicit theory (Lagnado et al., 2006).

These results, suggesting that learning on the WP task is explicit, require integration with previous results, as there is a discrepancy with Thesis 1. Thesis 1 states that implicit/procedural learning measured on the WP task is impaired in SLI, while Thesis 8 states that the WP task is explicit in adults. There are four possible resolutions to this controversy. On the one hand, it is a possible that learning on the WP task is implicit in children, and explicit in adults. Potential support for this hypothesis comes from previous studies showing that learning on the SRT task elicits subcortical activation in children and cortical activation in adults (Thomas et al., 2004). Other studies using the SRT task provided how many participants were able to develop explicit knowledge during the learning process, but did not compare the numbers directly across age-groups. Looking at these data across studies, we see that the rate of younger children and the rate of older children developing conscious access are very similar (50% in 10-year-olds, 37% in 7-year-olds Thomas & Nelson, 2001). This argues against the claim that children rely more on implicit learning, while adults are more prone to use explicit strategies. Clearly, more research is needed to disentangle the contribution of implicit/explicit knowledge to performance on these tasks.

The other possibility is that learning on the WP task is explicit in children as well. In this case, the impaired WP performance in SLI requires an alternative explanation. First, there is evidence that declarative memory is also impaired in SLI (Gathercole & Baddeley, 1990), which might account for the impaired declarative WP performance. Previous results also revealed severe executive deficits in SLI (Henry, Messer, & Nash, 2012). As clinical studies confirmed the relationship between the SRT task and executive functions (Jackson et al., 1995), the executive deficit might contribute to deficient learning on the WP task. A last possibility considers the nature of the deficit in SLI. Hsu and Bishop (2010) contrasts procedural and statistical learning deficits, and suggests that the two hypotheses are not identical. According to their view, statistical learning is the acquisition of general statistical information, regardless of the declarative versus procedural nature of the task. That is, statistical learning deficit may be apparent in both declarative and procedural tasks that require the extraction of statistical information. Here again, more research is needed to conclude the origins of the probabilistic categorization impairment observed in SLI.

6. Study 1: Impaired procedural learning in language impairment: Results from probabilistic categorization

Kemény, F., & Lukács, Á. (2010). Impaired procedural learning in language impairment: Results from probabilistic categorization. *Journal of Clinical and Experimental Neuropsychology, 32*(3), 249-258.

JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
2010, 32 (3), 249–258

Psychology Press
Taylor & Francis Group

Impaired procedural learning in language impairment: Results from probabilistic categorization

Ferenc Kemény[1] and Ágnes Lukács[1,2]

[1]Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary
[2]Research Institute of Linguistics, Hungarian Academy of Sciences, Budapest, Hungary

The Weather Prediction (WP) Task is a classical task of probabilistic category learning generally used for examining the dissociation of procedural and declarative memory. The current study focuses on performance of children with language impairment (LI) and compares their performance to that of typically developing (TD) children and adults with the aim of testing the procedural deficit hypothesis of LI (PDH; Ullman & Pierpont, 2005), which states that language impairment is not a specific linguistic phenomenon, but results from the dysfunction of a more general cognitive system: the procedural system. To test the generality of the procedural impairment, we needed a task that is dissimilar from language in that it does not build on sequential information. Children with language impairment show deficient learning on the Weather Prediction Task, which already appears at the early stages of the task. These results, in line with the PDH, point to the deficit of the procedural system in language impairment going beyond the language system. Whether this deficit is selective to the procedural system or is complemented by deficits in the declarative system is the subject of future studies.

Keywords: Implicit learning; Procedural system; Probabilistic categorization; Language impairment; Procedural deficit hypothesis.

Specific language impairment (SLI) is a developmental disorder usually characterized by focal disorder of the linguistic domain. Children with language impairment show a significant delay in language abilities in spite of not having any hearing deficits, neurological disorders, environmental deprivation, or mental retardation that could account for their language problems (e.g., Bishop, 1992; Leonard, 1997; Ullman & Pierpont, 2005). The core deficit concerns grammar, manifests itself as a difficulty in using suffixes and/or specific syntactic structures, and often persists into school years.

Children with a central deficit in the domain of language are often claimed to have *specific* language impairment, a term implying that language or grammar can be selectively impaired in an otherwise intact cognitive system. This view is problematic for several reasons. The disorder is very heterogeneous, and although many attempts have been made to identify specific subgroups (e.g., Aram, Morris, & Hall, 1993; Bishop & Adams, 1992; Vargha-Khadem, Watkins, Alcock, Fletcher, & Passingham, 1995; Whitehurst et al., 1991), a proper system of subcategorization with diagnostic validity is still missing. Another key problem is that specific language impairment in many (probably in most) cases turns out to be not as specific as claimed, and impairments in several nonlinguistic abilities tend to accompany the grammatical deficit, including motor control of oral and fine movements, hypothesis testing and categorization (on both linguistic and nonlinguistic material), mental rotation, sequencing, word retrieval, phonological

We are grateful to all adult participants and to the children and their teachers in the Dr. Nagy László Institute of Special Education in Kőszeg and in the primary school in Kerék Street in Budapest for their cooperation. Special thanks go to Attila Bukniez, István Kemény, and Bence Fadgyas for their help, to Dezső Németh and Szabolcs Kéri for their comments and suggestions, and to Karolina Janacsek and Gergely Éliás for their help in statistics. The research was funded by National Institutes of Health (NIH) R01DC00458 to Larry Leonard and the Országos Tudományos Kutatási Alapprogramok (Hungarian Scientific Research Fund, OTKA) TS 049840 to Csaba Pléh. During the research Ágnes Lukács was a grantee of the Bolyai János Research Scholarship of the Hungarian Academy of Science.

Address correspondence to Ferenc Kemény, Department of Cognitive Sciences, Budapest University of Technology and Economics (BME), Stoczek u. 2, Budapest 1111, Hungary (E-mail: fkemeny@cogsci.bme.hu).

http://www.psypress.com/jcen
DOI: 10.1080/13803390902971131

discrimination, simultaneous execution, and, perhaps most apparently, executive functions (Leonard, 1997; Ullman & Pierpont, 2005).

Besides theories that treat language or grammar as a functionally and anatomically isolated module with the potential for selective impairment, there have been several proposals for more basic mechanisms behind nonlinguistic impairments leading to language impairment. Most well known are claims about deficits in processing rapidly changing auditory input (Tallal & Piercy, 1973) and a decreased capacity of phonological short-term memory (Gathercole & Baddeley, 1990). Most relevant to this paper is a theory that tries to integrate all the linguistic and nonlinguistic deficits observed in SLI into a neurobiological model. The procedural/declarative model (Ullman & Pierpont, 2005) of language claims that there is a clear dissociation within language between the grammar and the lexicon, since they are functions of different memory systems. Grammar is a procedural function, while the lexicon is based on declarative memory. These two memory systems are dissociated not just functionally, but also anatomically, as is explained later in detail. This model has a specific prediction for the nature of language impairment called the procedural deficit hypothesis of specific language impairment (PDH; Ullman & Pierpont, 2005). On this view, language impairment is a result of abnormal development of brain structures underlying the procedural memory system responsible for learning cognitive and motor skills (for sequence and rule learning) and, among them, grammar. Developmental disorders of such a system should result in deficits of skills that rely on procedural learning within both the linguistic and nonlinguistic domains.

The PDH thus suggests that the developmental disorder termed specific language impairment is not specific to language, but is rather a deficit of the more general system of procedural memory. As it has already been stated the procedural system is responsible for the acquisition not only of motor skills, but also of cognitive skills (Knowlton, Mangels, & Squire, 1996a) like probabilistic category learning or rule learning, an example of which is learning and using grammatical rules (Ullman et al., 1997). Based on the model, if children with SLI indeed have a more general procedural deficit, they are expected to show lower performance on different sorts of tasks, linguistic and nonlinguistic, requiring the soundness of the procedural system. Accordingly, earlier research found deficient performance in implicit sequence learning in language impairment (Tomblin, Mainela-Arnold, & Zhang, 2007). The study compared the performance of adolescents with and without language impairment. Both groups showed learning on the serial reaction time (SRT) task, but the learning rate of language-impaired adolescents was significantly slower than that of the control group.

There are also results showing that language-impaired subjects have problems with the processing of rapid sequential information in the auditory domain (Tallal & Piercy, 1973) and also have deficits with other motor functions, like orofacial motor movements, that are best exemplified in speech movement (Vargha-Khadem et al., 1998; Watkins, Dronkers, & Vargha-Khadem, 2002a; Watkins et al., 2002b). At the same time manual praxis, generally, seems to be impaired too, but there is one study suggesting that manual praxis can be intact in certain cases (Vargha-Khadem et al., 1995). The comorbidity of language impairment and poor motor skills (both sequential and nonsequential) seems to be quite high (Hill, 2001).

What seems clear from the literature is that language impairment often involves a wide variety of motor deficits. There are several results indicating the comorbidity of sequence learning and language impairment (LI), and the implicit acquisition of nonsequential motor information also seems to be impaired. This suggests that language impairment is not properly characterized by linguistic features only and cannot be considered a pure linguistic phenomenon.

The implicit–explicit and the declarative–procedural distinction

The nature of implicit learning has been well studied throughout the last few decades, although not many conceptual works focused on the different aspects of multiple memory systems (see Poldrack & Foerde, 2008, for details). Implicit learning is the incidental acquisition of complex information with difficulty in explicitly recollecting the information acquired (Meulemans, van der Linden, & Perruchet, 1998). The concept of implicit learning has mostly been discussed in a multiple memory systems model (Knowlton, Squire, & Gluck, 1994), in which human memory is not homogenous: It has implicit and explicit functions. The former does not involve explicit awareness, while the latter does, so the distinction is based on the mode of recollection of knowledge and its availability to conscious information processing.

Although the definition cited above does not imply it, acquiring information in the traditional implicit learning tasks (like the serial reaction time task, artificial grammar learning, or probabilistic categorization) requires a lot more time and learning trials than acquiring information through explicit learning. This difference can be accounted for by the distinction of Squire, Knowlton, and Musen (1993), on the basis of the representations involved. The declarative system manipulates factual representations that have clear boundaries, while the procedural system uses more dynamic representations, which are like procedures and are mostly acquired in an incremental way.

The dissociation of the declarative and procedural systems (as explained earlier) cannot only be diagnosed by behavioral symptoms: The two systems are anchored in different brain regions (e.g., Knowlton & Squire, 1993; Ullman & Pierpont, 2005; though this view is not general, see, e.g., Voermans et al., 2004; Yin & Knowlton, 2006). While the functioning of the declarative system mainly relies on the activity of the mediotemporal lobe (Knowlton & Squire, 1993), the procedural system is based on more diverse regions, of which the most important are the basal ganglia, the fronto-striatal pathways, the cerebellum, Broca's area, and other areas handling movement, action planning, and motor execution (Ullman & Pierpont, 2005).

It has been known that the procedural system is responsible for the acquisition of motor skills (Knowlton & Squire, 1993), but the same system is responsible for the acquisition of nonmotor cognitive skills including the acquisition of a categories (Knowlton et al., 1996a) or the abstraction of a prototype (Nosofsky, Stanton, & Zaki, 2005). A wide variety of phenomena touching upon procedural learning are traditionally covered by the implicit learning literature: motor sequence learning (Meulemans et al., 1998), perceptual sequence learning (Remillard, 2003), artificial grammar learning (Aslin, Saffran, & Newport, 1999; Reber, 1967), probabilistic category learning (Knowlton et al., 1994), and so on.

There have been very few studies that directly addressed the problem of implicit/procedural learning in language impairment, but these scarce and controversial findings together with results from patients with neurodegenerative disorders of the procedural system (like Parkinson's and Huntington's syndrome) urge research in this field. There are results showing that sequence learning performance is decreased in dyslexia (Stoodley, Harrison, & Stein, 2006) and language impairment (Tomblin et al., 2007), and the same is true of Parkinson's (Siegert, Taylor, Weatherall, & Abernethy, 2006) and Huntington's syndrome (Knopman & Nissen, 1991). Artificial grammar learning (AGL) at the same time does not seem to be impaired in dyslexia (Rüsseler, Gerth, & Munte, 2006); there are studies finding both intact (Witt, Nuhsman, & Deuschl, 2002) and impaired (Smith & McDowall, 2006) AGL performance in Parkinson's syndrome. Research suggests that Huntington's syndrome does not involve impairment in AGL (Knowlton et al., 1996b). Results of one study exploring the link between language/learning disability and performance on the AGL task (Plante, Gómez, & Gerken, 2002) suggest that adults with language/learning disability show low performance on the AGL task, a task that is expected to model sensitivity to word order cues, which is displayed by even one-year-olds (Gómez & Gerken, 1999). This is another suggestion that language impairment is related to implicit learning abilities, but the nature of the deficit needs further investigation in different domains of implicit learning.

The Weather Prediction Task

One of the most frequently used probabilistic category learning tasks is the Weather Prediction (WP) Task (Knowlton et al., 1994). This is a dichotic decision-making task. Participants are presented with an image of a combination of one, two, or three of four cues (which can be either different tarot cards or geometrical shapes). They have to decide whether the pattern they see predicts SUNSHINE or RAIN and have to respond accordingly. As soon as they have made their choice a feedback appears to show whether they were right or wrong. This makes the WP task different from AGL and SRT: The WP task includes feedback-based incremental learning (Shohamy, Myers, Onlaor, & Gluck, 2004). This type of category learning is a cognitive skill involving activity in the procedural system (Knowlton et al., 1996a). There have been hardly any studies examining the relationship between the three traditional implicit learning tasks—that is, the SRT, AGL, and WP tasks. We are only aware of one (Aczel & Gonci, 2005) that found no correlation between performance on the three tasks.

The WP task is also similar to some other tasks measuring implicit learning. The Iowa Gambling Task (IGT) is an implicit learning task to measure risk taking (Bechara, Damasio, Damasio, & Anderson, 1994). The key feature of the task is to compute and clash the short- and long-term benefits and punishments within the task. A central neural structure behind performance is the ventromedial prefrontal cortex, which is proposed to merge the information from the emotional somatic markers supporting decision making. Though there have been studies focusing on the role of the prefrontal cortex in the WP task (Kincses, Antal, Nitsche, Bartfai, & Paulus, 2004), the primary areas active during probabilistic category learning are the basal ganglia (Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004; Knowlton et al., 1996a; Knowlton et al., 1994; Poldrack et al., 2001; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Shohamy et al., 2004). This difference might be due to the fact that a mistake in the WP task does not involve strong emotions, while emotions play a central role in the IG task; the IGT has been widely used as a research tool for studying the somatic marker hypothesis (Bechara et al., 1994; Dunn, Dalgleish, & Lawrence, 2006).

In the original version of the task (Knowlton et al., 1994) there are altogether four cues, and each cue has its own predictive value. Cue 1 predicts SUNSHINE in 77% of all cases, Cue 2 in 58%, Cue 3 in 42% and Cue 4 in 23%. The task consists of several blocks of 50 trials. Learning performance is measured by the difference between accuracy on the first and later blocks. People are expected to achieve a performance above 70% correct on this task (Gluck, Shohamy, & Myers, 2002; Knowlton et al., 1994; Knowlton et al., 1996b). According to the Rescorla–Wagner law (Knowlton et al., 1994; Rescorla & Wagner, 1972) participants are expected to learn the response to those cues first that have the best predictive value—that is, the cues that determine the outcome most. In the case of the original task these are the cues with 77% and 23% probability values. After the individual cues are learnt people are expected to integrate them into a whole pattern and (implicitly) start calculating the predictive values and give the answer according to the result.

The Weather Prediction Task proved to be a useful way of testing implicit learning in different adult neuropsychological disorders. In their study Knowlton and her colleagues (1994) compared the performance of amnesic patients with an age-matched control group. The etiologies of their subject were heterogeneous: The group included patients with Korsakoff syndrome, bilateral brain infarct, bilateral traumatic brain injury, and anoxia. The only feature that was shared by all patients was that they had amnesia associated with a mediotemporal lobe (MTL) impairment. During the first 50 trials there was no difference between the amnesic and control groups, both showing learning. Significant differences emerged in the later blocks only: The control group's performance was significantly better than that of the clinical group. Under this assumption, one would expect to see MTL activity during the second half of the task, a prediction that was borne out by Poldrack and his colleagues' positron emission tomography (PET) studies (Poldrack et al., 2001; Poldrack et al., 1999), which showed that during the early stages of the feedback-based Weather Prediction Task the striate, the caudate nucleus (NC), and the cortico-NC pathways are active while the activity of the MTL only increases in the second half of the task, probably as a result of verbalization.

In a subsequent study Knowlton and her colleagues (1996a) found a double dissociation between the two systems. Their results showed that people with Parkinson's syndrome showed significantly lower performance than amnesic and control participants in the first 50 trials. Their performance did not differ from chance at the beginning of the task, but later they started to improve. According to Knowlton and colleagues this is due to the fact that learning in the early stages relies on activation of the procedural system. At later stages of the task procedural activation decreases, while declarative processes emerge. With the emergence of declarative strategies, the rate of the procedural load decreases, and patients with impaired procedural functioning start to show better performance on the task—that is, Parkinson's patients start to improve. Due to the same reason the performance of amnesic patients declines after performing well on the first blocks. The theory of the early activation of the procedural system and the later declarative functioning has been tested in a PET study (Poldrack et al., 2001; Poldrack et al., 1999), which confirmed Knowlton and colleagues' (1996; 1994) results.

Possible strategies for solving the Weather Prediction Task

Unconscious strategies for solving the Weather Prediction Task are essential to understand the diversity of performance within the typical population and also to explain differences between clinical and control groups. There have been several attempts to identify the behavioral difference associated with the switch from procedural to declarative functioning. Since it is one of the key points of the present research, it is necessary to go into detail about these strategies. Gluck and colleagues (2002) were searching for an answer as to whether there are differences in robust strategies that are responsible for this variance. They identified three different strategies:

- Multicue strategy
- Singleton strategy
- One-cue strategy.

The multicue strategy is the most optimal way of solving the Weather Prediction Task. It is basically following all four cues and averaging them before decision. The use of this strategy throughout the task leads to more than 80% correct answers. This is the strategy we would probably use if the task was an explicit mathematical task—that is, if we are facing the combination of two cues, one with 23%, the other with 58% predictive strength for sunshine, we calculate the average (40.5%), decide whether it is above 50% (SUNSHINE) or below that (RAIN) and give an answer accordingly.

In the singleton strategy participants answer consistently if cues appear on their own, and they answer randomly when cues appear in combination. As a result, cue-based answers are given only if any of the cues appears alone—that is, seeing Cue 1 or Cue 2 alone, participants predict SUNSHINE, but if only Cue 3 or Cue 4 appears on the screen, they predict RAIN. Any other patterns (all combinations of cues) lead to random responses, resulting on these trials in an average performance of 50%. The optimal use of this strategy leads to about 70% correct performance.

The one-cue strategy is the consistent use of one cue as a predictor of an outcome. It means that, for example, the participant consistently answers SUNSHINE when Cue 1 is present (either alone or in a combination), but if it is not then his answers do not differ from chance. This strategy leads to a performance just above 60%.

Preliminary research with 30 participants showed that 27 used the singleton strategy, and 3 used the one-cue strategy on the first 50 trials. Later on more and more of them started to use the multicue strategy (Gluck et al., 2002).[1] Gluck and colleagues argued that the improvement of the participants' performance is not only quantitative, a proposal that is supported by PET studies (Poldrack et al., 2001; Poldrack et al., 1999) showing that different brain regions are active in the different stages of the task. The shift in activation patterns from the procedural to the declarative system can also be observed and distinguished behaviorally—that is, based on strategy use.

It has to be pointed out that from the perspective of the procedural–declarative dissociation the one-cue and the singleton strategies do not differ. According to Gluck and colleagues (2002) this behavioral difference is almost the same as the difference between the one-cue strategies—that is, they are different manifestations of the same learning process. The important difference lies in the distinction of the single (one-cue and singleton) and multicue strategies. Single strategies always rely on the appearance of a single cue (which is set by either the specifics of the cue, i.e., one can use a strategy relying on Cue 1 OR Cue 2 OR Cue 3 OR Cue 4 alone, or the number of the cue, i.e., any of the cues ALONE). Multicue strategies on the other hand involve the use of more cues and the comparison of cue combinations along their predictive value: Single strategies (i.e., one-cue and

[1]Gluck et al. (2002) and studies that followed used the predictive values of 80%, 60%, 40%, and 20% in the WP task.

singleton strategies) rely on the procedural system, while the multicue strategy relies on the declarative system.

To test the strategy-use hypothesis, Hopkins and colleagues (2004) repeated Knowlton et al.'s (1994) study in a later experiment with a more homogenous clinical group with only patients with specific bilateral MTL damage (Hopkins et al., 2004). The overall performance of the amnesic group was significantly lower than that of the control group. As expected, the difference between the clinical and control groups was not seen during the first 50 trials and only appeared in later phases of the task. In later phases of learning, control participants switched to the multicue strategy, but amnesic patients were unable to do so. While the multicue strategy seemed to be difficult for the hypoxic patients, all their performance patterns were fit properly to the single strategies. This confirms the hypothesis that the use of the multicue strategy is probably hippocampus dependent, while the single strategies can rely on the procedural system (Knowlton et al., 1994).

Shohamy and colleagues (2004) repeated another experiment of Knowlton et al. (1996a). They compared patients with mild Parkinson's syndrome with a control group. Both groups showed learning during the first 50 trials, and there were no group differences. Control participants—consistent with the earlier results—gradually switched from single-cue strategies to multicue strategy, while this change did not appear in Parkinson's patients. Shohamy and colleagues (2004) suggested that the discrepancy between their and Knowlton and colleagues' (1996a) result is due to the differences in the severity of the state of patients with Parkinson's disease participating in the two studies. These results are not in concert with the procedural–declarative model of probabilistic category learning, a discrepancy that we address in more detail in the Discussion section.

A critical review of literature on the Weather Prediction Task suggests that the introductory parts of the task rely on single strategies with procedural activity, while the later phases build on multicue strategies associated with activity in the declarative system. Deficits of the procedural system (like Parkinson's syndrome) are expected to be associated with a decreased overall performance, especially in the early phases of the task, which is manifested by an inability to use single strategies. As the declarative system starts to be active—between approximately the second and third blocks of 50 trials—improvement should be observed. The reason for this could be a compensatory reliance on declarative strategies for those who were unable to use single strategies due to procedural malfunction. If the procedural deficit hypothesis of language impairment (Ullman & Pierpont, 2005) is correct, we expect children with language impairment to show a performance pattern similar to that of Parkinson's patients and to show (a) no learning in the first 50 trials, and/or (b) learning in the third block of 50 trials, and/or (c) an inability to switch from single to multicue strategies. To test this hypothesis, we compared WP performance of a group of children with language impairment to typically developing children matched on chronological age. This study is the first to examine probabilistic category learning in children with LI (and also typically developing, TD, children). As we used a modified version of the original WP task to suit children, we first tested adults to set the baseline of learning and to make comparisons to other studies possible.

METHOD

Participants

Adults

A total of 16 adults (5 female, 11 male) participated in the study. All of them were recruited at the Budapest University of Technology and Economics and participated voluntarily in the study for credits. Their mean age was 20;5 with a standard deviation of 1;7. They were informed about the purpose of the research after the study. All participants provided written informed consent, in accordance with the principles set out in the Declaration of Helsinki and the stipulations of the local Institutional Review Board.

Children with language impairment

A total of 16 children (5 girls, 11 boys) were selected for the language-impaired group. Their mean age was 11;3 with a standard deviation of 1;3. All language-impaired children were students of an institute of special education in Kőszeg. All of these children met the criteria for LI. Each child scored above 85 on the Raven Colored Progressive Matrices (Raven, Court, & Raven, 1987), a measure of nonverbal intelligence. All children passed a hearing screening, and no child had a history of neurological impairment. Each child scored at least 1.5 standard deviations below age norms on at least two of four language tests administered. These four tests included two receptive tests and two expressive tests. The receptive tests were the Hungarian versions of the Peabody Picture Vocabulary Test (PPVT; Csányi, 1974; Dunn et al., 2006) and the Test for Reception of Grammar (TROG; Bishop, 1983). The expressive tests were the Hungarian Sentence Repetition Test (Magyar Mondatutánmondási Teszt, MAMUT; Kas & Lukács, 2007) and a nonword repetition test (Racsmány, Lukács, Németh, & Pléh, 2005).

Performance of the LI group was compared to that of a control group of 16 children matched individually on chronological age to children in the LI group (4 girls, 12 boys; all from a primary school in Budapest). Their mean age was 11;3 (*SD* 1;2). All children in the LI and in the TD groups were tested with the informed consent of their parents, in accordance with the principles set out in the Declaration of Helsinki and the stipulations of the local Institutional Review Board.

Procedure

Participants were presented with a special version of the Weather Prediction Task. In this version the procedure was the same as that explained earlier: Participants were shown a pattern of one, two, or three cues (in this case out of four simple geometric forms), and for each combination they had to tell whether it predicted sunshine or rain. Immediate feedback was given showing whether they were right or wrong. Each of the four cues had its own probability of determining SUNSHINE: 90%, 70%, 30%, and 10%. These probabilities were adjusted to make the task easier for children, as preliminary studies with the original weights did not show any learning in this age group. For the same purpose, we deviated from the routine of earlier studies in presenting participants with only three blocks of 50 trials, since we tried to lower the requirements for children. After every 25 trials there was a recess-slide, which was on as long as one of the buttons was pressed. Number and percentage of correct answers were measured for each block, and learning was signaled by improvement across the three blocks.

The outcome could be predicted in 84% based on the four cues that were present either alone or in combination. In 84% of the trials the expected answer and the given feedback are the same. In the remaining 16% the expected answer is not the same as the feedback given. In cases like 0110 in which Cue 2 and Cue 3 appear the average predictive value is 50%—that is, participants will probably respond by chance on these trials. There were trials where the feedback was not consistent with the expected answer. If a cue has 90% predictive value for sunshine that means that 1 out of 10 trials will have a feedback contradicting the usual outcome (i.e., RAIN instead of SUNSHINE). Since we did not want to investigate pure one-to-one associative learning, even those trials that presented Cue 1 alone, without any other cues, had to have some cases in which the feedback was RAIN instead of SUNSHINE. Altogether 84% of the feedbacks were predictable from the cues presented.

Strategy fitting

Three different strategies (multicue, singleton, and one-cue) were fit with regression to the performance of each participant. A strategy was assigned to a block of a participant if the average deviation from the expected performance based on that specific strategy did not exceed 0.1, an arbitrary criterion set by earlier literature (Gluck et al., 2002). For trials that had a predictive value (probabilities differing from 50%: all trials except those with Cue 2 and Cue 3 or Cue 1 and Cue 4—two opposing cues—present at the same time without any other cues in the multicue strategy) the square of the difference was calculated (1 or 0). For trials without predictive probabilities the sum of answers was drawn from the half of the trials (since we expect 50% accuracy on these trials), and this value was divided by the number of trials (to get the measure of difference/trial) and raised to the power of two (square of difference by trial). After this, the differences for predictive and nonpredictive trials were summed up and were divided by the complete number of trials (by 50 in each block). This way we were able to get an average deviance per trial. This value was calculated for all strategies—that is, the multicue strategy, the singleton strategy, and the 4 one-cue strategies. A participant was credited with using the multicue strategy if the value of average deviance did not exceed 0.1. If it did, the single strategies were fit, and the best fit strategy was assigned unless average deviance exceeded 0.1 from each the single strategy, in which case there was no strategy assigned to the block.

Note that earlier studies (Hopkins et al., 2004; Shohamy et al., 2004) have shown that from a functional point of view the difference between the singleton and the one-cue strategy is irrelevant (just as the differences between each one-cue strategy are), as these two strategies both build on the same underlying anatomical functioning—that is, both singleton and one-cue strategies rely on the declarative system, while the multicue strategy uses the procedural system.

RESULTS

Accuracy

A 3 (group) × 3 (blocks) repeated measures analysis of variance (ANOVA) was employed with group as between-subjects variable and blocks as within-subject variable to see whether the overall number of correct answers based on the predictive values of all four cues differs between blocks and groups. There was a significant main effect of group, $F(2, 46) = 15.584$, $\eta^2 = .409$, $p < .001$, showing that there is a significant difference between the groups with adults giving the most correct answers, followed by TD children, and children with LI giving the least. Post hoc least significant difference (LSD) tests revealed that the LI group differed from both the TD and adult groups significantly ($p < .001$ in both comparisons, see also Figure 1, Table 1), while performance in the two typically developing groups did not differ ($p = .109$). There was a significant main effect of block, $F(2, 46) = 7.361$, $\eta^2 = .141$, $p < .001$, showing that participants' performance improved with time. The Group × Block interaction did not appear to be significant, $F(4, 46) = 0.882$, $\eta^2 = .038$, $p = .478$ (Figure 2).

A one-way multivariate ANOVA was employed for three dependent variables (performance on Block 1, performance on Block 2, performance on Block 3) and one between-subjects variable (group). The ANOVA revealed a significant main effect of group on Block 1, $F(2) = 5.824$, $\eta^2 = .206$, $p < .01$. Post hoc LSD tests

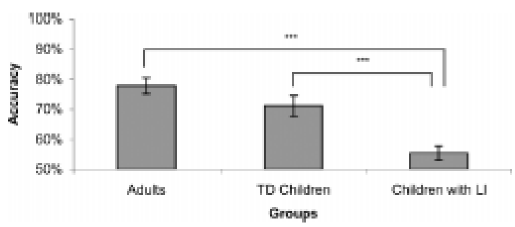


Figure 1. Overall performance on the PCL task by groups (% correct).

TABLE 1
Performance by groups: Descriptive statistics

	Block 1		*Block 2*		*Block 3*		*Total*	
Groups	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*	*Mean*	*SE*
Adults	71.09	3.84	79.56	3.07	83.07	3.46	77.91	2.64
TD children	66.02	4.58	73.31	3.94	74.09	3.68	71.14	3.53
Children with LI	53.91	2.11	54.43	3.70	57.94	3.94	55.43	2.52

Note. In percentages. TD = typically developing. LI = language impairment.

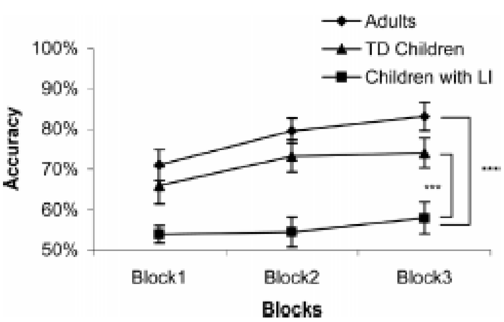


Figure 2. Performance of the three groups on the PLC task by blocks (% correct).

revealed a significant difference between LI and TD children ($p < .05$) and between LI children and adults ($p < .01$), but the difference between the two control groups was not significant ($p = .332$).

The ANOVA revealed a significant main effect of group on Block 2, $F(2) = 13.303$, $\eta^2 = .372$, $p < .001$. Post hoc LSD tests revealed a significant difference between LI and TD children ($p < .01$) and between LI children and adults ($p < .001$), but the difference between the two control groups appeared to be not significant ($p = .224$).

The ANOVA revealed a significant main effect of group on Block 3, $F(2) = 11.849$, $\eta^2 = .345$, $p < .001$. Post hoc LSD tests revealed a significant difference between LI and TD children ($p < .01$) and between LI children and adults ($p < .001$), and the difference between the two control groups was approaching significance ($p = .093$).

A striking difference between the LI and control groups is that the LI group does not show any evidence of improvement in the first two blocks; they seem to learn only in the third one, when the performance rises up to 57.94% (a performance level that is significantly better than chance; $t = 4.456$, $p < .001$). At the same time both control groups show greater improvement between the first and second blocks. Pairwise comparisons (LSD tests) at the .05 level confirm this difference for the control groups, but the difference is not significant between these blocks for the LI group (Figure 2).

Best fit strategies

Of the pool of 48 participants, altogether 15 switched to multicue strategy. Of the 16 adults, 10 arrived at the multicue strategy, and 5 of the 16 control children also managed to do so, but no children with LI applied this strategy at any point in solving the Weather Prediction Task. Among those who did not switch to multicue strategy, there were 6 adults, 9 control children, and 5 children with LI who used one of the single strategies (i.e., either singleton or one-cue strategy). A total of 2 control children and 11 children with LI failed to show any sign of strategy use (the average deviation exceeded 0.1). A two-way chi-square comparison of strategies by groups revealed a significant difference in strategy use between the groups, $\chi^2(2, N = 48) = 27.146$, $p < .001$ (see Table 2).

DISCUSSION

Results show that children with LI perform significantly worse on the Weather Prediction Task than either adults or typically developing children matched on chronological age. It is also clear that children with LI are less able to rely on strategies described by Gluck and colleagues (2002), and even those who seem to develop a one-cue strategy seem to be unable to switch to the more effective multicue strategy. The result that only 5 of the 16 children of the LI group showed any sign of strategy use suggests that children with language impairment have a more fundamental problem with making use of even the simplest single-cue strategies. Their results fall behind age-level expectancies, since typically developing children matched on chronological age learn and perform significantly better. TD children learn at a level closer to, but still significantly below, adult performance; they are less likely to switch to multicue strategy, and they use one of the single strategies instead.

TABLE 2
Strategy use by groups

	Adults	*TD children*	*Children with LI*
Switch to multicue	10	5	—
Single strategies	6	9	5
No strategies	—	2	11

Note. TD = typically developing. LI = language impairment. The distribution of strategy use by the three groups differed significantly. The distribution in the TD and LI groups also differed significantly, while there was no difference between TD children and adults.

The two control groups perform at a similar level and show the same rate of development across blocks on the Weather Prediction Task. The difference between the overall performance of adults and children seems to differ by only 5%. As can be seen in Figure 2, the difference between the two typically developing groups already appears in the first two blocks (although this difference might grow, since in the third block the performance of the two control groups was approaching significance). As we have explained earlier, early learning is linked to the procedural system, and the difference between children and adults is probably explained by late maturation of the fronto-striatal pathways (Casey, 2005; Thomas et al., 2004).

Our results on the performance of adults are not consistent with that of the Gluck et al. (2002) study. In the present study, 62.5% of adult participants switched to multicue strategy (10 out of 16), whereas in Gluck et al.'s study around 40% of all participants managed to switch to multicue by the 4th block (Figure 7, p. 416). This discrepancy can probably be due to the fact that in our result the predictive values of each cue were higher than that of the Gluck et al. (2002) study.

There is a great difference in accuracy, learning rate, and strategy use between control children and children with LI. Overall performance of LI children lags behind that of control children, and the difference from chance in their performance only reached significance by the end of the third block (and even then, it did not reach 60%). The performance of control children in the first 50 trials is already approximately the level of performance of LI children in the last 50 trials, which shows a clear deficit in the task performance already in the early phases, indicative of impairments in procedural functioning. The difference in accuracy is not the only marker of impaired learning in LI. Strategy analysis shows that children with LI are unable make use of any of the three possible strategies, and even best fit strategies seem to be less efficient and lead to lower performance for them. Besides failing to switch to the multicue strategy, children with LI are also less able to use the single strategies that would be the basis of switching.

As earlier behavioral (Hopkins et al., 2004; Knowlton et al., 1996a; Knowlton et al., 1994) and imaging (Poldrack et al., 2001; Poldrack et al., 1999) results show, the early stages of learning on the Weather Prediction Task, which include the singleton and one-cue strategies, seem to rely on the procedural system, while the later, representational stages of probabilistic categorization require hippocampal activity (Knowlton et al., 1994; Poldrack et al., 1999). This is shown by phenomena that the deficit of the declarative system in amnesia leads to an early learning using singleton and one-cue strategies and later deficits in the multicue strategy (Hopkins et al., 2004; Knowlton et al., 1994). Deficits of the procedural system—that is, Parkinson's syndrome, Huntington's syndrome—show impaired learning at the early stages, and, lacking single-cue strategies, patients do not get the opportunity of switching to multicue strategy.

Results of children with language impairment are similar to results of Parkinson's patients by Knowlton and colleagues (1996a) and Shohamy et al. (2004) for the early phases of learning. Language-impaired children—just like Parkinson's patients—show impaired performance on the Weather Prediction Task, already in its early stages of the first 50 trials. Unlike Parkinson's patients, though, they show very little learning and little evidence of strategy use through three blocks of 50 items. A major difference is that in Shohamy et al.'s study, Parkinson's patients were reported to be able to use single strategies, which were concluded to require declarative memory, and the authors interpreted results for patients with Parkinson's syndrome as an inability to switch to multicue strategy. In the light of other results from the literature, this conclusion is problematic (Hopkins et al., 2004; Knowlton et al., 1996a; Knowlton et al., 1994).

A reason for this inconsistency might be Shohamy and colleagues' (2004) focus on strategy analysis. Strategy analysis requires an arbitrary criterion to be set. Participants who reach the criterion are assigned the strategy. A total of 2 participants who were not assigned any strategies were excluded from the statistical analysis. Since there were only 12 patients with Parkinson's syndrome in the study, the exclusion of 2 patients is enough to lead to the incorrect assumption that patients with Parkinson's syndrome are impaired on later phases of the task, as almost 15% of the clinical group was unable to even start to solve the task. This is especially important considering the fact that the study tested patients with mild Parkinson's syndrome who were relatively well functioning.

Including those patients who did not fit any strategies would probably show a different picture, suggesting that Parkinson's patients are impaired on the early stages of the WP task, which would be consistent with earlier results. As the review of behavioral (Hopkins et al., 2004; Knowlton et al., 1996a; Knowlton et al., 1994) and PET (Poldrack et al., 2001; Poldrack et al., 1999) studies suggested, early phases of the WP task rely on the procedural system, while during the later phases activation shifts towards the declarative system. If Parkinson's patients are less able to fit single strategies, then they are impaired on the earlier phases, which is the procedural part of the task. Amnesic patients use single strategies properly (Hopkins et al., 2004), since these strategies do not require proper functioning of the declarative system. Taken together, these results imply that prototype learning relies on the procedural system, while specifying the individual characteristics and combining the predictive values is a declarative, hippocampus-dependent function. In the light of earlier research, our results indicate that children with LI show a procedural deficit: They do not learn on the initial stages of the WP task, and they also show a more severe inability to use strategies.

If, following Shohamy et al.'s (2004) procedure, children with LI who did not fit any strategies were excluded from our study, we could argue that the focal problem of children with LI on the WP task is that they are not able to switch to the use of the multicue strategy. In our case, it would imply the exclusion of 69% of the LI group. The exclusion of no-fits would lead us to the false conclusion that the central problem of children with LI in solving the Weather Prediction Task is a declarative deficit, manifest in the failure of switching to the multicue strategy.

The finding that children with LI show very little learning and strategy use throughout the task supports Ullman and Pierpont's (2005) hypothesis that children with language impairment have a more general cognitive problem in procedural learning going beyond language and argues against the specificity of language problems in SLI. Whether this deficit is selective to the procedural system, or is complemented by deficits in the declarative system (as suggested by severe vocabulary problems in LI), is the subject of future studies.

The procedural deficit hypothesis (Ullman & Pierpont, 2005) puts language impairment into a broader context by predicting impairments in nonlinguistic abilities, locating these deficits in the domain of procedural learning. Results of the present study are in concert with the PDH and, as such, draw attention to the importance of focusing on impairments outside the language domain as well, both in diagnosis and in training of LI.

Our results also raise a number of other questions suggesting possible lines of further research not only for language impairment, but also for probabilistic category learning. At what age does the difference in performance level and strategy use between children and adults disappear? Would this qualitative difference also disappear in LI children with age at a slower rate than in typical development, or is the qualitative difference maintained? What makes the Weather Prediction Task so much different from artificial grammar learning (Aslin et al., 1999) where children at very early ages can properly differentiate between legitimate and illegitimate sequences solely based on statistical properties of the stimuli? Would we get the same difference between LI and control children on the AGL or SRT tasks? These are yet open research questions with important consequences for both LI and implicit learning.

Original manuscript received 3 February 2009
Revised manuscript accepted 14 April 2009
First published online 22 June 2009

REFERENCES

Aczél, B., & Gönci, D. (2005). *Unity or fractionality of implicit learning: A methodological aspect*. Paper presented at CogSci2005—27th Annual Meeting of the Cognitive Science Society, Stresa, Italy, 21–23 July, 2005.

Aram, D., Morris, R., & Hall, N. (1993). Clinical and research congruence in identifying children with specific language impairment. *Journal of Speech and Hearing Research, 36*, 580–591.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1999). Statistical learning in linguistic and nonlinguistic domains. In B. MacWhinney (Ed.), *Emergence of language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7–15.

Bishop, D. V. M. (1983). *Test For Reception of Grammar*. Manchester, UK: Medical Research Council.

Bishop, D. V. M. (1992). The underlying nature of specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*, 3–66.

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment—literal and inferential meaning. *Journal of Speech and Hearing Research, 35*, 119–129.

Casey, B. J. (2005). Frontostriatal and frontocerebellar circuitry underlying cognitive control. In M. I. Posner, U. Mayr, E. Awh, & S. W. Keele (Eds.), *Developing individuality in the human brain: A tribute to Michael I. Posner*. Washington, DC: American Psychological Association.

Csányi, I. (1974). *Peabody Szókincs-teszt* [Peabody Vocabulary Test]. Budapest, Hungary: Bárczi Gusztáv Gyógypedagógiai Főiskola.

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience and Biobehavioral Reviews, 30*, 239–271.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection. *Journal of Memory and Language, 29*, 336–360.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*, 408–418.

Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

Hill, E. L. (2001). Non-specific nature of specific language impairment: A review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders, 36*, 149–171.

Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia, 42*, 524–535.

Kas, B., & Lukács, Á. (2007). *Magyar Mondatutánmondási Teszt* [Hungarian Sentence-Repetition Test]. Manuscript in preparation.

Kincses, T. Z., Antal, A., Nitsche, M. A., Bartfai, O., & Paulus, W. (2004). Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the human. *Neuropsychologia, 42*, 113–117.

Knopman, D., & Nissen, M. J. (1991). Procedural learning is impaired in Huntington's disease: Evidence from the serial reaction time task. *Neuropsychologia, 29*, 245–254.

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

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

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

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

Leonard, L. B. (1997). *Children with specific language impairment*. Cambridge, MA: The MIT Press.

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

Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition, 33*, 1256–1271.

Plante, E., Gómez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders, 35*, 453–462.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546–550.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews, 32*, 197–205.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*, 564–574.
Racsmány, M., Lukács, Á., Németh, D., & Pléh, C. (2005). A verbális munkamemória magyar nyelvű vizsgálóeljárásai [Hungarian diagnostic tools of verbal working memory functions]. *Magyar Pszichológiai Szemle, 60*, 479–505.
Raven, J., Court, J., & Raven, J. (1987). *Raven's Progressive Matrices and Raven's Colored Matrices.* London: H. K. Lewis.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.
Remillard, G. (2003). Pure perceptual-based sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 581–597.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. P. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Century Croft.
Rüsseler, J., Gerth, I., & Munte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology, 28*, 808–827.
Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*, 676–686.
Siegert, R. J., Taylor, K. D., Weatherall, M., & Abernethy, D. A. (2006). Is implicit sequence learning impaired in Parkinson's disease? A meta-analysis. *Neuropsychology, 20*, 490–495.
Smith, J. G., & McDowall, J. (2006). When artificial grammar acquisition in Parkinson's disease is impaired: The case of learning via trial-by-trial feedback. *Brain Research, 1067*, 216–228.
Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology, 44*, 453–495.
Stoodley, C. J., Harrison, E. P. D., & Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia, 44*, 795–798.
Tallal, P., & Piercy, M. (1973). Defects on non-verbal auditory perception in children with developmental dysphasia. *Nature, 241*, 468–469.
Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y. H., et al. (2004). Evidence of developmental differences in implicit sequence learning: An fMRI study of children and adults. *Journal of Cognitive Neuroscience, 16*, 1339–1351.
Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development, 3*, 269–293.
Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., et al. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*, 266–276.
Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41*, 399–433.
Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America, 92*, 930–933.
Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., et al. (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 12695–12700.
Voermans, N. C., Petersson, K. M., Daudey, L., Weber, B., van Spaendonck, K. P., Kremer, H. P. H., et al. (2004). Interaction between the human hippocampus and the caudate nucleus during route recognition. *Neuron, 43*, 427–435.
Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002a). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain, 125*, 452–464.
Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., et al. (2002b). MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain, 125*, 465–478.
Whitehurst, G. J., Arnold, D. S., Smith, M., Fischel, J. E., Lonigan, C. J., & Valdezmenchaca, M. C. (1991). Family history in developmental expressive language delay. *Journal of Speech and Hearing Research, 34*, 1150–1157.
Witt, K., Nuhsman, A., & Deuschl, G. (2002). Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. *Neuropsychologia, 40*, 1534–1540.
Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience, 7*, 464–476.

7. Study 2: The effects of feature analysis and transparency in probabilistic category learning in adults and children

Kemény, F., & Lukács, Á. (2009). The effects of feature analysis and transparency in probabilistic category learning in adults and children. *Learning & Perception, 1*(2), 199-213.

Learning & Perception 1 (2009) 2, pp. 199–213
DOI: 10.1556/LP.1.2009.2.9

THE EFFECTS OF FEATURE ANALYSIS, TRANSPARENCY IN PROBABILISTIC CATEGORY LEARNING IN ADULTS AND CHILDREN

FERENC KEMÉNY[1] * and ÁGNES LUKÁCS[2]

[1]Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary
[2]Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary; Department of Experimental Linguistics, Research Institute of Linguistics, Hungarian Academy of Sciences, Budapest, Hungary

The Weather Prediction (WP) Task is one of the most widely used tasks in probabilistic category learning research. Earlier studies mainly tested adults on similarly structured but different versions of the WP task, mostly without specific focus given to the differences between these tasks. The current paper focuses on the effects of stimulus organization on learning, manipulating two variables: 1) transparency of cues and outcomes and 2) combination of cues into features of a single image vs presenting them as distinct cues. Results show that different variables affect probabilistic categorization differently; cue-based (as opposed to holistic) presentation leads to better performance, and transparency also helps categorization in the early phases of the task, but this advantage turns into a disadvantage later. In general, adults' probabilistic categorization abilities surpass children's categorization, but the pattern of performance was similar regardless of age.

Keywords: probabilistic category learning, feature-based analysis, cue-based analysis, transparency, development

* Correspondong author; Department of Cognitive Sciences, Budapest University of Technology and Economics (BME), Stoczek u. 2, H-1111 Budapest, Hungary. Phone: +36 1 463 1072; E-mail: fkemeny@cogsci.bme.hu

THE EFFECTS OF FEATURE ANALYSIS, TRANSPARENCY AND AGE IN PROBABILISTIC CATEGORY LEARNING

Categorization is a task we perform numberless times each day. We maximalize similarities and minimalize differences between tokens at the required level of detail. These tokens may be objects or events. Sometimes there may be huge differences between members of the same category, while sometimes it is necessary to differentiate between two very similar entities: say the Chihuahua and the St. Bernard Dog are both dogs, but a slow-worm is a lizard and not a snake (as most people would think).

The example of the slow worm clearly illustrates one of the most important issues in categorization. Most studies of categorization deal with cases where features determine category membership, yet in most cases features are not completely predictive. Probabilistic categorization studies try to model these cases. In probabilistic categorization the link between cues and categories is not deterministic, but can be characterized by a predictive value anywhere between 0 and 1, i.e. the rate in which a given cue signals membership of a specific category (Meeter et al., 2008).

The current paper focuses on probabilistic category learning, and a specific task that measures probabilistic categorization. Our aim is to test the effect of different factors (feature analysis and transparency) and possible developmental differences by comparing adults' and children's performance in a probabilistic category learning task. The extensive research on probabilistic category learning has not so far devoted special attention to the effects of these specific factors on categorization. No studies addressed the question whether categorization is enhanced or reduced by a transparent link between categories and its members. There has been some research on categorization in Parkinson's syndrome testing the effect of feature combination, but without a clear definition of cue-combination (Shohamy et al., 2001). Apart from one of our former studies (Kemény and Lukács, 2010), earlier research also neglected the developmental aspect of probabilistic category learning. In the following, we will first introduce the Weather Prediction task (WP), the task most widely used for testing probabilistic categorization, then focus will be shifted to earlier studies varying combination and transparency in the WP task, and finally the developmental study of the WP task will be discussed, as an introduction to the current study. In the main part of the paper, we present a study with the aim of uncovering more specific mechanisms and developmental changes in probabilistic category learning.

THE WEATHER PREDICTION TASK

One of the most well-known probabilistic categorization tasks is the Weather Prediction (WP) task (Gluck and Bower, 1988; Knowlton, Squire and Gluck, 1994). In the task participants face one, two or three out of four different cues. After seeing the cues participants have to decide whether these cues predict SUNSHINE or RAIN. As soon as the participants make their choice, the correct outcome is revealed – this way feedback is provided whether their choice was right or wrong. Stimuli are presented in several blocks of 50 trials. Throughout a block of 50 trials the predictive value of each cue is set. The predictive value is the ratio of the cases when the given cue is associated with SUNSHINE and all appearances of the given cue. Chance level is 50%, values above chance level predict SUNSHINE, values below chance level predict RAIN. Categorization performance is characterized by the rate of correct answers. An answer is considered correct if it corresponds to the prediction based on the average predictive value of all cues present in the given combination, i.e. if a 70% and a 15% cue are present in a combination, then the expected correct answer is RAIN, since the average (42.5%) is below chance level (50%). Measuring accuracy gives an approximation of the extent to which participants acquired the given categorization. At the same time it gives no information on what participants base their predictions on and how categorization takes place (Meeter et al., 2008).

Another method of measuring categorization performance is to identify the strategy used throughout the task. Earlier studies (Gluck, Shohamy and Myers, 2002) identified three types of strategies: the multi-cue strategy, the singleton strategy and the one-cue strategy. The multi-cue strategy is the most optimal strategy. It is basically equivalent to the accuracy measure described above, i.e. participants using the multi-cue strategy are expected to answer based on the average of the predictive values of the cues that appear at the same time. The singleton strategy is somewhat less optimal. A singleton strategy user gives consistent answers when only one cue is present at the time, but if there are more cues on the screen, a random answer is given. Using the one-cue strategy participants focus on one of the cues, and if that specific cue appears, then they give a consistent answer, but in all other cases they answer randomly. Note that these strategies are not conscious calculations but implicit algorithms, and strategy use is defined by curve-fitting. Earlier results showed that participants usually use a singleton or a one-cue strategy in the early phase of the task, but in the later phases the number of multi-cue users increase (Gluck, Shohamy and Myers, 2002).

TRANSPARENCY AND FEATURE COMBINATION IN THE WP TASK

Though there is a 'canonical', most widely used version of the Weather Prediction task, a number of different versions appeared in its history over the past 20 years. As it is described below, researchers considered the different versions identical. The only variable that was proposed to have an effect on categorization was the mode of stimulus presentation (Shohamy et al., 2001). Though authors of the papers cited below imply that the modifications do not affect performance, reviewing the results suggests otherwise.

One of the first papers, which compared the performance of amnesic and healthy individuals (Knowlton, Squire and Gluck, 1994), used three different conditions of the PCL task. One condition was called the Weather Prediction with Cards (WP) task (pp. 111), the task that is in the focus of the current paper. Along with the WP task, two other versions were used with the same structure. One of them was Weather Prediction with Pictures, where the task of the participants was to decide whether it would be Sunshine or Rain on the basis of the appearance of individual pictures. In the same design as the cards were presented in the WP task, pictures of a Sailing boat, a Candle, a Butterfly and a Telephone appeared on individual cards in this condition. The third task that was a Medical Diagnosis task in which one, two or three out of four symptoms appeared, and the task of the participants was to decide which fictive disease these symptoms would predict. While the three tasks were considered to be identical by the authors, they were not matched on every variable (e.g. the WP Pictures task had only 50 trials), and results suggest that categorization performance might have been affected by variables that were not controlled for in Knowlton et al.'s design.

Shohamy and colleagues (2001) used a different task (the Ice-Cream task, IC) in which the different cues were not presented as distinct entities, but they were features of the same image: a puppet appeared which could wear a hat, a bow-tie, glasses or have a moustache, with any combination of these four features, and the goal of the participants was to guess, based on the appearance of the puppet whether it would ask for chocolate or vanilla ice-cream. Results showed that on the Ice-Cream task control participants can be categorized into two groups: solvers (who learn faster and reach high performance in a short time) and non-solvers (who learn slower and perform worse than previous studies with the original version of the task suggest).

Different studies report that Parkinson's patients perform on the WP task rather poorly (8% of 12 patients used multi-cue strategy, see Figure 5, Shohamy et al., 2004), while strategy use is better in the Ice-cream task (16% of 13 participants, Shohamy et al., 2004, p. 856). Another study by Hopkins and colleagues (2004) on hypoxic patients also found lower overall accuracy, but more advanced strategy use on the IC task. Unfortunately the discrepancy is not explained as the two tasks are not compared to each other, which might also be due to the fact that very low number of participants were employed.

Taken together, the different versions of the weather prediction task differ in a number of variables (arbitrariness of the link between cues and outcomes, cue combination), but the specific effects of these variables on learning have not been studied systematically and in detail so far, although they may result in rather different levels of performance (Hopkins et al., 2004; Knowlton, Squire and Gluck, 1994). The current paper focuses on two dimensions concerning the structure of the task: Combination and Transparency. Cue Combination is a variable along which the WP and the IC tasks are suggested to differ (Shohamy et al., 2001). In the WP task cues are presented as distinct entities, while in the IC task cues are features of a single object. Transparency on the other hand introduces another dimension on which tasks may vary from the WP task through the IC task to the Medical Diagnosis task. In the WP task cues and outcomes have nothing in common except for the statistical link established by the task. The statistical link is preserved in the other two tasks, but the IC task is a bit less arbitrary, while the medical diagnosis is truly transparent (though the notion of transparency is hard to quantify). In the study presented below, both variables (Combination and Transparency) can take two values (Combination: Holistic versus Cue-based; Transparency: Transparent versus Arbirtrary), adding up to four conditions. Although earlier studies in the literature suggest that performance is lower when cues are presented as features of a single image, these results are not conclusive. The tasks used earlier did not only differ on combination, but also on transparency, and previous papers never compared differences between different versions of probabilistic categorization tasks directly. The aim of the current research is to disentangle the effect of stimulus organization (transparency and combination) on probabilistic categorization performance in adults and children.

THE WP TASK, TRANSPARENCY AND COMBINATION IN CHILDREN

Previous results also suggest that performance of typically developing adults and children do not differ on the WP task (Kemény and Lukács, 2010; mean age of children was 11;3, SD: 1,3). The current study targeted a younger group of children from a narrower age range, to test whether the equivalence of performance still holds between a younger group of children and adults.

Transparency and combination are expected to modify categorization performance of adults. It is not clear, though, whether these changes have the same effect on children's performance. The use of less abstract stimuli in transparent conditions might further improve the performance of children compared to adults. Or from a different angle, an arbitrary setting, the use of abstract figures and abstract links between stimuli and outcomes might result in low categorization performance in children, while adults might be more prepared to use arbitrary stimuli and links in categorization.

METHOD

Participants

Altogether 217 people participated in the experiment: 134 adults and 83 children. Participants were randomly assigned to four different conditions. *Table 1* gives the number of participants in each condition and their ages. All adults were recruited at the Budapest Technical University, and participated voluntarily in the study for credits. They were informed about the purpose of the research after the study. All adult participants and parents of all children provided written informed consent, in accordance with the principles set out in the Declaration of Helsinki and the stipulations of the local Institutional Review Board.

Table 1. Number and age of participants by Group and by Condition

Condition	Adults			Children		
	Age	Std. Dev.	N	Age	Std. Dev.	N
Cue-based–Transparent	20;9	1;3	33	8;6	0;6	23
Cue-based–Arbitrary	21;6	1;4	31	8;9	0;4	20
Holistic–Transparent	21;0	1;9	35	8;8	0;3	20
Holistic–Arbitrary	21;3	1;11	35	8;7	0;4	20

Stimuli

Participants were assigned to four different groups. Each group faced one of four conditions. The four different conditions varied along two test dimensions independently, yielding 4 conditions altogether: *Transparency* (*Transparent* vs. *Arbitrary*) and *Combination* (*Holistic* vs. *Cue-based*). In the *Transparent* conditions there was a transparent link between cues and outcomes. This transparency was perceptually based (see below), while in the *Arbitrary* conditions the link between cues and outcomes was strictly due to feedback. In the *Cue-based* conditions cues were presented as single, distinct images, while in the *Holistic* conditions cues were different features of a single image.

The *Cue-based–Arbitrary* condition was the classical version of the WP task with modified predictive values for cues (as used in Kemény and Lukács, 2010). Cues were different geometrical forms – square, triangle, pentagon and rhombus – presented in different combinations simultaneously, and the role of the participants was to decide whether the given set of cues signal rain or sunshine.

In the *Cue-based–Transparent* condition cues were segments of line drawings. Participants had to guess whether the objects appearing on the screen belong to a boy or a girl. Each cue could be completed in two different ways to suit the two categories. Cue1 could be completed to yield either a bow-tie or a necklace, Cue2 could be completed into either a fishing-pole or a flower, Cue3 could turn into either a football or a basket, and Cue4 could become either a pair of shorts or a skirt (for all cues, the first object belonged to a boy, while the second object belonged to a girl, see *Figure 1*). As soon as participants answered, the cues were completed, and the gender of the owner (boy/girl) was revealed written up on the screen. To avoid the possibility that participants miscategorize the outcomes, at feedback each completed cue was labelled according to the object it formed (i.e. *szoknya* – 'skirt' in Hungarian – for an object belonging to a girl following Cue4).

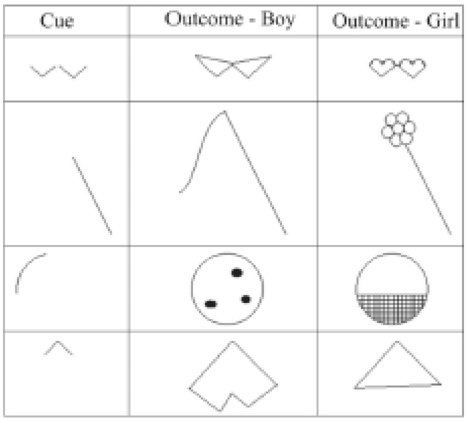


Figure 1. Cues, outcomes, and the transparent link between cues and outcomes in the *Cue-based–Transparent* condition

The *Holistic–Transparent* condition was very similar to the Cue-based–Transparent condition. The same cues were used, but they were combined into a single image. In the combined picture, the cues together with their outcomes appeared as part of a line drawing of a child. The role of the participants was to guess whether they would see a boy or a girl following the cues. As soon as they answered, the picture was completed, and the gender of the child was written below the picture (see *Figure 2*).

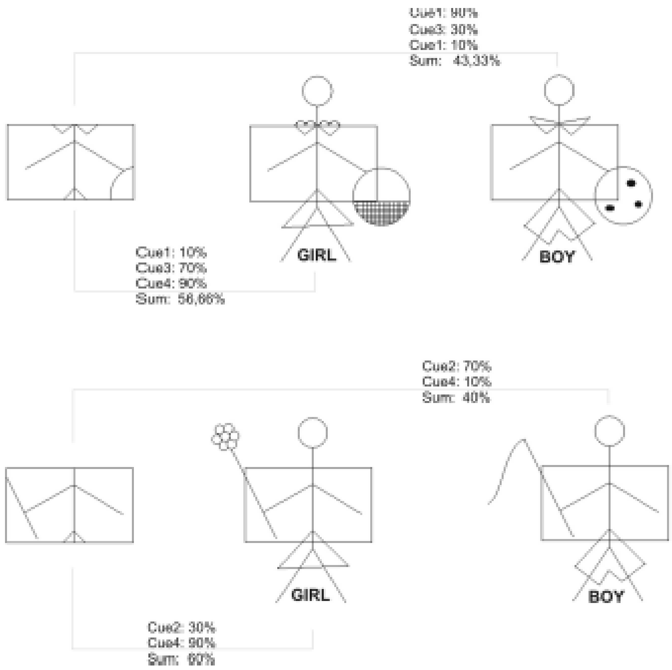


Figure 2. Cues, outcomes and the transparent link between them in the *Cue-based–Transparent* condition

The *Holistic–Arbitrary* condition (see *Figure 3*) was very similar to the Holistic–Transparent condition. The same images were used, and participants were asked to decide whether they predict sunshine or rain. As soon as they answered the correct outcome appeared on the screen along with the cues (similarly to the Cue-based–Arbitrary condition).

Apart from cue arrangements and the relationship between cues and outcomes, the four conditions were structurally identical. In each condition four blocks were administered. Each block included 50 items. In terms of cues and outcomes the order and feedbacks of the items were identical throughout the conditions – meaning for example that in item 15 in each task, Cue1 and Cue3 were present, and the outcome was always 1 (which was 'sun' in the arbitrary conditions and 'boy' in the transparent conditions). In all four conditions Cue1 lead to outcome 1 in 85,7% of all its appearances, Cue2 lead to outcome 1 in 70% of all its appearances, Cue3 predicted outcome 1 in 30%, while Cue4 predicted outcome 1 in 14.3% of all cases. In all other cases (14.3% of the appearances of Cue1, 30% of the appearances of Cue2, 70% of

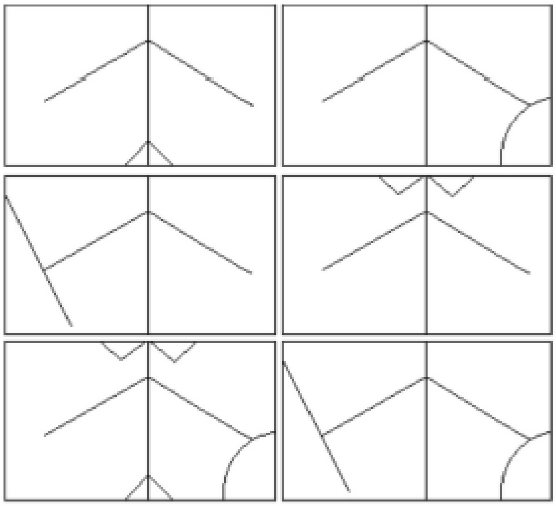


Figure 3. Cues of the *Holistic–Arbitrary* condition. The top four cues are the single cues, while the bottom row illustrates how cue-combinations are merged into single images

Cue3, and 85.7% of Cue4) appearances lead to outcome 0). Note that these predictive values are not identical to those used in the experiments of Knowlton et al. (these studies used 77%, 58%, 42% and 23% predictive values Knowlton, Mangels and Squire, 1996; Knowlton, Squire and Gluck, 1994) and Gluck, Shohamy and Myers (using 80%–60%–40%–20% predictive values Gluck, Shohamy and Myers, 2002; Hopkins et al., 2004; Shohamy, Myers, Onlaor and Gluck, 2004; Shohamy, Myers, Grossman et al., 2004). The reason for modifying the predictive values was because preliminary studies suggested that lower values were not suitable for testing children (Kemény and Lukács, 2007).

Procedure

Participants were seated in cubicles of the laboratory of the Department of Cognitive Sciences. The task was computerized, and ran on E-prime. Participants were instructed that they will see figures and have to decide (1) whether it will be rain or sunshine (arbitrary conditions); (2) whether they see a boy or a girl (holistic-transparent condition); or (3) whether the objects they see belong to a boy or a girl (cue-based–transparent condition). In all conditions there were four blocks, all blocks using the same pseudorandom organization of cues and outcomes. Each block consisted of 50 stimulus-outcome pairs. A stimulus appeared and was present until the participant responded using one of the keys on the keyboard (ENTER for outcome 1, SPACE for outcome 2). Immediately after the button press the correct answer appeared on the screen. In the arbitrary cases the correct answer was an icon depicting either sunshine or rain. The icon appeared below the stimulus, and was present along with the stimulus for 1500 msec. The transparent condition was similar, fragment of a picture (holistic) or fragments of several pictures (cue-based) were shown. The stimulus was present until response was given (by pressing either ENTER or SPACE). After response the correct answer appeared on the screen: the correct answer was the full picture which included the stimulus too (at its original place). Feedback was present for 1500 msec. After that the next stimulus appeared. Between each block of 50 trials a break-slide appeared. The break was over when the participant pressed a button on the keyboard.

RESULTS

Overall accuracy

Performance was measured as the overall rate of correct answers. A 2 × 2 × 2 × 4 Repeated Measures ANOVA was conducted on overall accuracies with Transparency (Transparent vs. Arbitrary), Combination (Holistic vs. Cue-based) and Group (Adults vs. Children) as between-subject variables and Block (1–4) as a within-subject variable.

The ANOVA revealed a significant main effect of Block ($F(3,216) = 46.932$, $\eta^2 = 0.183$, $p < 0.001$), a significant main effect of Combination ($F(1,216) = 4.445$, $\eta^2 = 0.021$, $p < 0.05$), a significant main effect of Group ($F(1,216) = 53.339$, $\eta^2 = 0.203$, $p < 0.001$), while the main effect of Transparency was not significant ($p = 0.421$ – for all main effects see *Figure 4*). The Block × Transparency ($F(6,216) = 6.398$, $\eta^2 = 0.030$, $p < 0.001$, *Figure 5*) and the Block × Group ($F(6,216) = 2.647$, $\eta^2 = 0.013$, $p < 0.05$, *Figure 6*) interactions were significant, while all other interactions (both 2-way and 3-way) were not significant (all $ps > 0.189$).

Block × Group Interaction

To disentangle the significant Block × Group interaction, two one-way repeated measures ANOVAs were conducted – one for each age-group. For adults, the ANOVA revealed a significant main effect of Block ($F(3,133) = 47.610$, $\eta^2 = 0.264$, $p < 0.001$). Bonferroni corrected post hoc testing revealed that except for Block 3&4 ($p = 0.13$) performance from all blocks significantly differed from each other (Block 2&3: $p < 0.01$, while for all other comparisons $p < 0.001$).

For children, the ANOVA revealed a significant main effect of Block ($F(3,82) = 10.526$, $\eta^2 = 0.114$, $p < 0.001$). Bonferroni corrected post hoc tests showed that differences between Block 1&2 ($p = 0.08$) 2&3 ($p = 0.32$) or 3&4 ($p = 0.49$) were not significant, but all other comparisons were significant (Block 2&4: $p < 0.05$, Block 1&3: $p < 0.01$, while Blocks 1&4: $p < 0.001$, see *Figure 6*).

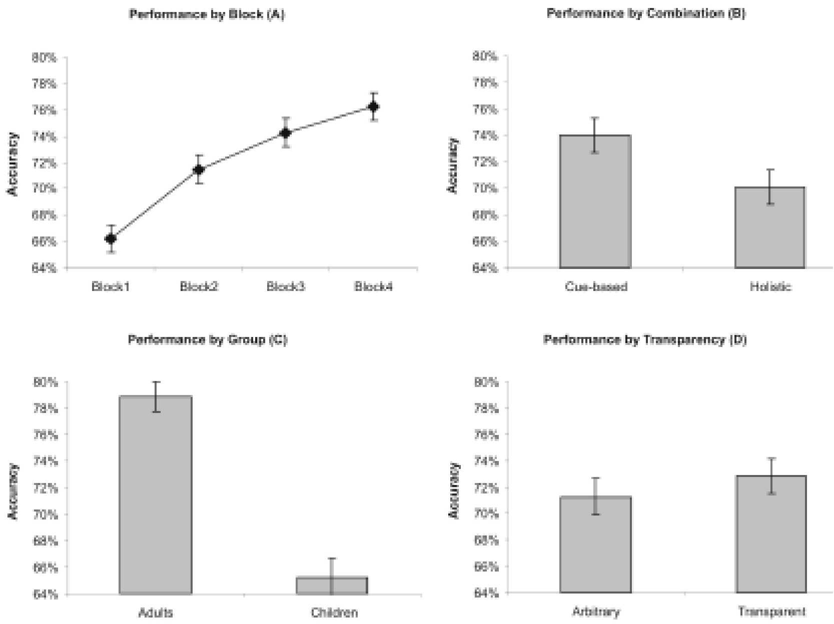


Figure 4. Main effects on overall accuracy. (A) illustrates the main effect of Block, (B) illustrates the main effect of Combination, (C) the main effect of Group, and (D) the main effect of Transparency. The significant main effect of Block reflects significant differences between all blocks (for Blocks 3 and 4 $p < 0.01$, all other *ps* < 0.001). Main effects of Combination ($p < 0.05$) and Group ($p < 0.001$) were also significant, while the main effect of Transparency was not significant. Error bars indicate Standard Error of Mean

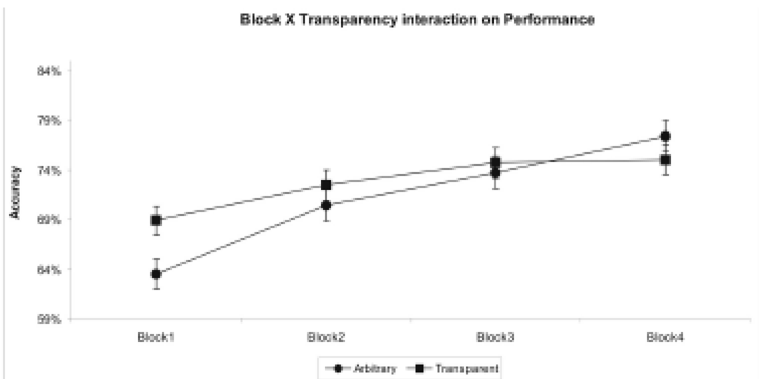


Figure 5. The effect of Block × Transparency interaction on overall accuracy. Error bars indicate Standard Error of Mean

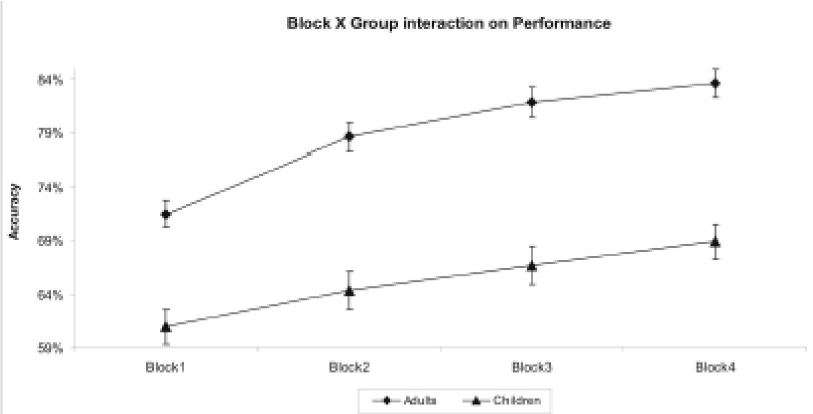


Figure 6. The effect of Block × Group interaction on overall accuracy. Error bars indicate Standard Error of Mean

Block × Transparency Interaction

To properly understand the significant Block × Transparency interaction, two one-way repeated measures ANOVAs were conducted. Participants differed along the Transparency variable. Members of the two *Arbitrary* conditions showed a significant main effect of Block ($F(3,105) = 37.041$, $\eta^2 = 0.261$, $p < 0.001$). Bonferroni corrected post hoc tests revealed that all blocks differed from each other significantly (Block 2&3: $p < 0.05$, 3&4 $p < 0.01$, all other *ps* < 0.001).

Performance of people participating in the *Transparent* condition also showed a significant main effect of Block $F(3,110) = 37.041$, $\eta^2 = 0.261$, $p < 0.001$). Bonferroni corrected post hoc pairwise comparisons revealed that the difference between Block 3&4 were not significant, all other differences were significant (for Block 2&3 and 2&4 $p < 0.05$, Blocks 1&2: $p < 0.01$, all other *ps* < 0.001).

Strategy fitting

A four-way hierarchical loglinear analysis was conducted on strategy use with Strategy use, Group, Transparency and Combination as within-subject variables. Results showed that two-way interactions were significant (η^2 (N = 217) = 57.915, df = 9, $p < 0.001$), higher-way interactions were not significant (all *ps* > 0.491). Two-way Chi-square tests revealed that the distribution of strategy use differed significantly depending on Transparency (η^2 (N = 217) = 11.343, df = 2, $p < 0.01$) and on Group (η^2 (N = 217) = 40.462, df = 2, $p < 0.001$). For distributions of strategy use by groups see *Table 2*.

Table 2. Distribution of strategy use by Group

		Cue-based		Holistic	
		Transparent	Arbitrary	Transparent	Arbitrary
Adults	Multi-cue	21	22	15	23
	Single str.	12	9	19	12
	No strat.	0	0	1	0
Children	Multi-cue	3	8	1	5
	Single str.	19	9	17	10
	No strat.	1	3	2	5

Group effect on the original WP task

A 2×4 ANOVA was conducted to compare the performance of children and adults. The ANOVA revealed a significant main effect of Block ($F(3,216) = 14.370$, $\eta^2 = 0.227$, $p < 0.001$), and a significant main effect of Group ($F(1,216) = 17.170$, $\eta^2 = 0.259$, $p < 0.001$), while the interaction was not significant ($F(6,105) = 0.316$, $p = 0.814$).

Strategy analysis revealed that the difference in the distribution of strategy use between children and adults was significant (η^2 ($N = 217$) $= 7.510$, $df = 2$, $p < 0.05$).

DISCUSSION

The aim of the current study was twofold. On the one hand, we tried to test whether holistic (as opposed to cue-based) presentation enhances learning on the WP task (Shohamy et al., 2001) or there may be other factors underlying this discrepancy. On the other hand, we tested whether children perform at the same level of adults in probabilistic categorization tasks, as suggested in an earlier study (Kemény and Lukács, 2010), or that may only apply to older children. Results showed that cue-based stimulus presentation leads to better probabilistic categorization performance. The early phase of categorization is enhanced by transparency between stimuli and outcomes, but this advantage turns into a disadvantage in the later phases of the task. Results also showed that while adults perform better than children, the variables in focus affect both adults and children similarly.

Though earlier results have been partially replicated, in general it is not obvious that the difference between holistic and cue-based tasks reported earlier (Shohamy et al., 2001) is entirely due to presenting cues in a combination, as parts of one figure. Our results did confirm findings from the IC task, that if all cues are presented as features of a single image, performance deteriorates (compared to the case when each cue is a single entity). At the same time our results suggest that this difference is not rooted in strategy use. Shohamy and colleagues suggest that in the case of holistic stimulus presentation typically developing healthy subjects may be grouped into "Solvers" and "Non-solvers". This differentiation of healthy subjects is not seen in our data of strategy use. We found that though overall accuracy is lower in the *Ho-*

listic conditions, strategy use is the same – i.e. a lower performance was achieved using the same strategies.

Transparency on the other hand shows a different pattern. In general, people participating in the *Transparent* conditions showed the same overall accuracy as participants of the *Arbitrary* conditions – though there was an interaction. In the first block there is an advantage of the *Transparent* conditions, but this advantage seems to decrease. By the fourth block performance on the *Arbitrary* conditions becomes higher (though this difference is not significant). This "turnover" of performances would not have been evident in a 3 block design, and even with this 4 block design, we are still unable to tell whether the disadvantage really turns into an advantage in the later phases or not. A design with more blocks would be needed to find out more about the Transparency × Block interaction.

Strategy use, on the other hand, does differ by Transparency. Results show that participants in the *Transparent* conditions show a lower ratio of multi-cue strategy use, but a higher ratio of using one of the single strategies; the number of people using no identifiable strategies is also lower. Our results suggest that transparency between categories and predictive features supports the use of single strategies, though the switch to multi-cue strategy is more difficult in the *Transparent* condition. Overall, perceptual transparency facilitates single cue strategies, though at the same time a meaningful relation between cues and outcomes seems to lock single strategies so that focusing on more cues becomes more difficult for participants.

At this point we can only speculate on the reason for this pattern. One possibility is that this result is rooted in the implicit/explicit nature of learning. As reviewed above, amnesic patients perform normally in the initial phases of the WP task, and their performance only decays later (Knowlton, Squire and Gluck, 1994), while patients with Parkinson's syndrome show impaired performance already from the beginning of the task (Knowlton, Mangels and Squire, 1996). On the authors' interpretation, these results (also confirmed by fMRI studies) imply that the early phases of probabilistic categorization rely on the procedural system, while the later phases rely on declarative functioning (Knowlton, Mangels and Squire, 1996; Poldrack et al., 1999; 2001). Later research on strategy use suggests that in the beginning of the task participants use one of the single strategies, i.e. the one-cue strategy or the singleton strategy, and they gradually develop the multi-cue strategy (Gluck, Shohamy and Myers, 2002). Taken together, these results imply that the single strategies should rely on the procedural, while the multi-cue strategy may depend on the declarative system. Though there may be theoretical differences (for more details see Sun, Slusarz and Terry, 2005), most researchers basically identify procedural knowledge with implicit knowledge (Meeter et al., 2008).

Transparency tends to move strategy use towards the single strategies, while arbitrariness seems to facilitate the integration of predictive values in the multi-cue strategy. Mapping this distinction onto the implicit-explicit dichotomy suggests that transparency co-occurs with implicit processes while arbitrariness facilitates explicit processes. This explanation implies that in the case of a transparent link, participants do not seek explicit hypotheses and strategies, since the motive for one or another outcome is transparent. On the other hand, the relationship between the cues and outcomes is not motivated in arbitrary setting but established only by the statistical link. In this case, the lack of a perceivable relationship might motivate explicit hypothesis testing more, thereby facilitating declarative strategies. This hypothesis also implies that awareness of the link between cues and outcomes will be higher in an arbitrary setting than in a transparent setting. Further research is required to test this hypothesis.

All in all our results suggest that the difference between performance on the Weather Prediction task and the Ice-Cream task (Hopkins et al., 2004; Shohamy et al., 2004; Shohamy et al., 2001) may not be entirely due to the holistic presentation of stimuli (and not entirely due to differences in transparency), but rather to an interaction between *Combination* and *Transparency*. Further research is needed to disentangle the similarities and differences between the tasks used by our experiment and the Ice-Cream task.

The second aim of the study was to test age-effects in probabilistic categorization. Kemény and Lukács (2010) compared the performance of typically developing adults, typically developing children and children with Language Impairment, and found that the two typically developing groups did not differ from each other significantly. On the other hand, our results show that children perform significantly worse on the task both in terms of accuracy and strategy use. The reason for not replicating the previous result may be rooted in sampling. In that study typically developing children's mean age was 11;3 (Std. dev, 1;3), while the current group of children was approximately two and a half years younger. They were also from a more homogenous age group, since all children were from the same year. Contrary to 11-year-olds, younger children performed worse on the WP task than adults. As there were no significant interactions with age, testing whether the discrepancy between the performance of adults and children differs by conditions was not motivated. Further research is required to draw a complete developmental curve for the WP task, and to test whether the performance of adults and children shows a similar or different pattern along other dimensions.

REFERENCES

Gluck, M. A., Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 225–244.

Gluck, M. A., Shohamy, D., Myers, C. (2002): How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*(6), 408–418.

Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., Gluck, M. (2004): Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia, 42,* 524–535.

Kemény, F., Lukács, Á. (2007): *Valószínűségi kategorizáció felnőtt- és gyerekkorban: napsütés vagy csokifagylalt?* Paper presented at the MAKOG XV., Eger.

Kemény, F., Lukács, Á. (2010): Impaired procedural learning in language impairment: Results from probabilistic categorization. *Journal of Clinical and Experimental Neuropsychology, 32*(3), 249–258.

Knowlton, B. J., Mangels, J. A., Squire, L. R. (1996): A neostriatal habit learning system in humans. *Science, 273*(5280), 1399–1402.

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

Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., Hopkins, R. O. (2008): Probabilistic categorization: How do normal participants and amnesic patients do it? *Neurosci. Biobehav. Rev., 32*(2), 237–248.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., Gabrieli, J. D. E. (1999): Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*(4), 564–574.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., Gluck, M. A. (2001): Interactive memory systems in the human brain. *Nature, 414*(6863), 546–550.

Shohamy, D., Myers, C. E., Onlaor, S., Gluck, M. A. (2004): Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*(4), 676–686.

Shohamy, D., Onlaor, S., Myers, C., Gluck, M. (2001, March): *Category learning in Parkinson's disease: The effect of stimulus presentation on performance.* Paper presented at the 8th Annual Meeting of the Cognitive Neuroscience Society, New York.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., Poldrack, R. A. (2004): Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain, 127*(4), 851–859.

Sun, R., Slusarz, P., Terry, C. (2005): The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychol. Rev., 112*(1), 159–192.

Yamada, T., Izyuuinn, M., Schulzer, M., Hirayama, K. (1990): Covert orienting attention in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry, 53*(7), 593–596.

8. Study 3: Perceptual Effect on Motor Learning in the Serial Reaction-Time Task

Kemény, F., & Lukács, Á. (2011). Perceptual Effect on Motor Learning in the Serial Reaction Time Task. *Journal of General Psychology, 138*, 110-126.

The Journal of General Psychology, 2011, *138*(2), 110–126

Perceptual Effect on Motor Learning in the Serial Reaction-Time Task

FERENC KEMÉNY
Budapest University of Technology and Economics

ÁGNES LUKÁCS
Budapest University of Technology and Economics
Hungarian Academy of Sciences

ABSTRACT. Although the Serial Reaction-Time Task has been an effective tool in studying procedural learning, there is still a debate as to whether learning in the task is effector-based, stimulus-based, or response-based. In this article, the authors contribute to this debate by contrasting response- and stimulus-based learning by manipulating them selectively and simultaneously. Results show that (a) participants learned response sequences in the absence of stimulus-specific perceptual sequence information but (b) not stimulus sequences without corresponding response information. In a third condition, response sequence and stimulus frequency information were in conflict, and each effect decreased learning in the other domain. Overall, our findings show that learning in these tasks is primarily motor-based, but it is also constrained by relatively salient perceptual information. Together with earlier findings in the literature, the findings also suggest a task and stimulus-arrangement-specific interaction between motor and perceptual learning, where relevance and salience of the specific information plays a crucial role.

Keywords: motor-based learning, perceptual-based learning, sequence learning, serial reaction-time task

IN EVERYDAY LIVES, PEOPLE OFTEN FACE or produce sequentially structured patterns. One follows a given sequence of turning left and right as he or she walks from one point to another. One follows a sequence of hand movements when writing a signature. One produces a sequence of oral movements when saying "Hello." These sequences are mostly acquired and fine-tuned in an implicit way, without explicit awareness and without the ability of conscious recall

The research was supported by Implab "Implicit Cognition" research grant. The authors are grateful to two anonymous reviewers and Dr. Paula Goolkasian for their helpful comments and suggestions.
Address correspondence to Ferenc Kemény, Department of Cognitive Science, Stoczek u. 2., 1111 Budapest, Hungary; fkemeny@cogsci.bme.hu (e-mail).

(Reber, 1993). In a controlled environment, these sequences can be modelled in different ways. One of the most well-known experimental paradigms of implicit sequence learning is the Serial Reaction Time (SRT; Nissen & Bullemer, 1987) task, a task in which a target stimulus can appear in one of four different locations, and participants are asked to react by pressing one of four buttons corresponding to the location of the target. Stimuli do not appear randomly. Rather, there is a fixed order that determines their appearance. This fixed order is usually a 10- to 12-element long sequence. The participants' average reaction times per block decrease throughout the consecutive training blocks. When the sequence is broken, and the target stimulus starts to appear randomly, reaction time increases, but as soon as the fixed order returns, the mean latency decreases again (e.g., Destrebecqz & Cleeremans, 2001).

Ever since the introduction of the SRT task, the issue of what is learned has been debated. There have been three accounts of the nature of learning: effector-based learning (Deroost, Zeeuws, & Soetens, 2006), perceptual-based learning (Remillard, 2003) and motor-based learning (Willingham, Wells, Farrell, & Stemwedel, 2000). Effector-based learning explains effects in the SRT task as learning by specific organs, and learning on this view can be characterized as the acquisition of an effector-movement sequence. Such a sequence may be the movement sequence of the fingers' muscles. This hypothesis implies that there should be no transfer observed from one hand to the other. Also, there should be no transfer from motor movement to verbal responses, where participants say aloud the location of the target stimulus. A number of studies argue against this hypothesis based on observations of transfer effects between effectors (Cohen, Ivry, & Keele, 1990; Grafton, Hazeltine, & Ivry, 2002; Keele, Jennings, Jones, Caulton, & Cohen, 1995).

The hypothesis of perceptual-based learning explains the sequence-correlated decrease of reaction times as a result of learning the stimulus sequence—that is, as stimulus–stimulus learning (Remillard, 2003). This theory suggests that in the SRT task, participants acquire the sequence as a row of images, and sequence learning lowers reaction times because participants predict the location of the next appearance. The proposal of motor-based learning predicts that the acquired sequences are response sequences (response–response learning)—that is, sequences of motor information (Willingham et al., 2000). This way, reaction time decrease is due to the prediction of the next response.

While stimulus- and perceptual-sequence learning are mostly used as interchangeable notions, the case is not as clear for motor learning. Most scholars use motor and response information interchangeably but contrasted with effector-based information. The major difference between motor and effector information is their level of abstractness. While effector information consists of specific muscle movements, motor information is usually considered at a more abstract level, free of effector-specific information (Mayr, 1996; Willingham, Nissen, & Bullemer, 1989; Willingham et al., 2000). This way, a motor sequence may be considered an abstract, goal-oriented motor program.

Several results suggest that learning is not based on either perceptual or motor sequences alone but is best understood as a combination of the two different sources of information. There are studies claiming that the convergence of sequential motor and perceptual information leads to a higher level of implicit learning (Robertson, Tormos, Maeda, & Pascual-Leone, 2001). Others show that in the case of probabilistic sequences, perceptual and motor information influence sequence learning to the same extent (Németh, Hallgato, Janacsek, Sandor, & Londe, 2009).

In the original version of the SRT task, stimuli and responses have a one-to-one relationship, resulting in converging stimulus and response sequences. This way the original "location-learning" version of the SRT task cannot decide between the above alternatives. There have been several approaches to modify the task in order to disentangle the effects of stimulus and response sequences on learning. In one modification, participants were exposed to a stimulus sequence but were only required to respond to the first pattern cycle in each block. In spite of the reduced motor information during learning, they showed the same sequence-specific reaction time pattern as participants of the SRT condition (Howard, Howard, & Mutter, 1992). This result was interpreted as an indication that perceptual-based learning does take place in the absence of sequential motor information, although, in this case, the implicitness of the task was questioned (Kelly, Burton, Riedel, & Lynch, 2003). However, it has also been shown that if the perceptual information is not the target dimension, that is, if it is not what participants focus on and respond to, then there is no implicit learning of pure perceptual information (Willingham, 1999; Willingham et al., 1989). There are also results showing the same reaction time patterns when participants have to name objects. In the naming task, there was no sequence in oral movements or in stimuli, but the categories of the pictures to be named followed a sequence, and sequence learning took place in such a setting as well (Goschke & Bolte, 2007).

The current article focuses on the nature of learning in the SRT task. We are especially interested in the perceptual structure of the information to be learned. The study presented below addresses the following questions: (a) Does sequence learning take place in the absence of stimulus-specific perceptual sequence information—that is, when the sequence is only present in the category of the stimuli to which participants respond; (b) does sequence learning take place based purely on perceptual information—that is, when all sequential information is coded in the stimuli themselves, and no sequential information is present in the response pattern; and (c) when the sequence is based on response information only, is it influenced by noncongruent but relatively salient perceptual structure? As already reviewed, theories of sequence learning usually treated stimulus and response learning mechanisms as autonomous and independent of each other. Previous studies mostly suggested that stimulus and response learning took place separately without any interaction between the different types of sequence information. The present study combines predictions of stimulus and response information to find out whether the different layers of learning are really independent of each other and whether the different types of sequence-specific information play similar or different roles in sequence learning. In contrast to earlier studies, the current experiment uses a two-dimensional (location vs. identification) setting, which makes it possible to compare the effect of stimulus (perceptual) and response (motor) learning with very similar tasks. Some researchers question the possibility of sequence learning in the case of multiple dimensions, though there are also studies suggesting that sequence learning may take place if both motor and perceptual information are varied (Mayr, 1996; Németh et al., 2009).

We designed an experimental setting that allows the simultaneous yet independent control of perceptual and motor information. In the experimental setting, pictures were shown to participants, and participants were asked to press one of four buttons according to the semantic category of the presented stimulus. Response sequences were elicited by sequences of categories. At the same time the location of the pictures were varied. Stimulus sequences consisted of a repeating pattern of stimulus location, irrespective of the identification of categories. Picture locations did not carry any response information, since participants did not have to respond to stimulus locations. If, as a number of articles suggest, learning in the classical deterministic serial reaction time task does not require attentional effort (Cohen et al., 1990; Curran & Keele, 1993; Jiminez & Mendez, 1999; Nissen & Bullemer, 1987), we might expect learning to take place even if the sequential information is not in the focus of attention—that is, it is not the information determining button presses. In this case, learning would be based on perceptual information alone, and a purely perceptual-based learning effect should be observed (Remillard, 2003). If learning on the SRT task does not rely on perceptual information, then we would expect sequence-specific learning to take place when stimuli lack sequence-specific information of stimulus location. Learning in such a case would be indicative of motor-based learning (Willingham, 1999). If learning on the SRT task is motor-based, and sequential perceptual information has no effect on reaction times, then a question arises whether perceptual information has an effect on sequence learning at all.

The current study uses three experimental conditions. In all conditions, pictures of four categories are presented, and the role of the participants is to respond quickly by pressing one of four buttons on the keyboard. Each button is associated with one of the categories. The Response condition tests whether sequence learning takes place to a response sequence in the absence of sequential stimulus location information. The Stimulus location condition tests if pure perceptual sequence learning takes place—that is, if sequence learning appears based on sequential perceptual information without the presence of a response (motor) sequence. In the Extra condition, a nonsequential stimulus location structure is added to the response sequence of the Response condition. The nonsequential stimulus location structure serves as predictive perceptual information. The target stimulus may appear in four different locations, and each location is associated with one high-frequency response and three low-frequency responses. Certain stimuli are more likely to appear at a given location and less likely to appear at other locations. The questions are (a) whether this stimulus location structure is acquired to the same extent in the presence or absence of the response sequence, and (b) whether response sequence learning takes place to the same extent in the presence or absence of the nonsequential stimulus structure. To answer this latter question, reaction times of the Extra condition are compared to those of a Frequency Control condition. The spatial information of the stimuli of the Frequency Control condition were identical to that of the Extra condition, but there was no underlying response sequence.

Earlier studies mostly suggested exclusivity of either motor or perceptual sequence learning (Remillard, 2003; Willingham, 1999; Willingham et al., 2000). Others argued that both components contribute sequence learning, and they do so to the same extent (Deroost et al., 2006; Nattkemper & Prinz, 1997; Németh et al., 2009). It is also possible that perceptual and motor factors play a different role in the SRT task. As a result, learning within the two domains may not be functionally equivalent. A goal of the current experiment is to selectively manipulate stimulus and response sequences. This way, we might be able to find out whether learning stimulus and response structures really take place independently, or whether there is an interaction between the processing of the two types of information, determined by the task and the specific arrangement of stimuli.

Method

Participants

Altogether, 110 young adults participated in the study (44 female, 66 male, *M* age = 22.1 years, *SD* = 2.25 years). Participants were randomly assigned to four different conditions. Thirty people participated in the Response condition, 30 participants were assigned to the Stimulus location condition, 24 people were in the Extra condition, and 26 participants were tested in the Frequency Control condition. All participants were recruited from the Budapest University of Technology and Economics, and they participated for credit. All participants provided written informed consent, in accordance with the principles set out in the Helsinki Declaration and the stipulations of the local Institutional Review Board.

Materials

To avoid one-to-one relationship between stimulus and response sequence information, participants were required to respond to categories instead of individual stimuli. Stimuli were presented in four categories. Participants in all four conditions were exposed to pictures of furniture, mammals, fruits and tools. Their task was to press the buttons corresponding to the semantic category of the given picture, i.e. to press *y* for furniture, press *c* for mammals, press *b* for fruit, and press *m* for tools. Note that Hungarian keyboards were used, where buttons *y*, *c*, *b*, and *m* are arranged horizontally, with one button between each neighboring pair.

The use of categories instead of tokens made it possible to introduce response sequences without the presence of a correlating stimulus sequence. Although categories formed a sequence, the specific items were selected randomly for each appearance out of 10 pictures for each category. The pictures of each category have been tested in a pilot study. Only those pictures were used in the experiments that were consistently recognized by all participants of the pilot study.

In the Response condition, the appearance of the categories followed a sequential structure, resulting in a 12-element response sequence ("ycymcb-mybcmb"). A response sequence fragment like "mybc" could be achieved through the presentation of any of the following stimulus sequences: "screwdriver–bed–strawberry–lion", "chainsaw–bed–apple–dog" or "hammer–wardrobe–pineapple–dog." No sequential perceptual information was present: Stimuli appeared in the center of the screen (excluding location effects), and with the use of categories instead of single items there was no one-to-one correspondence between the perceptual information and the given button presses. All images appeared in the center of the screen. Note that a repeating sequence of categories was present in the stimuli, determining the motor-response sequence.

In the Stimulus location condition, the appearance of the categories was random, but participants were still required to respond according to categories. This way, there was no sequence present in the responses. Although the order of categories was random, stimuli did not appear at a single location in the center but could pop up in any of the four corners of the screen. The location of the consecutive stimuli followed a 12-element sequence: "121423413243," in which *1* marks the top left corner location, *2* is the top right corner of the screen, *3* is the bottom left corner and *4* is the bottom right corner. For illustration[1] see Figure 1. The task of the participants was the same as in the Response condition: To press the button that corresponds to the given category.

The Extra condition was similar to the Response condition in that the sequential information was present in the responses. The sequence consisted of 12 consecutive elements ("ycymcbmybcmb"). Whereas the goal of the Response condition was to eliminate sequential perceptual information, the Extra condition was designed to test whether response learning is affected by any other kind of image-level perceptual information. As the name of the condition implies, a new feature was added to the structural organization of the method explained in the Response condition. Similar to the Stimulus location condition, pictures appeared in four different locations, but in this case, their appearance did not vary sequentially. Locations alternated randomly, but each category was associated with a high-frequency location and three low-frequency locations. Members of a given

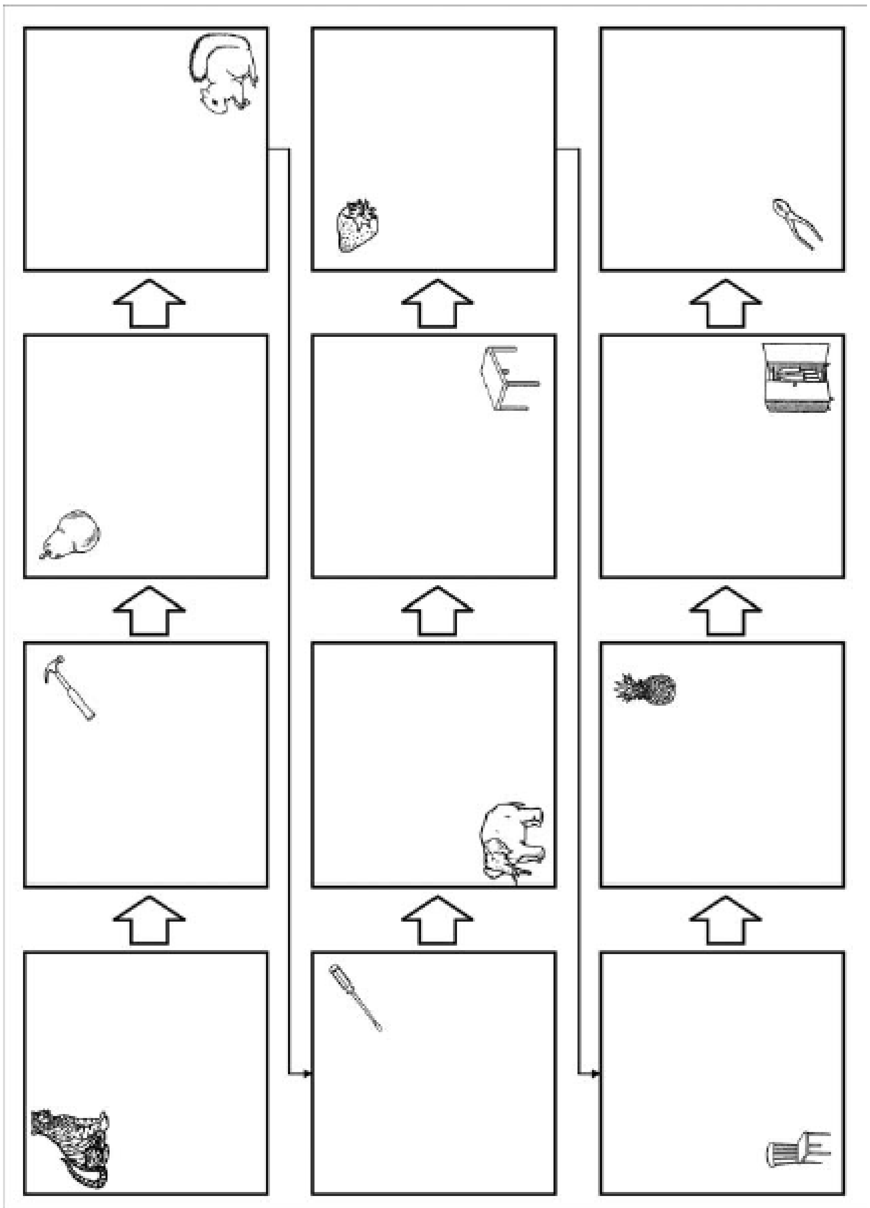


FIGURE 1. Design of the Stimulus location condition presenting the location sequence.

TABLE 1. The Design of the Extra and Frequency Location Conditions

Variable	Furniture (%)	Mammals (%)	Fruits (%)	Tools (%)
Location 1: Top left corner	55	15	15	15
Location 2: Top right corner	15	55	15	15
Location 3: Bottom left corner	15	15	55	15
Location 4: Bottom right corner	15	15	15	55

category appeared at the high frequency location in 55% of all appearances, while the remaining 45% pictures of that category were equally distributed among the remaining three low-frequency locations (15–15%). The overall design of the Extra condition is illustrated in Table 1. This way, each location had one high frequency (55%) and three low-frequency categories (15–15%). The task of the participants was the same as in the previous two conditions: to press the button that corresponded to the category of the presented image. By comparing performance in the Extra condition and performance in the Response condition, we wanted to test whether location frequency affects response sequence learning.

The Frequency Control condition was designed as a control to test the effect of location frequency without the presence of a response sequence. In this condition, we kept the location frequency information of the Extra condition, but eliminated the sequential structure from responses. The task again was to press the button that corresponded to the category of the given picture. Pictures were presented in the four corners of the screen. All categories had a high-frequency location, which was the place where the given category appeared in 55% of all appearances, and three low-frequency locations, where pictures of the category appeared in 15–15%.

In the three conditions with sequential structure (Response, Stimulus location, and Extra) each sequence consisted of 12 button-presses, and each block contained 10 sequences. There were 6 blocks in the experiment. In Blocks 1–4 and Block 6, the repeating sequence was present, while in Block 5, stimuli appeared in random order. Note that the repeating sequence was a stimulus sequence in the Stimulus location condition, and a response sequence in the Response and Extra conditions. The Frequency control condition consisted of 6 blocks of 120 pictures appearing randomly, with the location frequency constraint previously explained.

In all conditions, target stimuli were on screen until participants responded. After the response, there was a 250-ms delay before the next target stimulus appeared on the screen. Repetitions were not allowed in any of the random arrangements—that is, random locations, random appearance of categories, or random appearance of target stimuli within a category.

Apparatus

Participants were tested individually in two separate cubicles at the Cognitive Science laboratory of the Budapest University of Technology and Economics. There was a computer in each cubicle. One of the computers was a 3.00 GHz Intel Pentium 4. Stimuli were presented on a Samsung SyncMaster 796MB (17 in.) monitor, and responses were collected with a Genius K295 PS2 keyboard. The other computer was a 3.00 GHz Intel Core Duo. Stimuli were presented on a Samsung SyncMaster 750B (17 in.) monitor, and responses were collected using a Genius K295 PS2 keyboard. Distance was not measured, but participants sat approximately 60 centimeters from the screen. Both computers ran Windows XP SP3 operating systems. Experiments were run on E-prime 1.1 (Psychology Software Tools, Inc., Pittsburgh, PA).

Results

As accuracies showed a ceiling effect (above 96% on all conditions), no further comparisons were made on the percentage of correct answers.

Reaction Times

Reaction time (RT) analyses were conducted on median RTs for correct responses and computed for each block in each participant.

General Learning Effect

A 3 × 4 repeated measures analysis of variance (ANOVA) was conducted to explore the general learning effect. The ANOVA was conducted with Condition (Stimulus location vs. Response vs. Extra) as between-subject, and Block (1–4) as within subject variables. Results revealed a significant main effect of Block, $F(3,80) = 139.160$, $\eta_p^2 = 0.632$, $p < 0.001$, a significant main effect of Condition, $F(2,81) = 10.493$, $\eta_p^2 = 0.206$, $p < 0.001$, while the Block × Condition interaction was not significant, $F(6,78) = 0.306$, $p = 0.934$. Reaction times by Block by Condition are shown in Figure 2.

Post hoc tests with Bonferroni adjustment revealed that there was a significant difference in general learning between Response and Stimulus location ($p < 0.001$), while the other two comparisons were not significant ($ps > 0.071$).

Bonferroni adjusted post hoc tests revealed that performance on Block 1 significantly differed from performance of all other blocks ($p < 0.001$), Block 2 differed significantly from Block 4 ($p < 0.05$), while all other differences remained nonsignificant (all $ps > 0.338$).

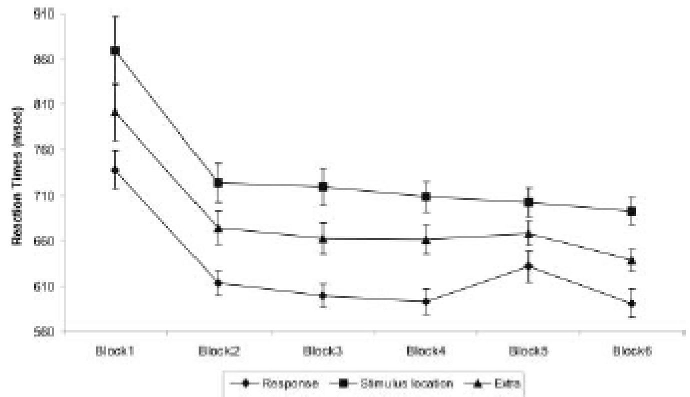


FIGURE 2. Means of median reaction times and standard errors across blocks for the Response, Stimulus location, and Extra conditions.

Sequence Learning Effect

Sequence-specific reaction time change was tested with the comparison of the random block (Block 5) and the two surrounding sequence blocks, Blocks 4 and 6. To test sequence-specific learning, a 2×3 repeated measures ANOVA was conducted with Block (Random block vs. Sequence blocks) as within and Condition (Stimulus location vs. Response vs. Extra) as between-subject variable. The ANOVA revealed a significant main effect of Block, $F(1,82) = 35.242$, $\eta_p^2 = 0.303$, $p < 0.001$, a significant main effect of Condition, $F(2,81) = 9.610$, $\eta_p^2 = 0.192$, $p < 0.001$, and a significant Block $\times$ Condition interaction, $F(4,78) = 11.632$, $\eta_p^2 = 0.223$, $p < 0.001$. Post hoc tests with Bonferroni adjustment revealed that the Response condition significantly differed from the Stimulus location condition ($p < 0.001$), while the other two comparisons (Stimulus–Extra and Response–Extra) were not significant (both $ps > 0.097$). Figure 3 shows the mean median reaction times and standard errors of the random and surrounding sequence blocks for the three conditions.

To disentangle the Block $\times$ Condition interaction, paired-sample t-tests were conducted for each condition. Performance on Block 5 (random) and Block 4 and 6 (sequence) was compared. Paired-sample t-tests showed a significant difference in the Response, $t(29) = -6.081$, $p < 0.001$, and the Extra conditions, $t(23) = -3.130$, $p < 0.01$. The difference between Blocks 5 and Blocks 4 and 6 was not significant in the Stimulus location condition, $t(29) = -0.342$, $p = 0.735$.

Since there were two conditions showing a sequence-specific increase in reaction times, a 2×2 repeated measures ANOVA was conducted to test whether participants showed the same degree of sequence learning in the Response and

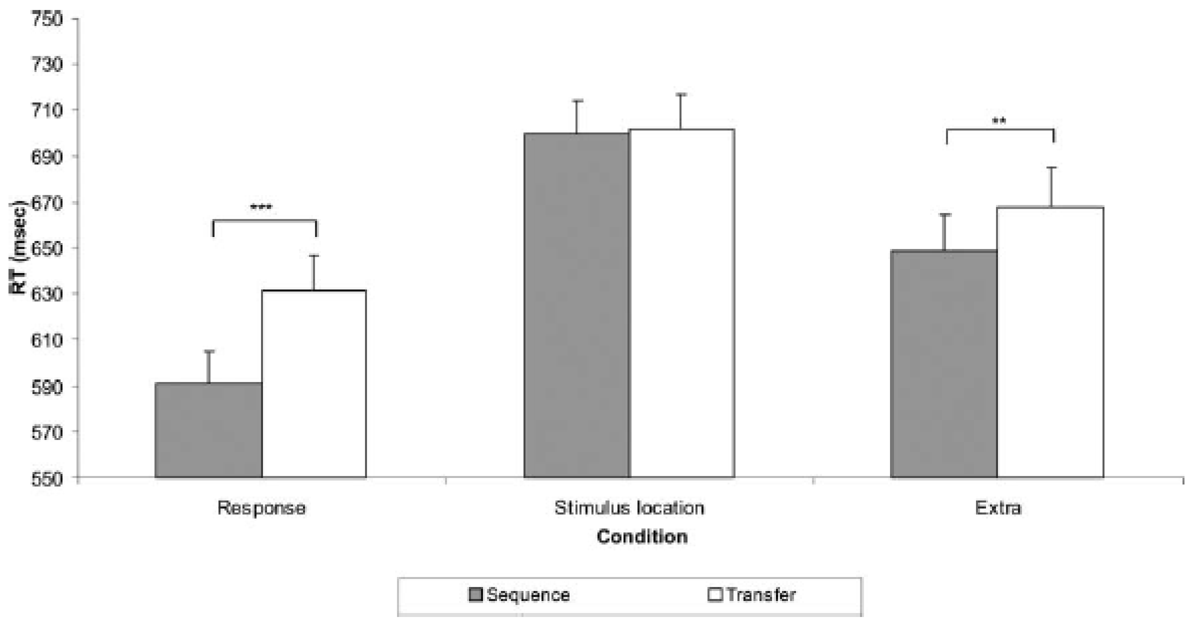


FIGURE 3. Means of median reaction times (RTs) and standard errors for the random block and the surrounding two sequence blocks in the Response, Stimulus location, and Extra conditions.

Extra conditions. The ANOVA was conducted with Condition (Response vs. Extra) as between and Block (Block 5 vs. the surrounding Blocks 4 and 6) as within subject variable. The ANOVA revealed a significant main effect of Block, $F(1{,}53) = 41.665$, $\eta_p^2 = 0.445$, $p < 0.001$, a significant main effect of Condition, $F(1{,}53) = 5.376$, $\eta_p^2 = 0.094$, $p < 0.05$, and a significant Block $\times$ Condition interaction, $F(2{,}52) = 5.505$, $\eta_p^2 = 0.096$, $p < 0.05$.

Location Frequency Effect

The Extra and the Frequency control conditions were similar in that they both manipulated the location frequency of the stimuli. The two conditions differed from each other in the presence vs. absence of a repeating sequence. In the Extra condition, there was a repeating response sequence, whereas there was no sequence in the Frequency control condition. A $2 \times 2 \times 6$ ANOVA was conducted to test whether the presence of the sequence affected the effect of location frequency. Within-subject variables were Block (Blocks 1–6) and Location frequency (High vs. Low), while the between-subject variable was Condition (Extra vs. Frequency control). Results showed a significant main effect of Block, $F(5{,}45) = 50.969$, $\eta_p^2 = 0.515$, $p < 0.001$, a significant main effect of Condition, $F(1{,}49) = 4.399$, $\eta_p^2 = 0.084$, $p < 0.05$, a significant main effect of Location frequency, $F(1{,}49) = 28.044$, $\eta_p^2 = 0.369$, $p < 0.001$, and a significant Location frequency $\times$ Condition interaction, $F(2{,}46) = 5.256$, $\eta_p^2 = 0.099$, $p < 0.05$, see Figure 4. All other interactions were non-significant (all *ps* > 0.231).

Discussion

General Learning

Results showed that although all conditions containing a repeating sequence showed a continuous decrease in reaction times, only the first block differed from all other blocks significantly, as well as the Block 2 from Block 4.

All three conditions differed from each other significantly. Mean median reaction times were smallest for the Response condition and largest for the Stimulus location condition. RT changes in the sequence blocks are twofold. The decrease of reaction times is partly due to general motor learning. This is a general acceleration of responses, not a sequence-specific effect. This general effect is complemented by a sequence-specific shortening of RTs, as participants learn the sequence, they are getting better and faster in predicting the next step. Based on the sequential organization, this might suggest that predictability of stimuli is highest for the Response condition, second best for the Extra condition, and lowest (not present) for the Stimulus location condition.

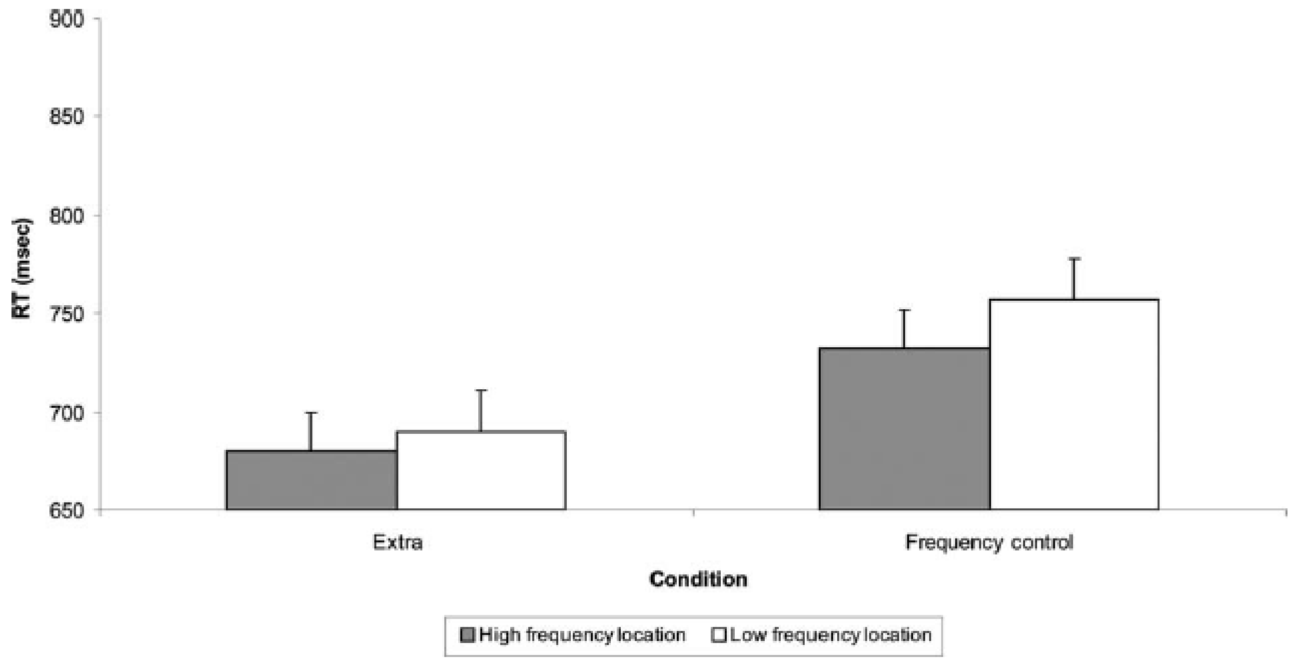


FIGURE 4. Means of median reaction times (RTs) and standard errors for high and low location frequencies in the Extra and Frequency control conditions.

Sequence Learning

RTs on the random and sequential blocks did not differ for the Stimulus location task. If the sequence is only present at the perceptual level, but not in the responses, learning may not take place in a short, six-block deterministic SRT task. This result suggests that sequence patterns are not necessarily extracted in all conditions when there is a sequence present at the perceptual level. This result does not confirm that pure perceptual learning takes place for all types of sequential stimulus information (Remillard, 2003).

There was a difference between reaction times in the random block and the surrounding sequence blocks in both the Response and Extra conditions. The Response condition basically replicated earlier results on the Serial Reaction Time task (e.g. Nissen & Bullemer, 1987). The difference between Nissen & Bullemer's study and ours is that in the present study, there was no one-to-one correlation between the perceptual and motor information, and perceptual information had no sequential structure. It is not clear whether the learning effect on this version of the task was due to pure response learning, since there was also a repeating sequence at the conceptual level, in which different categories followed each other in a fixed order, which could also contribute to sequence learning (Goschke & Bolte, 2007). These results show that sequence learning in the SRT task takes place in the absence of stimulus-based sequential perceptual information, suggesting that stimulus-based sequence information is not necessary for sequence learning.

Sequence Learning and Frequency Effect in the Extra Condition

The Extra condition confirms the findings of the Response condition—that is, that sequence learning may take place in the absence of stimulus-based perceptual sequences. The comparison of the two conditions revealed that the effect of sequence learning is less pronounced in the Extra condition than in the Response condition. This is due to the introduction of a perceptual structure: a location frequency effect.

The link between sequence learning and location frequency is not unidirectional. Comparing the Extra condition with the Frequency control condition revealed that the presence of a response sequence reduces pure frequency effect.

It seems plausible that the introduction of a new noise variable interferes with sequence learning, and reduces learning. Note that in the current setting, varying the location of stimuli can be considered noise in terms of sequence-learning, and a response sequence can be considered noise in terms of location frequency. But most studies of sequence learning suggest that learning on the SRT task does not require attentional effort (e.g., Jiminez & Mendez, 1999; Kelly et al., 2003). A more plausible explanation of the data is offered below.

Sequence Learning in General

All three experimental conditions argue against the hypothesis that sequence learning in the SRT task is necessarily based on perceptual information only (Remillard, 2003). Results of the Response and Stimulus conditions, however, are in concert with the motor-based learning hypothesis (Willingham et al., 2000). At the same time, the Extra condition argues for potential effects of perceptual information on sequence learning. Taken together, results of the current study suggest a more complex picture of learning on the Serial Reaction Time task, arguing against the exclusive effect of either perceptual or motor information. At the same time, we do not suggest that the two factors contribute to learning equally in all circumstances. While response learning may take place without a congruent stimulus sequence, it can be affected by irrelevant perceptual structure. The nature and size of these effects may depend on a number of factors that may be coded in the experimental design, including relative saliency of the stimulus information or the dimension of the stimuli that the participants have to respond to. Our results suggest that the motor learning component is more robust in the case of the early stages of the SRT task when participants have to respond by button presses, while sequential perceptual information does not necessarily lead to sequence learning by itself in this paradigm. At the same time, under specific arrangements, perceptual information does have an important effect on learning. For sequence learning to take place, perceptual and response information do not need to be correlated, but if some dimension of the perceptual information serves as a strong and salient predictive cue (as location frequency proved to be), it may interfere with the motor sequence and degrade motor sequence learning. This interference may be based on predictability. The location of the target stimulus in the Stimulus location condition is irrelevant information in the sense that it does not help the prediction of the button-press. In the Extra condition, location information serves as a probabilistic cue to the target category. The location of the target stimulus helps the response since participants may predict which category is more likely to appear at a given location. This may be true even if the location frequency effect seems to be smaller in the presence of a response sequence—for example, in the case of the Extra condition.

Our study hopefully adds an important contribution to the debate on the nature of sequence learning: perceptual sequences are neither always necessary (Response condition) nor always sufficient (Stimulus condition) for sequence learning. Importantly, our results also confirm earlier research suggesting that perceptual information can and does affect response learning. Whereas earlier studies suggested that convergent stimulus and response information may both be learned to the same extent (Németh et al., 2009) and may interact in a way that leads to higher efficacy of sequence learning (Robertson et al., 2001), our results show that if these types of information diverge, then learning decreases in both domains. Together with previous findings in the literature, these data show that the extent to which perceptual and motor information affect learning may depend on the task, the nature of the required response, and the organization of stimuli.

More research is needed to find out the nature of the interaction between stimulus and response information, and to find out which are the stimulus dimensions that affect response sequence learning. Is this interaction based on predictability, or does it rather depend on the salience of the given dimension? Research on salience and predictability may also be important in terms of stimulus learning: Would it be possible to vary either of the two dimensions in order to achieve the learning of a stimulus sequence? Is it possible that the lack of stimulus-based learning in the current paper is only due to a special combination of stimulus dimensions, or is stimulus-based learning missing in all circumstances in such a short, deterministic SRT task? How does domain and presentation mode of stimuli and responses affect learning? All these questions are open to further research.

NOTE

1. The line drawings are for illustration (drawings are collected from the International Picture Naming Project database: http://crl.ucsd.edu/~aszekely/ipnp/1studies.html, Bates et al., 2003); in the actual experiment color photos were used.

AUTHOR NOTES

Ferenc Kemény is currently a junior research fellow at the Department of Cognitive Science at the Budapest University of Technology and Economics. His research deals with the effect of stimulus characteristics and organization on sequence learning and probabilistic category learning. **Ágnes Lukács** is an associate professor at the Department of Cognitive Science of the Budapest University of Technology and Economics. Her main interests are language acquisition and language processing in typical and atypical development, with a special focus on the relationship between language and nonlinguistic cognitive abilities.

REFERENCES

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin and Review*, *10*(2), 344–380.

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

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

Deroost, N., Zeeuws, I., & Soetens, E. (2006). Effector-dependent and response location learning of probabilistic sequences in serial reaction time tasks. *Experimental Brain Research*, *171*(4), 469–480.

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

Goschke, T., & Bolte, A. (2007). Implicit learning of semantic category sequences: Response-independent acquisition of abstract sequential regularities. *Journal of Experimental Psychology-Learning Memory and Cognition*, *33*(2), 394–406.

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (2002). Motor sequence learning with the nondominant left hand: A PET functional imaging study. *Experimental Brain Research*, *146*(3), 369–378.

Howard, J. H., Howard, D. V., & Mutter, S. A. (1992). Serial pattern learning by event observation. *Journal of Experimental Psychology-Learning Memory and Cognition*, *18*(5), 1029–1039.

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

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

Kelly, S. W., Burton, A. M., Riedel, B., & Lynch, E. (2003). Sequence learning by action and observation: Evidence for separate mechanisms. *British Journal of Psychology*, *94*(3), 355–372.

Mayr, U. (1996). Spatial attention and implicit sequence learning: evidence for independent learning of spatial and nonspatial sequences. *J Exp Psychol Learn Mem Cogn*, *22*(2), 350–364.

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

Németh, D., Hallgató, E., Janacsek, K., Sándor, T., & Londe, Z. (2009). Perceptual and motor factors of implicit skill learning. *Neuroreport*, *20*(18), 1654–1658.

Nissen, M. J., & Bullemer, P. (1987). Attentional Requirements of Learning—Evidence from Performance-Measures. *Cognitive Psychology*, *19*(1), 1–32.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay in the cognitive unconscious*. Oxford, England: Oxford University Press.

Remillard, G. (2003). Pure perceptual-based sequence learning. *Journal of Experimental Psychology-Learning Memory and Cognition*, *29*(4), 581–597.

Robertson, E. M., Tormos, J. M., Maeda, F., & Pascual-Leone, A. (2001). The role of the dorsolateral prefrontal cortex during sequence learning is specific for spatial information. *Cerebral Cortex*, *11*(7), 628–635.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition*, *27*(3), 561–572.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology-Learning Memory and Cognition*, *15*(6), 1047–1060.

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

Original manuscript received June 12, 2010
Final version accepted November 15, 2010

9. Study 4: Self insight in Probabilistic Category Learning

Kemény, F., & Lukács, Á. (in press). Self-insight in Probabilistic Category Learning, *Journal of General Psychology*.

Abstract

The Weather Prediction (WP) task is one of the most extensively used Probabilistic Category Learning tasks. Although it has been usually treated as an implicit task, its implicit nature has been questioned with focus on the *structural knowledge* of the acquired information. The goal of the current studies is to test if participants acquire explicit knowledge on the WP task. Experiment 1 addresses this question directly with the help of a subjective measure on *self-insight* in two groups: an experimental group facing the WP task, and a control group with a task lacking predictive structure. Participants in the experimental group produced more explicit reports than the control group, and only on trials with explicit knowledge was their performance higher. Experiment 2 provided further evidence against the implicitness of the task by showing that decreasing stimulus presentation times extends the learning process, but does not result in more implicit processes.

Keywords: probabilistic category learning, multiple-cue learning, self-insight, implicit versus explicit learning

Self-insight in Probabilistic Category Learning

Forecasting the weather is not an easy task – especially for lay people: a number of different factors may be considered. There are several cues that may help us in proper prediction, or may make our prediction more difficult. These cues include brightness, clouds, wind, the colours that appear, the humidity we feel, even the smell of the air, or maybe the pain in our back and a number of other cues. Still, people can differentiate between dark cloudy days, and may more or less properly predict whether it will be raining. If one is asked how they managed to make a proper prediction, they may come up with some kind of explanation highlighting one or another cue, or might simply just say they worked from intuition.

The Weather Prediction (WP) task models such inferences based on multiple cue probability learning – but in fact has nothing to do with real weather. It is designed for proper control on a set of cues that are included: there are four cues, each of them predicting a specific outcome with a given probability, and the outcome is to be predicted based on the cues that appear. Participants are shown 1, 2 or 3 out of four possible cues, then they are prompted to decide whether, based on the cues, there would be rain or sunshine. The cues of the WP task are arbitrary cues that are not linked to the weather in real life: usually tarot cards (Gluck, Shohamy, & Myers, 2002), pictures (Knowlton, Squire, & Gluck, 1994), geometric shapes (Keri, Szlobodnyik, Benedek, Janka, & Gadoros, 2002), or fragments of line-drawings (Kemény & Lukács, 2009). The link between cues and outcomes is based on their statistical associations. The statistical association is based on feedback provided throughout the task. Each cue has a different predictive value: there are usually two strong cues, one predicting each outcome with a high association rate, and two weak cues, one predicting each outcome with a low association rate. As participants in the beginning have basically no idea on what they have to do, they guess. After each answer the proper response is revealed: following the decision they are shown for example that the appearance of a triangle and a rhombus in fact led to rainy weather. Participants usually face four blocks of 50 items. Performance on Block 1 is generally only slightly above chance level, whereas in Block 4 it could reach even 80% (Knowlton, Mangels, & Squire, 1996).

The implicit/procedural versus explicit/declarative aspects of the WP task

The WP task was first used in the framework of the explicit/declarative vs. implicit/procedural distinction, with clinical groups being impaired on one or the other memory systems. Results from amnesic patients revealed that during the early phases of the task their performance is identical to healthy control participants', whereas healthy participants have a significant advantage on the later phases (Knowlton et al., 1994). Parkinson's patients show a deficit on the WP task already in the early phases, and their performance hardly rises above chance level even later (Knowlton, Mangels et al., 1996). Also, amnesic patients were shown to be unable to answer debriefing questions about the task, whereas Parkinson's patients were able to give correct answers to questions like 'How many cues could have appeared on the screen simultaneously?' (Knowlton, Mangels et al., 1996). Since patients with a procedural deficit (PD patients) show impairment in the early phases of the task, and patients with declarative deficit (amnesic patients) show a deficit in the later phases of the task, solving the WP task is concluded to rely on the procedural system in the early stages, and the declarative system in the later phases (Knowlton, Mangels et al., 1996). This was further confirmed by an imaging study showing that there is a rapid MTL deactivation in the beginning of the task suggesting the lack of MTL activation, hence a lack of declarative functioning (Poldrack, Prabhakaran, Seger, & Gabrieli, 1999).

Later studies, instead of measuring the overall categorization performance, focused mainly on the way participants solve the task. Three basic strategies were identified initially

(Gluck et al., 2002). The One-cue strategy is a sub-optimal strategy when participants focus on one cue. If the cue in focus is present, participants give a consistent answer, while in the absence of the cue their response is random. The Singleton strategy is also a sub-optimal strategy. Singleton strategy users only give consistent answers if one cue is present at a time. In the case of cue-combinations they respond randomly. The third strategy, called Multi-cue strategy, is the optimal strategy. Multi-cue strategy users focus on all cues that are present at the same time, and they respond according to the combined predictive values of the cues. That is if cue 1 and 3 are present at the same time, the prediction is made based on the average predictive value of the two cues: (85,7% + 30%) / 2 = 57,85%. If the combined value is above 50%, the expected result is SUN, while if it is below 50%, the expected result is RAIN. Note that the first two strategies may be grouped together as they require focus on only one cue at the time (Single strategies), whereas the Multi-cue strategy requires focus on several cues simultaneously.

Gluck and colleagues (2002) also showed that these strategies are implicit in the sense that participants' self-report on strategy use do not match their real performance. They also showed that strategy use throughout the task changes. In the earlier phases participants are more likely to use one of the Single strategies (One-cue or Singleton), while in the later phases more and more participants manage to develop a Multi-cue strategy. Based on two assumptions, Gluck and colleagues (2002) hypothesize that Single strategies rely on the declarative system, whereas multi-cue strategy relies on the procedural system. On the one hand, single strategies are easy to verbalize, hence it is most likely that participants will be aware of their strategy use. At the same time it is difficult to verbalize multi-cue usage, hence it is thought to be implicit. On the other hand, imaging data show that the earlier hypothesized MTL deactivation might not in fact be significant, suggesting that there is MTL activation in the earlier phases of the task (Poldrack et al., 2001).

Strategy use was also tested in clinical populations. Results showed that Parkinson's patients (Shohamy, Myers, Onlaor, & Gluck, 2004) were unable to switch to Multi-cue strategy, whereas hypoxic patients (with hippocampal malfunctions, Hopkins, Myers, Shohamy, Grossman, & Gluck, 2004) were unable to develop any strategies. These results are in concert with the hypothesis that the declarative versus procedural nature of solving the WP task differs by strategy use.

Either due to a priori definition – and comorbidity with declarative deficits – (Knowlton et al., 1994; Knowlton, Squire et al., 1996; Hopkins et al., 2004), or due to dissociation with verbal reports (Gluck et al., 2002), early studies suggested that the WP task is at least partially implicit. On the other hand the implicit nature of the WP task has been questioned by a number of papers – on three grounds. In a meta-analysis Zaki (2005) suggested that amnesic patients do in fact show deficits on categorization. Lagnado, Newell, Kahan and Shanks (2006) tested structural knowledge both blockwise and itemwise and found that participants are able to report each cue-strength and differentiate between cues based on their importance already in the early phases. Newell, Lagnado and Shanks (2007) and Price (2009) used dual-task paradigms, and found that concurrent tasks that are expected to reduce implicit performance do not affect learning, whereas concurrent tasks that are expected to reduce explicit hypothesis testing caused a serious decay in categorization performance. As the latter three studies used the WP task, we will discuss them in detail in the following section.

In their first experiment, Lagnado et al. (2006) asked participants to provide the experienced predictive values of each cue, and to report the subjective importance of a specific cue in the predictions. Results showed that – similarly to suggestions by Rescorla and Wagner (1972) – participants were able to report the predictive values of stronger cues already in the first block of 50 trials, whereas predictive values of weaker cues were only accessible later (probability ratings). Subjective measures also showed that as the task progressed, participants reported that they relied more on stronger cues (cue-usage ratings). Trial-by-trial subjective reports in Experiment 2 showed the same results: participants first learned to use stronger card, and from Block 2 they rated the stronger cards more important (page 171 & Figure 11 of Lagnado et al., 2006).

Newell et al. (2007) used a dual task design, and found that performance on the Weather Prediction task decreased due to the presence of a concurrent numerical stroop task, suggesting that the task is declarative. In Experiment 2, Newell et al. used the same trial-by-trial reports employed by Lagnado et al. (2006) for both a declarative (observational) and a procedural (feedback) version of the task. The procedural version was identical to the one introduced earlier, whereas in the declarative version, participants were shown cues and outcomes simultaneously, and were instructed to memorize the associations[1]. Results showed that participants of the observer and feedback groups showed a similar performance, and also their cue-ratings were very similar. In term of cue-reliance, both groups reported that they rely on strong cues more than weak cues very early (Block 6 in the observation and Block 9 in the feedback task, with 5 trials in each block). Newell et al. suggest that this rather small difference should not be interpreted as evidence for separate systems underlying the different tasks, but rather that both tasks rely on the same system, with the observation task being somewhat easier. In sum, neither Lagnado et al. (2006), nor Newell et al. (2007) suggest that learning on any form of the WP task is implicit. Note that both the Lagnado et al. (2006) and the Newell et al. (2007) studies focused on the perception and recollection of cue-outcome

[1] Note that the method was introduced by Shohamy et al (2004). The basic difference is that Shohamy et al considered the observational WP task a procedural method due to the fact that the link between cues and outcomes are probabilistic, and showed that Parkinson's patients are only impaired on the feedback-based WP task, which is identical to the procedural WP task of Newell et al (2007).

contingencies, a task that was in part correctly solved by amnesic patients (Reber, Knowlton, & Squire, 1996).

Two experiments by Price (2009) tested whether manipulations designed to disrupt implicit (Exp1) or explicit (Exp2) learning interferes with learning on the WP task. In Price's Experiment 1, the delay between cue and feedback was manipulated, as earlier studies suggested that if the cue-feedback units are broken, implicit categorization declines, but explicit hypothesis testing remains intact (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005). In Experiment 2, feedback processing was disrupted: a short, one-item memory-scanning task was administered either immediately (short feedback condition) or after 2500 msec (long feedback condition) following feedback presentation. This way there was either a 0 or 2500 msec window for feedback processing. Previous research on disrupted feedback processing showed that it impairs explicit hypothesis testing, but does not affect implicit categorization (Maddox, Ashby, Ing, & Pickering, 2004). Results showed that delayed feedback had no effect on learning on the WP task, whereas feedback-processing disruption caused a serious decrease in categorization performance. In line with Lagnado et al. (2006) and Newell et al. (2007) this result suggests that the WP task relies on explicit processes.

In sum, there are three different hypotheses outlined in the earlier literature based on the implicit versus explicit nature of the WP task. The implicit-first theory suggests that the early stages of the WP task are implicit, while it becomes explicit in the later blocks (Knowlton, Mangels et al., 1996; Knowlton et al., 1994). The strategy-theory suggests that Single strategies are explicit, whereas Multi-cue strategy use is implicit (Gluck et al., 2002; Meeter, Myers, Shohamy, Hopkins, & Gluck, 2006), while, based on experimental psychological results, the explicit-theory suggest that the task is explicit (Lagnado et al., 2006; Newell et al., 2007; Price, 2009)

Structural knowledge versus Self-insight

So far, research has focused on structural knowledge (task knowledge), and its conscious nature. Structural knowledge is basically the knowledge that leads to one answer or another (Dienes & Scott, 2005). A piece of structural knowledge may be that in the presence of Cue1 the weather will be sunshine. These bits of structural knowledge may be conscious or unconscious. A piece of unconscious structural knowledge is when one is not aware of the fact that Cue1 leads to sunshine, yet performance suggests that this knowledge lies behind performance. On the other hand, structural knowledge is conscious when one provides answers based on a piece of structural knowledge, and able to report it. Reporting how a prediction relies on a cue is also a piece of structural knowledge. These two measures, probability ratings (the perceived probability of a given cue or cue-combination, explained earlier) and cue-usage ratings (the subjective reliance on the appearing cues) were measured in papers by both Lagnado et al. (2006) and Newell et al. (2007).

Self-insight (Lagnado et al., 2006) on the other hand concerns the reported access to the learnt information. According to Rosenthal (1986; 2005), a mental state may be conscious if we are conscious of being in that mental state. If we have a (higher order) thought on the mental state (Dienes & Scott, 2005), we do not only behave as if we had structural knowledge (this would be the first-order representation), but we also report to have and use it. As there have been previous studies on the structural knowledge of participants of the WP task, the goal of the current paper is to test self-insight, to obtain subjective measures on how participants made their decision, and to what extent are they aware of that given decision. Participants of the WP task are asked to classify whether their decision was based on GUESSING, INTUITION, whether they report 'I think I know the answer' (THINK answer), or they suggest to rely on remembering (REMEMBER answer) or knowledge of the rule (RULE answer). If the number of responses is higher in the first three categories that may indicate less explicit knowledge, while if participant report to rely on remembering or rule-knowledge that may be an indicative of more explicit knowledge. Note, that Dienes and Scott (2005) also used 5 categories, but instead of the ‘I think I know the answer’ their categories included a ‘pre-existing knowledge’. They did not report results on pre-existing knowledge, as there were no answers within that category. We included the ‘I think I know the answer’ type of response to make the transition smoother between the statements of ‘Intuition’ and ‘I remember’ answers.

We present two experiments testing the implicitness of learning on the WP task by measures of self-insight. Experiment 1 employed a modified version of the Weather Prediction task, with more extreme predictive values (Kemény & Lukács 2010) together with a control condition where stimulus-outcome associations were random. In both conditions, we collected subjective data on self-insight after each decision. Experiment 2 was designed to test whether prolonging the learning period (by using shorter presentation times) affect self-insight.

As explained above, there are three competing theories of learning on the WP task. The implicit-first theory (Knowlton et al., 1994; Knowlton, Squire et al., 1996), the strategy-theory (Gluck et al., 2002; Meeter et al., 2006; Meeter, Radics, Myers, Gluck, & Hopkins, 2008) and the explicit theory (Lagnado et al., 2006; Newell et al., 2007; Price, 2009). The implicit-first theory predicts that learning is implicit in the beginning, and becomes explicit later. According to the implicit-first theory, participants of both conditions should provide similar self-insight reports in the earlier phases of the task, while they should differ in later phases. In terms of performance, the implicit-first theory predicts that in the early phases, participants show higher performance with implicit self-report, whereas in the later phases they show higher performance with explicit self-report.

The strategy-theory predicts self-insight based on different strategies. It predicts single strategy users to provide more explicit self-reports than multi-cue users. As single strategies develop earlier in time than multi-cue strategy, strategy theory also predicts more explicit self-reports to appear in the early phases of the task, than in the later phases.

In the distribution of answers, the explicit theory predicts a similar pattern than the implicit-first theory: participants provide more implicit answers in the beginning than in the later phases. However, the explicit theory predicts that participants' performance goes above chance only for those items where decision is associated with an explicit self-report. Our experiments contrasts the three theories, and tests whether either of the predictions fit the data observed.

Experiment 1

Method

Participants. Altogether 56 subjects (26 female and 30 male) participated in Experiment 1. They were randomly assigned into either the Experimental or the Control condition, with 29 participants in the former and 27 in the latter. The mean age of participants of Experiment 1 was 21.44 years (SD = 1.68 years). All participants were recruited from the Budapest University of Technology and Economics, and participated for credit points. All participants provided a written informed consent, in accordance with the principles set out in the Helsinki Declaration and the stipulations of the local Institutional Review Board.

Procedure. All participants completed a computerized version of the Weather Prediction task. The task ran on E-prime 1.2 (Psychology Software Tools Inc., Pittsburgh, PA). First, participants received their instructions:

“Hi,

you will be the weather forecaster. You will see strange pictures, and your task will be to decide whether it will be SUNSHINE or RAIN! Click the icon that corresponds to your prediction.

After your choice the computer reveals what the weather really was. After that, please report how sure you were in your judgement: you will see a line, and your task is to click to the point that best characterizes your decision. We defined five points to help your decision.

Press any key to continue!”

In each item participants received 1, 2 or 3 out of four cues. Their goal was to predict whether there will be sunshine or rain. The cues were a square (Cue1), a triangle (Cue2), a pentagon (Cue3) and a rhombus (Cue4). A 640 X 480 display setting was used, the size of the cues were identical, they were fit into a 120 X 120 square with a narrow white border. The cues were always presented vertically 96 pixels from the top. If a cue appeared alone, it was situated in the vertical centre line, in the case of two cues, the cues were on the two sides of the vertical centre line, while if three cues were present, the central cue was located in the centre line and the two other cues appeared on each side. The icons of the two outcomes were also present: two slightly smaller icons (100 X 100) appearing below the cues, 255 pixels from the top. The SUN icon was always on the left side, 219 pixels from the left edge, whereas the RAIN icon appeared on the right, 423 pixels from the left edge of the screen. If the participants predicted sunshine, they had to click on the SUN icon using a two-buttoned mouse, if their prediction was rain, they had to click on the RAIN icon. The cue (or cue-combination) was present until the participant responded, and until the feedback was present. Immediately after response, a feedback was given: only the icon of the correct answer remained on screen with the cues. The feedback only revealed the outcome; the participants' choice did not appear on the screen. Note that as cues and outcomes have a probabilistic relationship, outcomes are not necessarily the same as the expected correct answers. The feedback was presented for 1500 msec along with the cue(s), then it disappeared.

After feedback a question appeared: 'How sure were you in your judgement?'. Below the question there was a continuous line with five statements above the line. The role of the participants was to click on the line below one of the statements. The five statements were the following (left to right): 'I was guessing', 'Intuition', 'I think I knew it', 'I remembered the answer' and 'I know the rule' (Since participants were native speakers of Hungarian, both the question and the five statements appeared in Hungarian). These are conventional, everyday statements in Hungarian, so further instructions were not given. In general, Intuition seems to be the most problematic one; however its Hungarian equivalent is "Megérzés", which literally means a decision based on gut feeling without rational arguments. As soon as participants clicked on the line, the question disappeared, and a new item (cue or cue-combination) appeared for prediction.

There were two conditions. The two conditions – Experimental and Control – differed on the predictive values of the cues. While in the case of the Experimental condition the cues and outcomes had a predictive structure, the link between cues and outcomes was at chance level in the Control condition. In the Experimental group, Cue1 (square) was associated with sunshine in 85.7% of its appearances, Cue2 (triangle) led to sunshine in 70% of all its appearances, Cue 3 (pentagon) led to rain in 70%, and Cue 4 (rhombus) was associated with rain in 85.7% of its appearances. Note that in the case of the remaining appearances the cue led to the other outcome: rain for cues 1 & 2 and sunshine for cues 3 & 4. For both conditions there were four blocks, and each block contained 50 predictions. At the same time, in the control condition, in half of the appearances of each cue the outcome was sunshine, while for the other half the outcome was rain. Table 1 shows the predictive values for each cues and cue-combinations in both the Experimental and Control conditions. Note that the structure of the task is identical to that used in Kemény and Lukács (2010), while the subjective measurement is adapted from Dienes and Scott (2005).

Strategy analysis. To avoid repeated analyses on the same data, strategy use was calculated for Block 2. Previous studies of strategy use suggest that strategies do not fluctuate, but there are sudden, clear-cut changes (Shohamy, Myers, Kalanithi, & Gluck, 2008). Strategy analysis though requires a wide window, which obviously causes difficulties in finding the exact point of change. However, similarly to previous tasks, we used a whole block of 50 trials to decide strategy use (Hopkins et al., 2004). There are altogether 6 different strategies: the four one-cue strategies, the singleton strategy and the multi-cue strategy. Strategy analysis is identical to previous studies (Gluck et al., 2002). Each of the six strategies predicted the expected number of 'sun' answers for each pattern differently. The model score for a strategy was computed as the sum of the squared difference between the number of expected sun answers in the pattern and the number of sun answers the participants gave for that specific cue or cue-combination. This score was divided by the sum of squares of the number of presentation of each pattern. The computation of the model score is illustrated in equation (1). A strategy was assigned if the model score was below 0.1 (criterion identical to Gluck et al, 2002).

(1)

$$ModelScore_M = \frac{\sum_P (\#sun\_expected_{P,M} - \#sun\_actual_P)^2}{\sum_P (\#presentations_P)^2}$$

Results

Data analysis. First the number of responses was compared by Block and by Condition to see whether the distribution of self-insight reports were the same in the two groups, and whether the difference changed by time. In the second analysis we compared categorization performance by the accompanying self-insight report, by Block and by Condition. This required participants to have data in all 2 x 5 cells (two blocks and five different types of responses). There were however hardly any participants having all five responses in both blocks. In fact, none of the experimental condition participants, and only 5 of the control participants had values in all 10 cells. This event is not completely unexpected as Dienes and Scott (2005) found the same anomaly. They combined implicit and explicit responses: implicit responses were the ones where participants report that they have no clear representation on the source of their knowledge, whereas explicit responses were the ones where the source of knowledge is identified. This clustered GUESS, INTUITION and THINK answers into the implicit category, and REMEMBER and RULE answers into the explicit category. These labels are hypothesis driven, but will be used throughout the Results section to enhance clarity. 21 Experimental and 23 Control participants had data in all four cells. In the following analyses, only data of these participants were included, except when indicated otherwise.

Learning performance. Learning performance was analysed only on the Experimental group with a repeated-measures ANOVA with Block (1 throught 4) as within-subject design. There was a monotonic improvement in learning performance, revealed by a significant linear trend, $F(1, 20) = 11.996$, $p < 0.01$, $\eta^2_p = 0.375$. Figure 1 illustrates blockwise performance of both the Experimental and Control conditions.

Comparing early and later blocks in terms of self-insight by Condition. Knowlton et al. (1994, 1996) suggests that the early phase of the WP task is procedural and the later phase is declarative, whereas Gluck et al. (2002) surmises the reverse. As the difference is rooted in the implicit nature of the phases, we have decided to compare self-insight between the Experimental and Control conditions by phases. Since there is no clear indication on the identification of phases in previous literature, we have decided to compare the 1st and 4th blocks of 50 items.

Self-insight measures were compared between the Experimental and the Control group on both the early and later blocks. A 2 x 2 x 2 repeated-measures ANOVA was conducted with Block (Early vs. Late) and Answer-type (Implicit vs. Explicit) as within-subject variables, and Condition (Experimental vs. Control) as a between subject variable. Note, that data for Block and Condition are invariant (there are 50 answers in each block, and 100 answers in each condition), so the main effect of Block and Condition, and the Block x Condition interaction are not applicable.

The ANOVA revealed a significant Answer-type x Condition interaction, $F(1, 42) = 16.267$, $p < 0.001$, $\eta^2_p = 0.279$, and a significant Block x Answer-type interaction, $F(1, 42) = 17.847$, $p < 0.001$, $\eta^2_p = 0.298$. Neither the Answer-type main effect ($p = 0.813$), nor the Block x Answer-type x Condition interaction ($p = 0.140$) was significant. Figure 2A shows the number of Implicit and Explicit self-reports by Condition and by Block.

The Answer-type x Condition interaction reflects that the number of implicit self-reports decrease from Block 1 to Block 4, while the number of Explicit self-reports increase.

This increase in Explicit self-reports is numerically the same as the decrease in the Implicit self-reports, as the number of answers in each block is always 50.

To further explore the Answer-type x Condition interaction, separate paired-sample t-tests were conducted for each Condition with Answer-type as within-subject variable. The t-tests revealed that while for the Experimental condition the number of explicit answers were significantly higher, $t(20) = -2.459$, $p < 0.05$, Control participants provided more implicit self-reports, $t(22) = 3.307$, $p < 0.01$.

Categorization performance associated with Self-insight type by Condition and by Block. So far we tested how participants' self-insight differs by answer-type in early vs. later blocks. In the next step, self-insight and performance are directly mapped onto each other. The following ANOVA was conducted on the average categorization performance that was associated with a given category. A 2 x 2 x 2 repeated measures ANOVA was conducted with Block (Early vs. Late) and Answer-type (Implicit vs. Explicit) as within-subject variables and Condition (Experimental vs. Control) as a between subject variable. Results revealed that performance associated with explicit-type answers was higher, confirmed by a significant main effect of Answer-type, $F(1, 42) = 49.304$, $p < 0.001$, $\eta^2_p = 0.540$. A significant main effect of Condition, $F(1, 42) = 29.063$, $p < 0.001$, $\eta^2_p = 0.408$, revealed that the Experimental condition showed higher performance than the Control condition. There was also a significant Answer-type x Condition interaction, $F(1, 42) = 50.330$, $p < 0.001$, $\eta^2_p = 0.545$. No other main effects or interactions were significant (all ps > 0.195).

To further explore the Answer-type x Condition interaction, data in the two blocks was merged, and both performance with implicit answers and performance with explicit answers were compared between the two conditions. A multivariate ANOVA was used with implicit and explicit performance as dependent variables and Condition as between-subject variable.

The MANOVA revealed that performance associated with explicit answers were significantly higher in the Experimental condition, $F(1, 42) = 64.892$, $p < 0.001$, $\eta^2_p = 0.607$, while there was no difference in performance associated with implicit answers, $p = 0.517$. Figure 3A shows the average categorization performance by Answer-type and by Condition.

Self-insight by Strategy use in Block 2. Knowlton et al. (1994, 1996) suggests that the differentiation is strictly time-based, i.e. the task relies on the procedural system in the early and the declarative system in the later blocks. On the other hand, Gluck et al. (2002) suggested that relying on different memory systems is not based on the time passed during the task, but strategy use. Gluck et al. proposed that singleton and one-cue strategy users rely on the declarative system, whereas multi-cue users rely on the procedural system. To avoid overlapping analyses of the same data, Block 2 strategy use was analysed. Only results of the Experimental condition will be considered, since the control condition lacks a predictive structure, hence all strategies yield random answers. Out of the 21 participants whose data was analysed previously, 13 participants were using multi-cue strategy, 5 participants were using one of the single strategies in Block 2. The remaining 3 participants did not use any identifiable strategy, and hence were excluded from the present analysis.

A 2 x 2 repeated-measures ANOVA was conducted with Answer-type as within-subject, and Strategy use as between-subject variable. Once again, the numbers in each strategy are invariant; hence Strategy main effect is not applicable. The ANOVA revealed a significant main effect of Answer-type, $F(1, 16) = 21.792$, $p < 0.001$, $\eta^2_p = 0.577$, and a significant Answer-type x Strategy use interaction, $F(1, 16) = 14.559$, $p < 0.01$, $\eta^2_p = 0.476$. Figure 4 shows the number of Implicit and Explicit answers by Strategy use.

To further analyse the interaction, the two group of strategy users were compared on the number of Implicit answers provided (note that the number of Explicit answers were not analysed, as it equals 50 minus the number of Implicit answers). An independent samples T-test revealed that single strategy users provided significantly more Implicit self-insight reports than multi-cue users, $t(16) = 3.816$, $p < 0.01$.

Categorization performance associated with Self-insight type by Strategy use in Block 2. A 2 x 2 repeated measures ANOVA was conducted with Answer-type (Implicit vs. Explicit) as within-subject variable, and Strategy use (Single vs. Multi-cue strategy) as between subject variable. The dependent variable was WP score; that is the percentage of answers that match the predictive values of the cues (and not whether the participant's answer matched the real outcome). Performance of the Single strategy user group was 34.53% (SE = 5.82%) with implicit self-report, while 88.85% (SE = 2.65%) with explicit self-report. The two values for Multi-cue users were 51.99% (SE = 5.44%) for items without and 91.21% (SE = 1.98%) with reported self-insight. Performance appeared to be higher when explicit self-report was associated with the decision, as confirmed by a significant main effect of Answer-type, $F(1, 14) = 69.872$, $p < 0.001$, $\eta^2_p = 0.833$. Strategy use main effect and Answer-type x Strategy use interaction were not significant (both ps > 0.242).

Discussion

Experiment 1 sought an answer to the question whether higher order knowledge on the decisions develops in the Weather Prediction task, and whether judgements can be considered explicit or implicit in this sense. Results in general show that self-insight on the Weather Prediction task differs from that of a control task: participants of the Experimental conditions gave less Implicit and more Explicit answers than the Control condition. A Block x Answer-type interaction also showed that across trials, participants reported more reliance on memories and rule knowledge, even in the case of the Control condition. This change in answer-type though did not differ by Condition.

First of all, our results show that both groups shifted towards more explicit responses by the end of the task. This is not in line with proposals suggesting that the early blocks of the task can be more explicit than the later blocks. However, these results on their own should be handled with care, as also Control participants reported more explicit answers in the later phases.

The presence of group and block differences even in the absence of an interaction could be interpreted in two ways: learning on the task is strictly explicit, or the earlier blocks of the task are implicit, and explicit knowledge emerges later. To test this hypothesis, performance was mapped onto implicit versus explicit types of self-reports. Results showed that the Experimental group achieved a higher performance, but this higher performance was only due to items with explicit self-report. Items lacking self-insight lead to chance level performance in both conditions. These results suggest that the higher numbers of explicit self-reports in the fourth block are an artefact of the learning task, and reflect a familiarity effect. And there is no emergence in the explicit processes throughout learning: it is already explicit from the early stages.

Gluck et al. (2002) suggest that reliance on the different systems may not be time dependent, but rather depends on the strategy in use. This hypothesis was tested on Block 2 data. Results showed that multi-cue users report more self-insight than single strategy users. Mapping performance and self-insight reports showed that multi-cue strategy users and single strategy users did not differ from each other in average performance. At the same time, only Explicit answers led to above chance performance. The combination of the two types of analysis shows that better learning is achieved with explicit answers, and there are more explicit answers in multi-cue strategy. This result suggests that both strategies rely on explicit processes, but Multi-cue strategy users provide numerically more explicit predictions. In general, these results are not in line with the strategy hypothesis (Gluck et al, 2002), which predicts more explicit decisions in the Single strategies.

In sum, results showed that learning on the WP task cannot be characterized as an implicit process. In fact, analyses revealed that neither time-based, nor strategy-based differences in task-specific awareness can be observed. This result is in line with earlier experimental psychological studies of the WP task (Lagnado et al., 2006; Newell et al., 2007; Price, 2009). Conclusions from the data should be handled with care however, as there was hardly any improvement throughout the task. Participants in the Experimental condition showed a 65.7% categorization performance already in the first block, and they reached an average of 76.8% by Block 4. It might be the case that the early stage of learning is already over within Block 1, and characteristics of later stages also appear already here. In Experiment 2 we used a version of the WP task where stimulus presentation time was limited, which we expected to make explicit availability more difficult. This way we expect learning performance to be prolonged in time.

Experiment 2

The goal of Experiment 2 was to prolong the learning process, and test whether any signs of implicit knowledge appear. We aimed to prolong learning on the WP task by increasing task-difficulty: with the reduction of stimulus presentation time. The structure and procedure of Experiment 2 is identical to that of Experiment 1 with the exception that stimulus exposure is limited. Also, hypotheses were identical to those of Experiment 1, according to the implicit-first hypothesis (Knowlton et al, 1994, 1996), learning should be implicit in the early blocks of the task, according to the strategy-hypothesis (Gluck et al, 2002, Meeter et al, 2006, 2008), learning should be explicit under the use of single strategies, while implicit under the use of the multi-cue strategy, whereas according to the explicit-hypothesis (Lagnado et al, 2006, Newell et al, 2007), learning should be explicit. As it was hypothesized in Experiment 1, we expect that learning in Block 1 will be closer to chance level, than to Block 4 performance (which was not the case in Experiment 1).

Method

Participants. Altogether 52 subjects participated in the study (27 females and 25 males). Their mean age was 20.65 years (SD = 1.37 years). All participants were recruited from the Budapest University of Technology and Economics, and participated for credit points. All participants provided a written informed consent in accordance with the principles of the Helsinki declaration and the stipulations of the Institutional Review Board. Both groups included 26 randomly assigned participants (see below).

Procedure. The procedure was very similar to the WP task explained in Experiment 1, except that participants saw the target cues for a limited time only. Stimulus presentation time was 200 milliseconds per cue: that is 200 milliseconds if only one cue was present, 400 milliseconds in the case of a combination of two cues, and 600 milliseconds if there were three cues present at the same time. While the time of exposure was proportional to the number of cues, cue combinations – as in Experiment 1 – appeared simultaneously.

Just like in Experiment 1, there were two conditions, an Experimental and a Control condition. In the case of the Experimental condition, the task had a predictive structure: Cue1 (square) lead to sunshine in 85,7% of its appearances, Cue2 (triangle) lead to sunshine in 70%, Cue3 (pentagon) lead to sunshine in 30%, while Cue4 (rhombus) was associated with sunshine in 14.30% of its appearances. Cues of the Control condition had 50% predictive value, i.e. and equal chance to be associated with either outcomes, see Table 1. The procedure was identical to the Procedure in Experiment 1 (see description thereof).

Strategy analysis. Strategy use was computed for Block 2, and strategy analysis was identical to the method explained in Experiment 1.

Results

As in Experiment 1, some participants did not have a response in all four cells (Block 1 Implicit, Block 1 Explicit, Block 4 Implicit, and Block 4 Explicit): 24 Experimental and 24 Control condition participants had valid data. All data analyses were conducted on these participants, except when indicated otherwise.

Learning performance. Learning performance was tested on the Experimental group using a repeated-measures ANOVA with Block (1 through 4) as within-subject variable. There was a monotonic increase in performance, confirmed by a significant linear trend, $F(1, 23) = 57.784$, $p < 0.001$, $\eta^2_p = 0.715$. Figure 1 illustrates categorization performance by Experiment and by Condition.

Comparing early and later blocks in terms of self-insight by Condition. Blocks 1 and 4 were compared to test whether either the early or the later blocks of the task may be considered as implicit based on self-insight reports. This analysis included data of both the Experimental and Control conditions.

A 2 x 2 x 2 repeated-measures ANOVA was conducted with Block (Early vs. Late) and Answer-type (Implicit vs. Explicit) as within-subject variables, and Condition (Experimental vs. Control) as a between subject variable. Note, that data for Block and

Condition are invariant (there are 50 answers in each block, and 100 answers in each condition), so the main effects of Block and Condition, and the Block x Condition interaction are not applicable.

The ANOVA revealed a significant Answer-type x Condition interaction, $F(1, 46) = 6.078$, $p < 0.05$, $\eta^2_p = 0.117$, a significant Block x Answer-type interaction, $F(1, 46) = 41.962$, $p < 0.001$, $\eta^2_p = 0.477$, and a significant Block x Answer-type x Condition interaction, $F(1, 46) = 6.505$, $p < 0.05$, $\eta^2_p = 0.124$. Figure 2B illustrates the number of Implicit and Explicit answers by Block and by Condition.

To further analyse the Block x Answer-type x Condition interaction, a separate 2 x 2 repeated measures ANOVA was conducted for each Block with Answer-type (Implicit vs. Explicit) as within-subject variable and Condition (Experimental vs. Control) as between subject variable. For Block1, the ANOVA revealed a significant main-effect of Answer-type, $F(1, 46) = 17.358$, $p < 0.001$, $\eta^2_p = 0.274$. This shows that there were more Implicit than Explicit answers in Block 1. The Answer-type x Condition interaction was not significant, suggesting that in terms of the number of self-insight reports, the Experimental and Control conditions did not differ from each other.

The ANOVA for Block 4 revealed that there were more Explicit answers, as confirmed by a significant main-effect of Answer-type, $F(1, 46) = 6.556$, $p < 0.05$, $\eta^2_p = 0.125$, and there was also a significant Answer-type x Condition interaction, $F(1, 46) = 9.494$, $p < 0.01$, $\eta^2_p = 0.171$. The interaction shows that in the Control condition the number of Implicit and Explicit answers are comparable, as confirmed by a paired samples t-test, $t(23) = 0.319$, $p = 0.752$, while in the case of the Experimental condition, the number of Explicit answers are significantly higher, $t(23) = 4.872$, $p < 0.001$.

Categorization performance associated with Self-insight type by Condition and by Block. A 2 x 2 x 2 repeated measures ANOVA was conducted with Block (Early vs. Late) and Answer-type (Implicit vs. Explicit) as within-subject variables and Condition (Experimental vs. Control) as between subject variable.

The ANOVA revealed that performance on Block 4 was significantly higher than performance on Block 1, $F(1, 46) = 7.447$, $p < 0.01$, $\eta^2_p = 0.139$. Performance with self-insight was higher than without self-insight, as it was revealed by a significant Answer-type main effect, $F(1, 44) = 101.869$, $p < 0.001$, $\eta^2_p = 0.689$. Also, performance for the Experimental condition was significantly higher than that of the Control condition, confirmed by a significant main effect of Condition, $F(1, 46) = 42.909$, $p < 0.001$, $\eta^2_p = 0.482$.

Apart from the main effects there was a significant Answer-type x Condition interaction, $F(1, 46) = 85.758$, $p < 0.001$, $\eta^2_p = 0.651$, and a significant Block x Conditions interaction, $F(1, 46) = 9.244$, $p < 0.01$, $\eta^2_p = 0.167$. Block x Answer-type ($p = 0.146$) and Block x Answer-type x Condition ($p = 0.898$) interactions were not significant. We do not provide further analysis on the Block x Condition interaction, as it only touches upon differences in categorization performance.

To further explore the Answer-type x Condition interaction, data in the two blocks was merged, and both performance associated with implicit answers and performance associated with explicit answers were compared between the two conditions. A MANOVA was employed with implicit and explicit performance as dependent variables and Condition as between-subject variable. Results revealed that performance associated with explicit answers were significantly higher in the Experimental condition, $F(1, 46) = 336.848$, $p < 0.001$, $\eta^2_p = 0.880$, while performance associated with implicit answers did not differ by Condition, $p = 0.399$. Figure 3B illustrates performance associated with Implicit versus Explicit self-insight.

Self-insight by Strategy use in Block 2. We tested whether participants using different strategies report different levels of self-insight. Self-insight data and strategy use data was only analysed in Block 2. Again, only participants with a developed single or multi-cue strategy were considered: 10 participants used one of the single strategies, and 12 participants used multi-cue strategy. The remaining 2 participants did not use any of the strategies defined by Gluck et al. (2002).

A 2 x 2 repeated-measures ANOVA was conducted with Answer-type (Implicit vs. Explicit) as within-subject, and Strategy use (Single vs. Multi-cue) as between-subject variable. The numbers in each strategy are invariant, hence Strategy main effect is not applicable. The ANOVA revealed a significant Answer-type x Strategy use interaction, $F(1, 20) = 5.361$, $p < 0.05$, $\eta^2_p = 0.211$. The main effect of Answer-type was not significant, $p = 0.185$. Figure 4 illustrates the number of responses in each category by Strategy use.

Categorization performance associated with Self-insight type by Strategy use in Block 4. We also tested whether learning performance differs based on strategy use and self-insight type. A 2 x 2 repeated measures ANOVA was conducted with Answer-type (Implicit vs. Explicit) as within-subject variable, and Strategy use (Single vs. Multi-cue strategy) as between subject variable. The dependent variable was WP performance. The WP performance with explicit self-reports were 89.97% (SE = 3.20%) for Single and 95.14% (SE = 1.33%) for Multi-cue strategy users. For items without reported self-insight, Single strategy users achieved 44.20% (SE = 4.24%) accuracy, whereas Multi-cue users scored 50.45% (SE = 5.92%). Performance was higher in the case of explicit self-reports. This was confirmed by a significant main effect of Answer-type, $F(1, 20) = 127.261$, $p < 0.001$, $\eta^2_p = 0.864$. Performance of multi-cue strategy users and single strategy users were comparable ($p = 0.191$). And also the Answer-type x Strategy use interaction was not significant ($p = 0.894$)

Discussion

Results showed that Experiment 2 achieved its aim: categorization performance in Block 1 did get closer to chance level. Also, results showed that limiting stimulus presentation did change self-insight reports in a number of ways. However in general, results seem to point into the same direction as Experiment 1. Based on the number of answers in each self-report category we may not conclude that the Experimental and Control conditions differed in Block 1. However, considering performance associated with the different Answer-types we see that although the numbers are comparable, performance on the Explicit items are higher for the Experimental condition. Performance associated with Implicit self-reports were also comparable between the two Conditions. These results are again in line with the Explicit theory (Lagnado et al., 2006), and do not confirm the Implicit-first theory (Knowlton, Mangels et al., 1996).

Similarly to Experiment 1, multi-cue strategy users gave less Implicit and more Explicit answers than users of the single strategies. In terms of performance, Explicit self-report lead to higher performance. At the same time there were no differences by Condition, and the interaction was not significant either. Combining the analyses suggests that multi-cue strategy users give more Explicit answers that in turn leads to higher performance. This suggests that both strategies lead to explicit decisions and lack implicit decisions, the only difference is the more frequent usage of explicit decisions in Multi-cue strategy. The results do not confirm the predictions of the Strategy hypothesis (Gluck et al., 2002).

General Discussion

Earlier studies of the role of implicit versus explicit processes in the Weather Prediction task focused on structural knowledge: knowledge of the associations between cues and outcomes. As it has been reviewed in the introduction, amnesic participants show learning on the WP task together with structural knowledge above chance (Reber et al., 1996). As patients with severely impaired declarative system can report some kind of structural knowledge, it is not clear how structural knowledge may be interpreted in terms of explicit versus implicit processes. This way, we decided to focus on self-insight instead of structural knowledge. Participants were exposed to the WP task, and were asked to characterize each of their decisions, whether it was based on guessing, intuition, some unidentifiable knowledge ('I think I know the answer'), memories of earlier cue-outcome presentations or the knowledge of the rule.

Experiment 1 showed that participants gave more implicit type answers in the Control condition than in the Experimental condition. Performance was higher for the Experimental condition, but only for items associated with an explicit self-report (Remember or Rule answers). In general, learning was only associated with explicit answers. Strategy use was also examined, and analysis of Experiment 1 found that more explicit answers are provided by Multi-cue users than by Single strategy users. Implicit performance was near chance level, and deviations from chance performance were only manifest in items with an explicit self-report. In sum, the advantage in the overall performance of Multi-cue users is only associated with a higher number of explicit decisions. Results of Experiment 1 were in line with the Explicit theory (Lagnado et al., 2006), contradicting both the Implicit-first (Knowlton, Mangels et al., 1996) and Strategy theories (Gluck et al., 2002). The Implicit-first hypothesis differentiates between the early and later blocks of the task. In Experiment 1 however, categorization performance on Block 1 was already nearer to performance on Block 4 than to chance level. For this reason we tried to prolong the learning trajectory in a new design in Experiment 2.

In Experiment 2, the WP task was made more difficult by a reduction of cue-presentation time. We expected, and found an extended learning trajectory: it took longer for participants to reach the maximum learning performance of Experiment 2. Prolonging the learning process lowered performance on Block 1. In spite of this difference, Experiment 2 confirmed results of Experiment 1: participants did not show any evidence of implicit learning, either time-based, or strategy-based. That is, just like Experiment 1, Experiment 2 did not provide any evidence in favour of either the implicit-first (Knowlton, Mangels et al., 1996; Knowlton et al., 1994) or the strategy-theory (Gluck et al., 2002). Instead, learning seemed to rely on explicit processes in both experiments, reflected in the fact that performance associated with implicit self-reports was at chance level, while performance associated with explicit self-reports was always above chance.

Several concerns should be addressed with the methodology of the current paper. A major methodological issue is the reliability of introspection. There are numerous events when participants are not aware of their performance. This is well explored by implicit learning studies (Frensch, Lin, & Buchner, 1998). As we were interested in Higher Order Thoughts (Rosenthal, 1986; 2005) on performance, it was necessary to use introspective results. However, to minimise possible distortion caused by participants not being aware of their performance, data was compared to a baseline. This way explicit performance was not defined as an absolute criterion in the number of explicit answers, but as a deviation from baseline, which was provided by a control group who completed a task without a predictive structure. This does not solve the problem of introspection, but controls for it across the two conditions: since both the experimental and the control conditions share the bias induced by asking for introspective answers, any difference that we observe between the two conditions is due to the experimental manipulation.

An interesting, closely related phenomenon is the lack of Block x Answer-type x Condition interaction in both experiments, which suggests that there is a change in the distribution of self-reports for participants in the control condition with random associations as well. A potential reason is that participants are trying to find the 'rule' in an experimental setup regardless of whether there is one. In both conditions, they develop a strategy and might end up with using a rule consciously, although in the control condition, this would never lead to an increase in performance. Further research is required to clarify this point too.

A major methodological concern is about the time of self-report collection. For each item the sequence of events was the following: cues appeared, then the participant's response was collected, after which feedback was given. The self-insight question only appeared after the feedback. The reason for this design was to keep the cue and outcome units together, and not to increase the time between cues and outcomes, as it is documented to interfere with implicit categorization performance (Maddox & Ing, 2005). On the other hand, this could modify the self-report behaviour of participants: participants might report guessing if their prediction was incorrect, and might report rule-knowledge or memory use if the prediction was correct. There are results arguing against this concern. Such a strategy would predict that participants' performance without self-insight would be around 0 per cent, whereas performance with self-insight would be at 100%. This is not the case, though: in both experiments, performance without self-insight is around 40%, while performance with self-insight is around 90%. This suggests that although there are more wrong than correct predictions on items where no self-insight is reported, a relatively high percentage of correct answers is also associated with a lack of self-insight. However, more research is required to decide whether this methodological problem applies.

Conclusion

Altogether, our results are not in line with earlier hypotheses, suggesting either that learning is implicit in the early stages of the Weather Predicition task (Knowlton, Mangels et al., 1996; Knowlton et al., 1994), or that implicit and explicit systems are differentially employed by different strategies (Gluck et al., 2002). In general, results from two experiments show that solving the WP task does not rely on implicit processes at all, and are compatible with a single-systems approach to the WP task (Lovibond & Shanks, 2002; Lagnado et al., 2006; Newell et al., 2007; Price, 2009).

Implicit hypotheses of the WP task are mainly based on neuropsychological evidence (Hopkins et al., 2004; Keri et al., 2000; Keri et al., 2002; Knowlton et al., 1994; Knowlton, Squire et al., 1996; Reber et al., 1996; Shohamy et al., 2004), proposing that learning is implicit if participants with a severe declarative impairment are still able to show improvement on the task. As reviewed above, similarly to our results, other papers (Lagnado et al, 2006; Newell et al, 2007; Price; 2009) showed that healthy participants do not provide evidence for the implicitness of learning on the WP task. The discrepancy here may have two explanations. One possibility is that patients with amnesia are in fact impaired on categorization tasks: a meta-analysis by Zaki (2005) found that in most studies comparing patients with amnesia and healthy controls, the control groups show a numerical advantage. The meta-analysis showed that patients with amnesia are impaired on categorization (Zaki, 2005). The other possibility is that patients with amnesia and healthy participants use different strategies: patients with amnesia use their non-declarative system due to impairment in the declarative system (Knowlton & Squire, 1993), whereas healthy participants do not have such constraints. Healthy participants may use their declarative systems since it is available, which might lead to better performance (eg. Reber, 1992; or Willingham, 1998). This may be important in the case of the WP task, as participants are explicitly asked to try to infer the outcome based on the presented stimuli. Further research on both neuropsychological patients and different manipulations on the task in typical populations would be required to test how learning of patients with amnesia and healthy participants really differ.

References

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*(5-6), 338-351.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological Research-Psychologische Forschung, 61*(2), 83-98.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*(6), 408-418.

Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia, 42*(4), 524–535.

Kemény, F., & Lukács, Á. (2009). The effects of feature analysis and transparency in probabilistic category learning in adults and children. *Learning & Perception, 1*(2), 199-213.

Keri, S., Kelemen, O., Szekeres, G., Bagoczky, N., Erdelyi, R., Antal, A., Benedek, G., & Janka, Z. (2000). Schizophrenics know more than they can tell: probabilistic classification learning in schizophrenia. *Psychological Medicine, 30*(1), 149-155.

Keri, S., Szlobodnyik, C., Benedek, G., Janka, Z., & Gadoros, J. (2002). Probabilistic classification learning in Tourette syndrome. *Neuropsychologia, 40*(8), 1356-1362.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*(5280), 1399-1402.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262*(5140), 1747-1749.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*(2), 106-120.

Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., Swenson, M., & Butters, N. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology, 10*(4), 538-548.

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General, 135*(2), 162-183.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28*(1), 3-26.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 650-662.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition, 32*(4), 582-591.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 100-107.

Meeter, M., Myers, C. E., Shohamy, D., Hopkins, R. O., & Gluck, M. A. (2006). Strategies in probabilistic categorization: results from a new way of analyzing performance. *Learning & Memory, 13*(2), 230-239.

Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: how do normal participants and amnesic patients do it? *Neuroscience and Biobehavioral Reviews, 32*(2), 237-248.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review, 14*(3), 505-511.

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

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*(4), 564-574.

Price, A. L. (2009). Distinguishing the contributions of implicit and explicit processes to performance of the weather prediction task. *Memory & Cognition, 37*(2), 210-222.

Reber, A. S. (1992). The Cognitive Unconscious - an Evolutionary Perspective. *Consciousness and Cognition, 1*(2), 93-133.

Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience, 110*(5), 861-871.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectivness of reinforcement and nonreinforcement. In A. H. Black & W. P. Prokasy (Eds.), *Classical Conditioning II: current research and theory* (pp. 64-99). New York: Appleton Century Croft.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49*(3), 329–359.

Rosenthal, D. M. (2005). *Consciousness and mind*. Oxford: Clarendon.

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioral Reviews, 32*(2), 219-236.

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*(4), 676-686.

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

Zaki, S. R. (2005). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin & Review, 11*(6), 1048–1054.

Table 1. Types and occurrences of cues or cue-combinations per blocks of 50 trials in the two conditions. The first column (Cues) shows which cues are present in a given combination: A is cue1, B is cue2, C is cue3, D is cue4. Frequency is the number of appearances within a block of 50 trials. The third and fourth columns provide the probability that the given cue or combination leads to sunshine. The same structure applies for both Experiments 1 and 2.

Cues	Frequency	Experimental condition	Control condition
A	8	0,875	0,5
B	4	0,75	0,5
C	4	0,25	0,5
D	8	0,125	0,5
AB	8	0,875	0,5
AC	1	1	1
BC	2	0,5	0,5
BD	1	0	0
CD	8	0,125	0,5
ABC	2	1	0,5
ABD	1	1	1
ACD	1	0	0
BCD	2	0	0,5

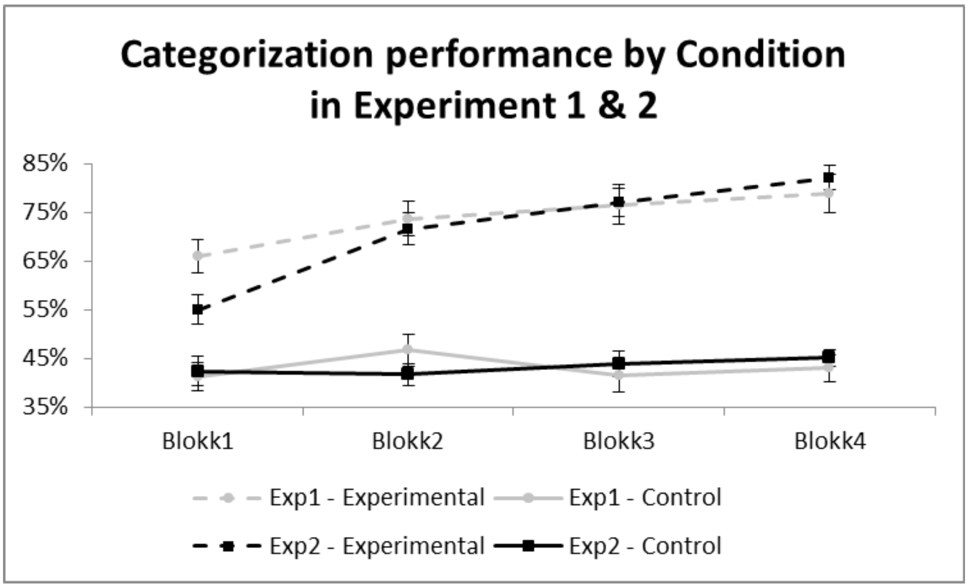
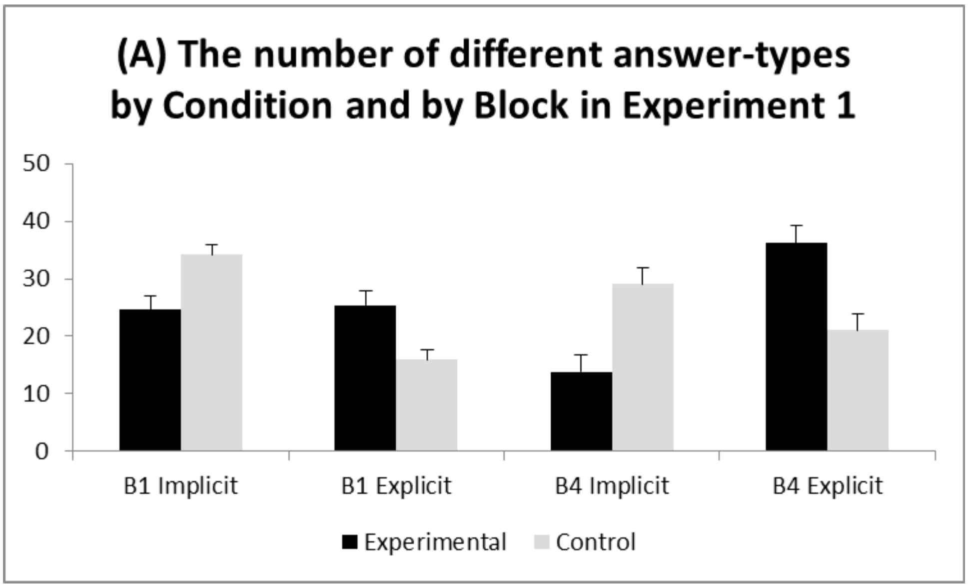


Figure 1. Learning performance by Experiment and by Condition. Gray lines are for data in Experiment 1, while black lines are for data in Experiment 2. Broken lines represent Experimental conditions, while solid lines represent Control conditions. Error bars indicate Standard Errors



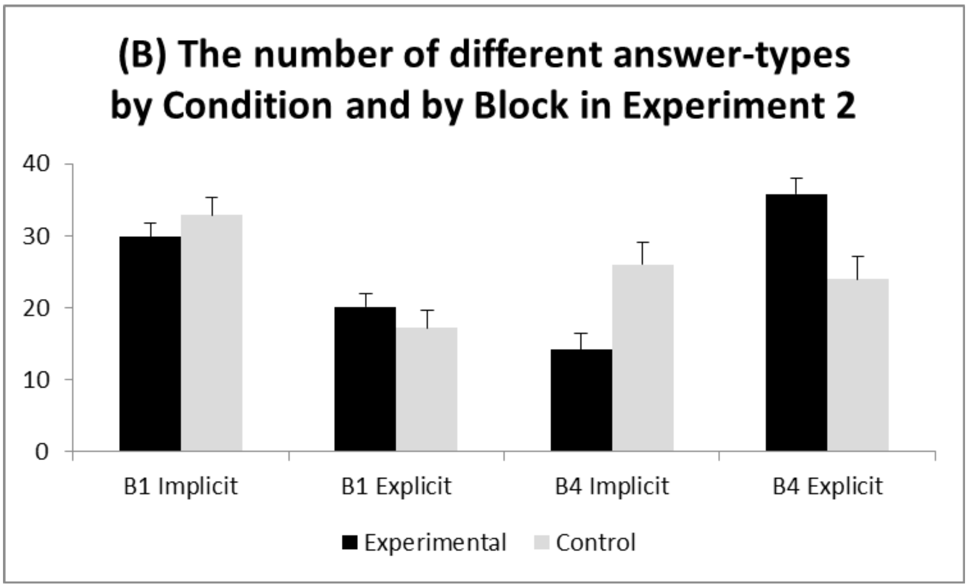
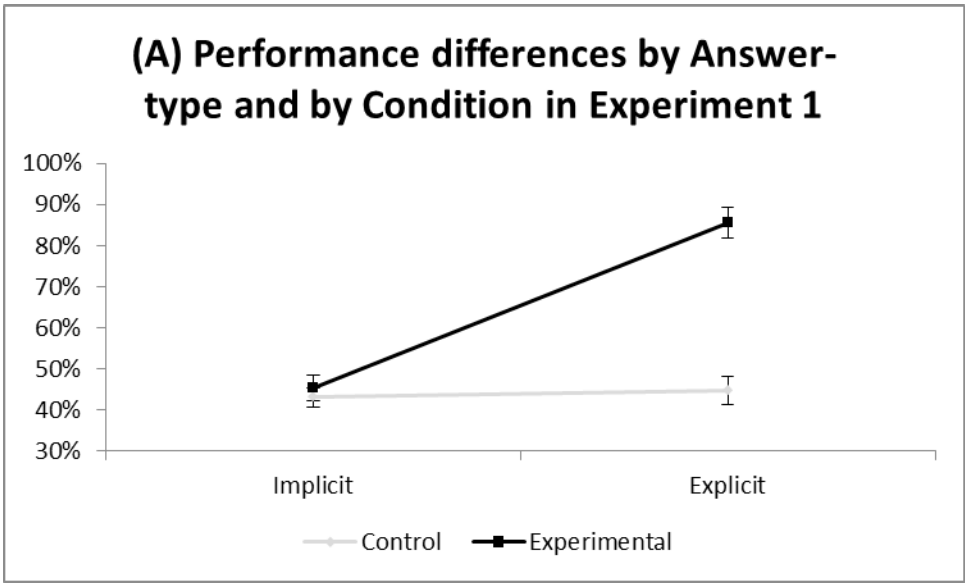


Figure 2. Number of answers in each category by Phase and by Condition in Exp 1 (A) and Exp 2 (B). Error bars indicate Standard Errors


(A) Performance differences by Answer-type and by Condition in Experiment 1
100%
90%
80%
70%
60%
50%
40%
30%
Implicit
Explicit
Control
Experimental

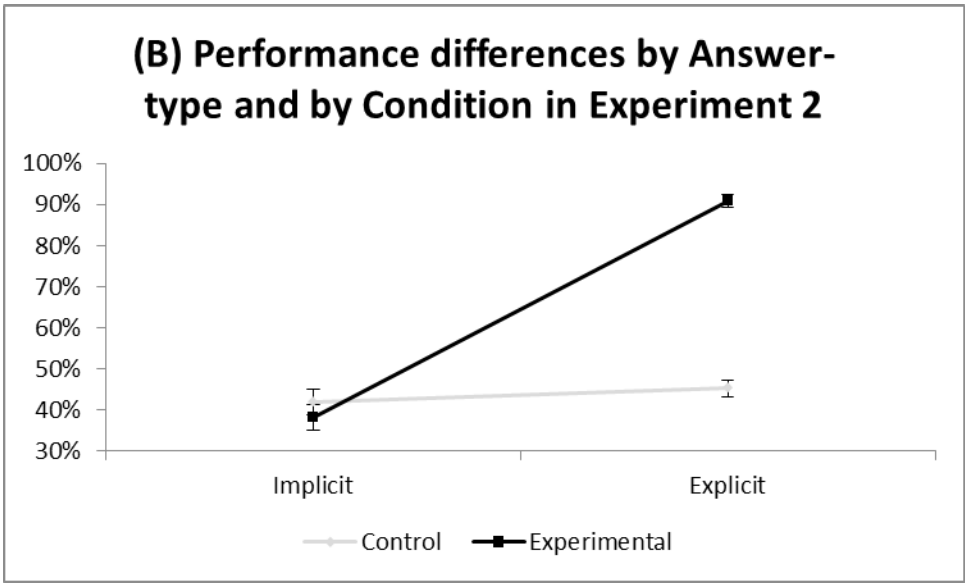

(B) Performance differences by Answer-type and by Condition in Experiment 2
100%
90%
80%
70%
60%
50%
40%
30%
Implicit
Explicit
Control
Experimental

Figure 3. Performance associated with the different Answer-types by Condition in Exp 1 (A) and Exp 2 (B). Error bars indicate Standard Error of Mean.

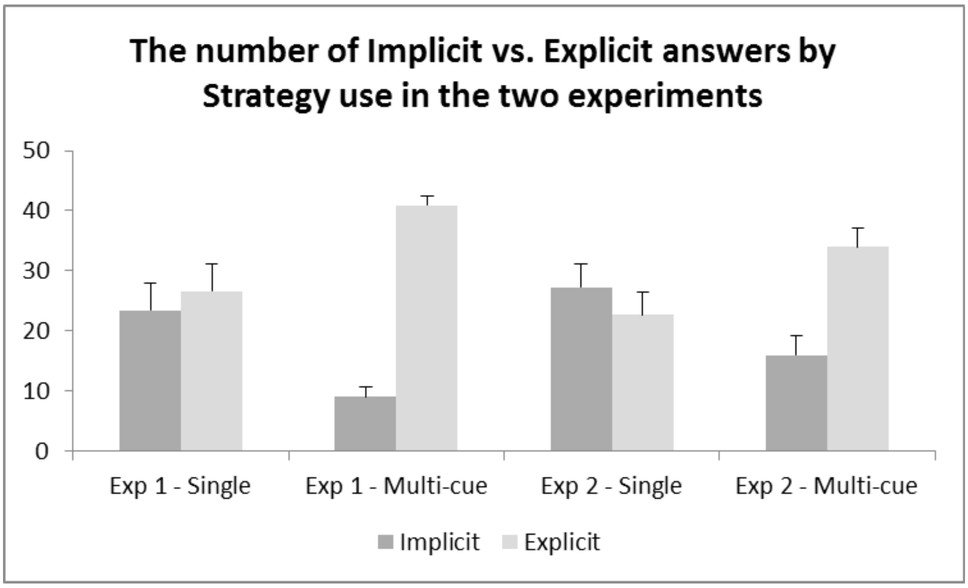


Figure 4. The number of Implicit (darker gray) and Explicit (lighter gray) answers in Block 2 by Strategy use. Both Exp 1 (left) and Exp2 (right) data are presented. Error bars indicate Standard Error of Mean.

10. General discussion

The dissertation presents four studies using the Weather Prediction task or the Serial Reaction-Time task. The goal of these studies is two gain better insight into the process of learning by examining IL from different aspects. Study 1 further extended the earlier findings in the neuropsychological literature on IL, and related procedural learning to language use by showing that probabilistic categorization using the WP task is affected in Specific Language Impairment. Studies 1 and 2 reported data from both typically developing children and adults, allowing for comparisons concerning the development of procedural learning. Study 2 and 3 tapped into the question whether procedural learning is affected by the systematic changes of stimulus sets. Study 4 tested the role of self-insight in probabilistic categorization.

The goal of Study 1 was to test whether the pattern shown by children with SLI is in concert with the Procedural Deficit Hypothesis (Ullman & Pierpont, 2005). As reviewed in the introduction, the PDH suggests that SLI is not a language specific impairment, but is due to more general deficits in the procedural system. As Parkinson's syndrome can also be characterized with a deficit in the procedural system, we hypothesized that children with SLI show a similar performance to PD patients. Our results showed that children with SLI show a near chance level performance, with the inability to develop any strategies. Results are in partial accordance with earlier neuropsychological literature testing Parkinson's disease. Earlier studies of Parkinson's patients showed that performance on the WP task is near chance level (Knowlton, Mangels et al., 1996). Later studies highlighting the integrative role of basal ganglia showed that PD patients are not able to switch to multi-cue strategy (Shohamy et al., 2004). Considering a procedural deficit in SLI the former results are replicated, while the latter are not. Developmental differences observed in Studies 1 and 2 may provide a sufficient explanation to this incongruence.

Both Studies 1 and 2 tested both typically developing children and adults. Data from study 1 showed that while young adults' performance has a numerical advantage over that of TD children, there was no significant difference between the two groups. Considering strategy use, adults were more likely to develop a multi-cue strategy, as 10 out of 16 participants used this response pattern. TD children on the other hand mostly used one of the single strategies with some multi-cue users and two no strategy users. While the mean age of TD children was 11;3 (Sd: 1;3) in Study 1, children participating in Study 2 were younger. In Study 2, year 2 primary-school students were tested. Their mean age was between 8;6 and 8;9 (with Sds: 0;3 - 0;6 in the four group of children). Results revealed a massive age-group effect in both performance and strategy use, with adults showing higher performance and more multi-cue usage, whereas children were more likely to use one of the single strategies. Compared to TD children, children with SLI show a severe impairment on the WP task. This implies that the decrement in WP performance is expected to manifest also in strategy use. That is, we should not expect children with SLI to use one of the single strategies – which is what we see in Parkinson's syndrome (Shohamy et al., 2004) – as their peers who perform much better show this pattern.

Taken together, Study 1 supports claims that IL and language learning are closely associated. Earlier studies have already reported that Specific Language Impairment is not necessarily specific to language, but may involve a number of different non-linguistic deficits too. These deficits include motor control of oral and fine movements, hypothesis testing and categorization, mental rotation, sequencing, word retrieval, phonological categorization, simultaneous execution and executive functions (Leonard, 1997). The Procedural Deficit Hypothesis on the other hand provides a direction to the co-occurring deficits, suggesting that linguistic difficulties are the result of malfunctions in the procedural system (Ullman &

Pierpont, 2005). Our results only confirm the co-occurrence of procedural and linguistic deficits; we however provided no data on the direction of the relationship.

While there are ample studies on age-related changes on the SRT and AGL tasks, the WP literature lacked studies testing children. Both studies 1 and 2 reported comparisons of typically developing children and adults. As discussed above, a prolonged gradual change was observed. Children at 8 years of age showed significantly lower performance than adults, and also used less advanced strategies. At the same time children over the age of 11 did not differ statistically from adults, but in the light of Study 2 data we can conclude that their numerical disadvantage and more frequent single strategy usage is an intermediate step throughout development. Our studies in general are in line with earlier studies suggesting a continuous increase with development (Fletcher et al., 2000). A number of issues are open yet. Thomas et al (2004) found that children between 7 and 11 perform significantly lower than adults due to potential differences in fronto-striatal brain circuitry. In Study 1 we tested only slightly older children. The question arises whether the maturation of the same fronto-striatal circuitry should be critical in the WP task as identified by Thomas et al's fMRI study. The question also rises whether we find the same developmental trajectory as in the SRT task, or whether there is an earlier peak in the developmental curve of the WP task. More systematic data collection is required to answer these questions. Systematic data collection would also answer a number of questions in connection with ageing, as previous studies are not conclusive on the issue. There are results suggesting that sequence learning is impaired in older ages (Howard & Howard, 1997) as well as data in favour of preserved skill learning (Gaillard, Destrebecqz, Michiels, & Cleeremans, 2009).

The main focus of Study 2 was the modification of the inner structure of cues and the link between cues and outcomes. Data confirmed that both Combination and Transparency play an important role in probabilistic categorization. The novelty in this study is that contrary to previous research it focused on the inner structure of stimuli and not on transfer over domain- or modality borders. The advantage of Cue-based over Holistic presentation can be explained by perceptual effects. If each cue is a separate image, it is easier to perceive and utilize them. At the same time, the identification of holistically presented cues is more difficult. This may be a peripheral effect causing a decrease in performance. However, testing whether this effect is peripheral, or is at the heart of procedural learning requires more research.

Transparency between cues and outcomes leads to better learning early in the task, but this advantage turns into a disadvantage later on. The explanation to such phenomenon could be that transparent associations are easy to verbalize, and may be acquired faster, but they may not be subject to gradual change. A possible explanation is rooted in the implicit versus explicit functioning throughout the task. It might be that easily verbalizable strategies are explicit, whereas strategies that are difficult to verbalize are implicit. This is in concert with hypotheses by Gluck and colleagues (2002). They however considered the declarative strategies as the first step towards a later emerging procedural representation. Incorporating task difficulty may change the picture though. It is possible that single strategies are only declarative if the cue-outcome associations are easy to learn. In such a case though, the declarative strategy does not lead to a more advanced procedural representation. It is also important that declarative representations are not subject to gradual, feedback-based changes. This explains why the early advantage turns into a disadvantage later. On the other hand, in the case of difficult to learn cue-outcome associations, participants may develop a procedural strategy in the beginning, and as soon as they have explicitly learned the cue-outcome associations, they are already able to combine cues and outcomes. More research is required to test the effect of conscious awareness on strategy use. As a first step, in Study 4 we adapted subjective self-insight measures of the Artificial Grammar Learning task (Dienes & Scott, 2005) to the WP task.

Easy strategies lead to quickly acquired representations that later become sub-optimal. The same pattern is observed in Study 3, where we tested sequence learning in three different conditions. Response sequences were shown to be learnable in the lack of one-to-one mapping between motor and perceptual information, on the other hand, stimulus sequences were not learnable if they were not in the attended, target domain. These results could have suggested that learning on the SRT task is purely response-based, and perceptual information has no effect on it. Another condition was added to test this possibility. A probabilistic perceptual structure was added to the response sequence in the Extra condition. While participants were responding with sequentially organized responses to categories, the pictures locations varied in a systematic way. The location of the stimuli were random, but each location had a high frequency category (55% of all appearances), and three low frequency categories (15% each). This way the categorical identity of the given picture could be predicted by its location. Again, there is a highly predictive complex structure (response sequence), and a less predictive simple structure (location frequency). Results showed that the location frequency is acquired, but the response sequence is not. That is, similarly to Study 2, the less complex prediction is the one learned. It is also important that, similarly to Study 2, the clash between two predictive structures cause a decrement even in the acquisition of the less complex one: confirmed by the Frequency control condition, the location frequency effect is higher if there is no response structure in the task. In sum, results of both studies suggest that procedural learning is decreased in the case of multiple structures.

Study 2 suggested that there may be differences in the explicit versus implicit nature of probabilistic categorization on the WP task. So far though, no direct measures of explicitness have been developed to the task. In Study 4 we adapted the subjective self-report method of Dienes and Scott (2005). This method was originally designed for the Artificial Grammar Learning task. During the WP task, after each decision, participants were prompted to categorize the basis of their decision. They could report reliance on Guessing, Intuition, 'I think I know the answer'-type knowledge, Memory and Rule knowledge. The first three can be clustered to the implicit whereas the last two to the explicit category. This is an introspective measure, and using introspection is not considered completely reliable. The best example is implicit learning, which is knowledge without conscious access to the knowledge. For this reason we employed a control condition where participants faced the same task including the prompt about the source of knowledge, but the cue-outcome associations were random. This way there could not appear any form of learning in the task, which could serve as a baseline for comparison.

Results showing no sign of implicit processes in Study 4 question the validity of Study 1. These issues may be explained in five different ways. On the one hand, even if healthy participants rely on explicit strategies, learning could still be implicit in clinical groups. This could be the explanation for earlier studies finding that patients with amnesia show improvement on the task (Knowlton et al., 1994). The competing memory systems approach suggests that if learning is explicit than there is no need for the activation of the implicit system (Foerde et al, 2007, 2007). The explicit system is available in healthy adults, so they use it, but patients with an impaired explicit learning system have to rely on implicit learning. The second explanation is that it might be the case that the WP task is only explicit for adults, and not for children. This way learning reported in Study 1 is implicit learning, and the deficit of children is due to malfunctions in their procedural systems. More research is required for this possibility too: especially on the self-insight of children. The third and fourth possibilities are that the task is an explicit learning task even for children. In this case we need an explanation why children with SLI show decreased learning. On the one hand they might be impaired in declarative functioning. Memory deficits have indeed been documented before (Gathercole & Baddeley, 1990). On the other hand, decreased probabilistic categorization may be due to decreased executive functions. The link between executive functions and implicit learning have previously documented in Parkinson's syndrome (Jackson et al., 1995). On the other hand, children with SLI were also shown to be impaired on executive functions (Henry et al., 2012). Evaluating the plausibility of the last two possibilities also require further research. A last explanation concerns the methodology used by Study 4. Item-by-item self-insight prompting might call up on a more explicit approach to the task. This leads to a conclusion that highlights explicit processes in the task, while it is possible that, in the absence of item-by-item self-reports, participants normally use implicit learning. More research, maybe with blockwise self-report-collection is required for better explanations.

The four studies in the dissertation offer new insight in the nature of IL in several important respects, especially for probabilistic category learning, where studies on different learning effects are relatively few. Our results showed that probabilistic categorization using the WP task is impaired in Specific Language Impairment (Study 1), separate cue presentation is generally better than holistic presentation, and a transparent link between cues and outcomes lead to better early and worse later performance (Study 2). Children's data also revealed that probabilistic categorization is subject to a slow improvement during development (Studies 1 and 2). Self-insight data collected from young adults questioned the presence of implicit processes in the WP task (Study 4). Results from the SRT task showed that response sequences are acquired in the absence of a correlating perceptual sequence, while unattended perceptual sequences are not acquired. At the same time, a simple probabilistic perceptual structure has a detrimental effect on response sequence learning (Study 3). Besides these new insights, the results also raise a great number of further questions. A number of these are methodological: can we distinguish between pure-perceptual and unattended learning in the SRT (Study 3); can trial-by-trial prompting measure implicit processes without an explicit contamination (Study 4). Another lot are theoretical: what is the direction of the relationship between procedural learning and language (Study 1); are there differences in the explicitness of strategy use if stimulus sets are varied (Study 2); can learning processes be implicit if participants (including patient with amnesia) are able to report some kind of structural knowledge (Study 4). These questions require further studies in connection with both implicit learning and linguistic abilities.

11. References

Aczél, B., & Gönci, D. (2005). *Unitiy or Fractionality of Implicit Learning: A Methodological Aspect.* Paper presented at the Proceedings of the CogSci2005 - 27th Annual Meeting of the Cognitive Science Society, Stresa, Italy.

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality Independence of Implicitly Learned Grammatical Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 899-912.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1999). Statistical Learning in Linguistic and Nonlinguistic Domains. In B. MacWhinney (Ed.), *Emergence of Language*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition, 32*, 1369-1378.

Bennett, I. J., Howard, J. H., Jr., & Howard, D. V. (2007). Age-related differences in implicit learning of subtle third-order sequential structure. *Journal of Gerontology: Psychological Sciences, 62*(2), P98-103.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: theoretical and empirical issues*. Hove: Lawrence Erlbaum.

Bishop, D. V. M. (1992). The Underlying Nature of Specific Language Impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33*(1), 3-66.

Buchner, A., Steffens, M. C., Erdfelder, E., & Rothkegel, R. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *Quarterly Journal of Experimental Psychology, 50A*(631-663).

Buchner, A., Steffens, M. C., & Rothkegel, R. (1998). On the role of fragmentary knowledge in a sequence learning task. *Quarterly Journal of Experimental Psychology, 51A*, 251-281.

Channon, S., Shanks, D., Johnstone, T., Vakili, K., Chin, J., & Sinclair, E. (2002). Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia, 40*(12), 2185-2197.

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

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24-39.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities - Pitting abstract against stimulus-specific representations. *Psychological Science, 17*(10), 905-912.

Curran, T. (1997). Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience, 9*(4), 522-533.

Daselaar, S. M., Rombouts, S. A., Veltman, D. J., Raaijmakers, J. G., & Jonker, C. (2003). Similar network activated by young and old adults during the acquisition of a motor sequence. *Neurobiology of Aging, 24*(7), 1013-1019.

Dennis, N. A., & Cabeza, R. (2011). Age-related dedifferentiation of learning systems: an fMRI study of implicit and explicit learning. *Neurobiology of Aging, 32*(12).

Deroost, N., & Soetens, E. (2006a). Perceptual or motor learning in SRT tasks with complex sequence structures. *Psychological Research, 70*, 88–102.

Deroost, N., & Soetens, E. (2006b). Spatial Processing and Perceptual Sequence Learning in SRT Tasks. *Experimental Psychology, 53*(1), 16-30.

Deroost, N., Zeeuws, I., & Soetens, E. (2006). Effector-dependent and response location learning of probabilistic sequences in serial reaction time tasks. *Experimental Brain Research, 171*(4), 469-480.

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

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*(5-6), 338-351.

Ferraro, F. R., Balota, D. A., & Connor, L. T. (1993). Implicit Memory and the Formation of New Associations in Nondemented Parkinsons-Disease Individuals and Individuals with Senile Dementia of the Alzheimer Type - a Serial Reaction-Time (Srt) Investigation. *Brain and Cognition, 21*(2), 163-180.

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

Fletcher, J., Maybery, M. T., & Bennett, S. (2000). Implicit learning differences: a question of developmental level? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 246-252.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences of the United States of America, 103*(31), 11778-11783.

Foerde, K., Poldrack, R. A., & Knowlton, B. J. (2007). Secondary-task effects on classification learning. *Memory & Cognition, 35*(5), 864-874.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science, 12*(1), 13-18.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America, 103*(7), 2458-2463.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 99*(1), 529-534.

Fu, Q., Fu, X., & Dienes, Z. (2008). Implicit sequence learning and conscious awareness. *Consciousness & Cognition, 17*(1), 185-202.

Gaillard, V., Destrebecqz, A., Michiels, S., & Cleeremans, A. (2009). Effects of age and practice in sequence learning: A graded account of ageing, learning, and control. *European Journal of Cognitive Psychology, 21*(2-3), 255-282.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection. *Journal of Memory and Language, 29*, 336-360.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 225–244.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*(6), 408-418.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*(5), 178-186.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 501-518.

Hartman, M., Knopman, D. S., & Nissen, M. J. (1989). Implicit learning of new verbal associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1070-1082.

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nat Rev Neurosci, 11*(7), 523-532.

Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry, 53*(1), 37-45.

Higham, P. A., Vokey, J. R., & Pritchard, J. L. (2000). Beyond dissociation logic: evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General, 129*(4), 457-470.

Hopkins, R. O., Myers, C. E., Shohamy, D., Grossman, S., & Gluck, M. (2004). Impaired probabilistic category learning in hypoxic subjects with hippocampal damage. *Neuropsychologia, 42*(4), 524–535.

Howard, D. V., & Howard, J. H., Jr. (1989). Age differences in learning serial patterns: direct versus indirect measures. *Psychology and Aging, 4*(3), 357-364.

Howard, J. H., Jr., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging, 12*(4), 634-656.

Hsu, H. J., & Bishop, D. V. M. (2010). Grammatical Difficulties in Children with Specific Language Impairment: Is Learning Deficient? *Human Development, 53*(5), 264-277.

Jackson, G. M., & Jackson, S. R. (1995). Do Measures of Explicit Learning Actually Measure What is Being Learnt in the Serial Reaction Time Task? A Critique of Current Methods, *Psyche* (Vol. 2).

Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard, C. (1995). Serial Reaction-Time Learning and Parkinsons-Disease - Evidence for a Procedural Learning Deficit. *Neuropsychologia, 33*(5), 577-593.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513-541.

Janacsek, K., Fiser, J., & Németh, D. (2012). The best time to acquire new skills: age related changes in implicit learning across human life span. *Developmental Science, 15*(4), 496–505.

Jimenez, L., & Vazquez, G. A. (2005). Sequence learning under dual-task conditions: alternatives to a resource-based account. *Psychological Research, 69*(5-6), 352-368.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Kemény, F., & Lukács, Á. (2009). The effects of feature analysis and transparency in probabilistic category learning in adults and children. *Learning & Perception, 1*(2), 199-213.

Kemény, F., & Lukács, Á. (2010). Impaired procedural learning in language impairment: Results from probabilistic categorization. *Journal of Clinical and Experimental Neuropsychology, 32*(3), 249-258.

Kemény, F., & Lukács, Á. (2011). Perceptual Effect on Motor Learning in the Serial Reaction Time Task. *Journal of General Psychology, 138*, 110-126.

Kemény, F., & Lukács, Á. (2012). Self-insight in Probabilistic Category Learning, *manuscri*: BME Department of Cognitive Science.

Keri, S., Szlobodnyik, C., Benedek, G., Janka, Z., & Gadoros, J. (2002). Probabilistic classification learning in Tourette syndrome. *Neuropsychologia, 40*(8), 1356-1362.

Knopman, D. S., & Nissen, M. J. (1987). Implicit learning in patients with probable Alzheimer's disease. *Neurology, 37*(5), 784-788.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*(5280), 1399-1402.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact Artificial Grammar Learning in Amnesia - Dissociation of Classification Learning and Explicit Memory for Specific Instances. *Psychological Science, 3*(3), 172-179.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262*(5140), 1747-1749.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 169-181.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*(2), 106-120.

Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., Swenson, M., & Butters, N. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology, 10*(4), 538-548.

Koehler, J., Chia, A., & Lindsey, S. (1995). The random match probability in DNA evidence: Irrelevant and prejudicial? *Jurimetrics Journal, 35*, 201-209.

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple-cue learning. *Journal of Experimental Psychology: General, 135*(2), 162-183.

Leonard, L. B. (1997). *Children with specific language impairment*. Cambridge, Mass.: The MIT Press.

Lieberman, M. D. (2000). Intuition: a social cognitive neuroscience approach. *Psychological Bulletin, 126*(1), 109–137.

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience, 16*(3), 427-438.

Lukács, Á., & Kemény, F. (2012). Development of different forms of implicit learning throughout the lifespan: BME Department of Cognitive Science.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 650-662.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition, 32*(4), 582-591.

Maddox, W. T., & Filoteo, J. V. (2001). Striatal contributions to category learning: quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society, 7*(6), 710-727.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 100-107.

Marcus, G., Johnson, S., Fernandes, K., & Slemmer, J. (2004). Rules, statistics and domain-specificity: Evidence from prelinguistic infants, *29th Annual Meeting of the Boston University Conference on Language Development*. Boston, MA.

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

Mayr, U. (1996). Spatial attention and implicit sequence learning: evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 350-364.

Meulemans, T., Peigneux, P., & Van der Linden, M. (1998). T-12. Preserved Artificial Grammar Learning in Parkinson's Disease. *Brain and Cognition, 37*(1), 109-112.

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

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

Neal, A., & Hesketh, B. (1997). Episodic knowledge and implicit learning. *Psychonomic Bulletin & Review, 4*(1), 24-37.

Nemeth, D., Hallgato, E., Janacsek, K., Sandor, T., & Londe, Z. (2009). Perceptual and motor factors of implicit skill learning. *Neuroreport, 20*(18), 1654-1658.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review, 14*(3), 505-511.

Nissen, M. J., & Bullemer, P. (1987). Attentional Requirements of Learning - Evidence from Performance-Measures. *Cognitive Psychology, 19*(1), 1-32.

Nissen, M. J., Willingham, D., & Hartman, M. (1989). Explicit and Implicit Remembering - When Is Learning Preserved in Amnesia. *Neuropsychologia, 27*(3), 341-352.

Owen, A. M., Roberts, A. C., Hodges, J. R., Summers, B. A., Polkey, C. E., & Birbaumer, N. (1993). Contrasting mechanisms of impaired attentional-shift in patients with frontal lobe damage or Parkinson's disease. *Brain, 116*, 1159–1175.

Pascual-Leone, A., Grafman, J., Clark, K., Stewart, M., Massaquoi, S., Lou, J. S., & Hallett, M. (1993). Procedural learning in Parkinson's disease and cerebellar degeneration. *Annals of Neurology, 34*(4), 594-602.

Peigneux, P., Meulemans, T., Van der Linden, M., Salmon, E., & Petit, H. (1999). Exploration of implicit artificial grammar learning in Parkinson's disease. *Acta Neurologica Belgica, 99*(2), 107-117.

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

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119*, 264–275.

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

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews, 32*(2), 197-205.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*(4), 564-574.

Price, A. L. (2009). Distinguishing the contributions of implicit and explicit processes to performance of the weather prediction task. *Memory & Cognition, 37*(2), 210-222.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*(4), 463-475.

Rauch, S. L., Savage, C. R., Brown, H. D., Curran, T., Alpert, N. M., Kendrick, A., Fischman, A. J., & Kosslyn, S. M. (1995). A PET investigation of implicit and explicit sequence learning. *Human Brain Mapping, 3*(4), 271-286.

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

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An essay in the cognitive unconscious*. Oxford: Oxford University Press.

Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience, 110*(5), 861-871.

Reber, P. J., Martinez, L. A., & Weintraub, S. (2003). Artificial grammar learning in Alzheimer's disease. *Cognitive, Affective, & Behaviorial Neuroscience, 3*(2), 145-153.

Reber, P. J., & Squire, L. R. (1994). Parallel Brain Systems for Learning with and without Awareness. *Learning & Memory, 1*(4), 217-229.

Remillard, G. (2003). Pure perceptual-based sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 581-597.

Riedel, B., & Burton, A. M. (2006). Auditory sequence learning: differential sensitivity to task relevant and task irrelevant sequences. *Psychological Research, 70*(5), 337-344.

Roediger, H. L., & Mcdermott, K. B. (1995). Creating False Memories - Remembering Words Not Presented in Lists. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*(4), 803-814.

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language, 47*, 172–196.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105*(3), 669-680.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron, 37*(6), 1013-1025.

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

Seger, C. A., Prabhakaran, V., Poldrack, R. A., & Gabrieli, J. D. E. (2000). Neural activity differs between explicit and implicit learning of artificial grammar strings: An fMRI study. *Psychobiology, 28*(3), 283-292.

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

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*(4), 676-686.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A. & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain, 127*(4), 851-859.

Shohamy, D., Onlaor, S., Myers, C., & Gluck, M. (2001, March). *Category learning in Parkinson's disease: The effect of stimulus presentation on performance.* Paper presented at the 8th Annual Meeting of the Cognitive Neuroscience Society, New York.

Siegert, R. J., Taylor, K. D., Weatherall, M., & Abernethy, D. A. (2006). Is Implicit Sequence Learning Impaired in Parkinson's Disease? A Meta-Analysis. *Neuropsychology, 20*(4), 490-495.

Skosnik, P. D., Mirza, F., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2002). Neural correlates of artificial grammar learning. *Neuroimage, 17*(3), 1306-1314.

Smith, J., Siegert, R. J., & McDowall, J. (2001). Preserved implicit learning on both the serial reaction time task and artificial grammar in patients with Parkinson's disease. *Brain and Cognition, 45*(3), 378-391.

Smith, J. G., & McDowall, J. (2006). When artificial grammar acquisition in Parkinson's disease is impaired: The case of learning via trial-by-trial feedback. *Brain Research, 1067*(1), 216-228.

Squire, L. R. (1986). Mechanisms of Memory. *Science, 232*(4758), 1612-1619.

Squire, L. R., Knowlton, B., & Musen, G. (1993). The Structure and Organization of Memory. *Annual Review of Psychology, 44*, 453-495.

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y. H., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: An fMRI study of children and adults. *Journal of Cognitive Neuroscience, 16*(8), 1339-1351.

Thomas, K. M., & Nelson, C. A. (2001). Serial reaction time learning in preschool- and school-age children. *Journal of Experimental Child Psychology, 79*(4), 364-387.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition, 92*(1-2), 231-270.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*(2), 266-276.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex, 41*, 399–433.

van den Bos, E., & Poletiek, F. H. (2006, July 26-29). *Implicit Artificial Grammar Learning in Adults and Children.* Paper presented at the The 28th Annual Conference of the Cognitive Science Society, Vancouver, British Columbia, Canada.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8*(1), 168-176.

Watanabe, D., Savion-Lemieux, T., & Penhune, V. B. (2007). The effect of early musical training on adult motor performance: evidence for a sensitive period in motor learning. *Experimental Brain Research, 176*(2), 332-340.

Westwater, H., McDowall, J., Siegert, R., Mossman, S., & Abernethy, D. (1998). Implicit learning in Parkinson's disease: evidence from a verbal version of the serial reaction time task. *Journal of Clinical and Experimental Neuropsychology, 20*(3), 413-418.

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

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition, 27*(3), 561-572.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the Development of Procedural Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1047-1060.

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

Witt, A., & Vinter, A. (2012). Artificial grammar learning in children: abstraction of rules or sensitivity to perceptual features? *Psychological Research, 76*(1), 97-110.

Witt, K., Nühsman, A., & Deuschl, G. (2002). Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. *Neuropsychologia, 40*(9), 1534-1540.

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35*(6), 1380-1398.

Zhuang, P., Dang, N., Warzeri, A., Gerloff, C., Cohen, L. G., & Hallett, M. (1998). Implicit and explicit learning in an auditory serial reaction time task. *Acta Neurologica Scandinavica, 97*, 131-137.