Implicit Learning: Development, Individual Differences, and Educational Implications

Implicit Learning:

Development, individual differences, and educational implications

Priya Kalra

Committee:
Jon R. Star, Chair
Andrew D. Ho
John D.E.Gabrieli
Jennifer M.Thomson

A Thesis Presented to the Faculty
of the Graduate School of Education of Harvard University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education
2015

Acknowledgements

I would like to thank my committee for their support, patience, attention to detail and adherence to high standards. I thank my fiancée, Heidi Beeler, for her moral, emotional, and occasionally financial support through this process. This project would not have been possible without the help of all of the following: Amy Finn, Stacy Peazant, Shu-Ling Chen, Larsen Family Foundation, Harvard Student Health Insurance Program, Drs. Sara Kimmel and Katherine LaPierre, and the Federal Student Loan Program. I especially thank my Berkeley SLC students for reminding me why this research is interesting to me.

Table of Contents

Abstract

This dissertation attempts to link models from cognitive neuroscience with problems and models from education research as well as to advance our understanding of implicit learning. In addition to a review of the current understanding of implicit learning from psychology and neuroscience, an essay on the potential applications of implicit learning to education and two empirical studies comprise this document.

The first study compares implicit learning in adults and children to address the question of developmental invariance in implicit learning. One novel aspect of this study is the use of a battery of implicit learning tasks, as well as comparison explicit learning tasks. Although gross differences were not found between adults and children in the implicit learning tasks, nevertheless first-level item analysis revealed that children and adults may differentially exploit stimulus frequency information to perform the tasks.

The second study uses parallel forms of multiple implicit learning tasks to determine the reliability of implicit learning tasks for adult participants. Contrary to the prevailing view of implicit learning, stable individual differences were found. Correlations between individual implicit learning and certain non-cognitive traits (such as conscientiousness) were found, but IQ was not correlated with implicit learning.

Finally, the implications of these findings for basic research as well as for the possibility of applying implicit learning to K-12 instruction are discussed.

Chapter One: Overview

This dissertation was inspired by my curiosity about possible interactions between implicit and explicit memory systems and the representations they create, and how these interactions could be leveraged to optimize instruction in K-12 education. Before I could attempt to investigate inter-system relationships, I found that some fundamental questions about implicit learning needed to be addressed. In particular, Arthur Reber's (1993) proposed characteristics of implicit learning have been both influential but untested—with the outstanding exception of evidence for robustness to disease and injury:

1. Robustness to injury or disease
2. Age independence
3. IQ independence
4. Low individual variability
5. Conservation across phylogeny

In this dissertation I empirically address the proposed characteristics of age-independence, IQ-independence, and low individual variability. In Chapter 2, I review what is known about implicit learning and how it is defined in the context of the multiple memory systems model of human memory. This chapter may be particularly helpful to readers who are unfamiliar with the multiple memory systems model specifically, or unfamiliar with the cognitive neuroscience more generally.

In Chapter 3, I discuss how what we already know about implicit learning could be applied to instructional strategies, as well as how the implicit/explicit distinction in psychology and neuroscience compare to the procedural/conceptual distinction in education research.

Chapter 4 presents an empirical study comparing implicit learning in 10-year-old children and adults. This study directly addresses the 2nd of Reber's proposed characteristics. In contrast to other studies that have compared implicit learning in adults and children, this study uses multiple tasks and furthermore presents first-level item analysis for two of the tasks.

Chapter 5 is another study, this time to address the questions of individual variability and intelligence (i.e. intelligence quotient, IQ) independence. Through the lens of classical test theory, this question can be seen as a question of reliability—can individual differences in implicit learning be reliably measured? Again, multiple tasks are used. The participants in this study are neurotypicals (healthy) young adults.

Finally, Chapter 6 provides a synthesis of the findings and an update on what the state of our understanding of implicit learning and its potential applications to education might be in the light of these findings.

Together, these theoretical and empirical matters are intended to form the keystone of a structure that will reach from research in cognitive neuroscience, cognitive psychology, and educational psychology on one side to instructional strategies and curriculum design on the other. Much work remains to be done for such a structure to become viable, but this dissertation represents a first step. I hope to be able to contribute further to this bridge with my future work in this area.

Chapter Two: Introduction to Implicit Learning

Key theoretical and empirical issues related to defining and operationalizing implicit learning are discussed.

The history of psychology in the 20th century is often described in three phases: the era of introspectionism and psychoanalysis, followed by the total domination of Behaviorism, followed by the “cognitive revolution” (Baars, 1986). The acceptability of the unconscious as a subject of study underwent drastic changes with each of these movements. In their epistemological severity, the Behaviorists rejected many proposed mental structures and contents, including the Freudian conception of the unconscious. The subsequent cognitive paradigm allowed for the possibility of unconscious processes, but this notion of a *cognitive* unconscious differed radically from the Freudian unconscious. Rather than a sea of neurosis-inducing urges, the cognitive unconscious is imagined simply as a set of mental processes that operate below the threshold of awareness.

Cognitive psychology and educational psychology have several intersections and areas of overlap, but implicit or unconscious processes have rarely featured in educational psychology research[1]. Given the modern emphasis on students as engaged learners actively constructing understandings, it may seem strange to suggest that an unconscious form of learning could be useful for education purposes. A key goal of this dissertation is to explore the potential utility of implicit learning to educational theory and practice. However, before that goal can be approached, the fundamentals of research on implicit learning carefully explained.

---

[1] Exceptions will be discussed in Chapter 3.

What is Implicit Learning?

Once the cognitive unconscious was deemed suitable for serious study, a flood of dual system theories followed, positing "implicit" and "explicit" or "fast and "slow[2]" systems across a wide variety of cognitive domains including reading, reasoning, decision-making, and long-term memory (Evans, 2008). Of these dual- or multiple system theories, the one best supported by converging lines of evidence (from behavioral studies, neuropsychology, animal research, and human neuroimaging) is the multiple memory system model, which includes an implicit/explicit distinction. Although the multiple memory system taxonomy does not overtly include "implicit learning" as a proposed construct, nevertheless, the concept of "implicit learning" is not incompatible with the multiple memory systems model.

Using widely accepted definitions and criteria, we can locate implicit learning within the multiple memory systems taxonomy.

Definitions. Psychologist Arthur Reber may have been the originator of the term "implicit learning" in his work on artificial grammar learning (Reber, 1967, 1993). According to Reber (1993), "Implicit learning is the acquisition of knowledge that takes place largely independently of conscious attempts to learn and largely in the absence of explicit knowledge about what was acquired" (p. 5). Cognitive neuroscientist Carol Seger (1994) further delineated three criteria for implicit learning:

> "The first criterion is that the knowledge gained in implicit learning is not fully accessible to consciousness, in that subjects cannot provide a full (or, in many cases, any) verbal account of what they have learned. […] The second criterion

[2] Nobel-prize-winner Daniel Kahneman's recent book, Thinking Fast and Slow has recently brought these ideas into the public spotlight, but dual-system models in psychology date back at least to the 1960s and 70s (Reber, 1967; Schnieder & Schiffrin, 1977; Evans, 2008).

> is that subjects learn information that is more complex than a single simple association or frequency count. […] . The third criterion is that implicit learning does not involve processes of conscious hypothesis testing but is an incidental consequence of the type and amount of cognitive processing performed on the stimuli" (p. 164).

This definition and these criteria will allow us to locate implicit learning within the multiple memory systems model of long-term memory.

Implicit vs. explicit: The multiple memory systems taxonomy. The distinction between implicit and explicit memory emerged from the observation that patients with anterograde amnesia were able to demonstrate learning on certain types of tasks. Anterograde amnesia is the inability to form new memories for facts and events, though retrieval of most existing memories is usually intact. In almost cases, anterograde amnesia is caused by damage to a brain structure called the hippocampus and other structures in the medial temporal lobe (MTL) of the brain. Indeed, bilateral surgical removal[3] of the hippocampus and related MTL structures resulted in one of the clearest and most profound cases of anterograde amnesia in the patient formerly known as H.M. H.M. and patients like him cannot remember events that occurred minutes before, or new words they learned minutes before. They must be constantly re-introduced to the doctors and nurses working with them.

Nevertheless, it was found that H.M. and other amnesic patients were able to demonstrate certain forms of learning and memory at levels similar to normal controls. Despite the fact that they do not remember previously learning or performing the task,

[3] Since his recent death, the identity of H.M. has been revealed to be Henry Molaison. Molaison suffered from crippling seizures and the removal of the hippocampus was an attempt to stem the occurrence of seizures. (Corkin 1968, 1997, 2002, 2013).

amnesics show improvement over time (as indicated by performance measures such as reaction time and accuracy) in a variety of complex tasks such as the Tower of London puzzle (Shallice, 1982), mirror reading (Cohen & Squire, 1980), and even the video game Tetris (Stickgold, Malia, Maguire, Roddenberry, & O'Connor, 2000).

The learning that amnesics demonstrate is not dependent on the medial temporal lobe, and is largely not accessible to consciousness. These non-MTL-dependent types of learning are also present in neurologically typical individuals ("normal controls" or "neurotypicals"), but had not been observed in isolation since normal controls still have the ability to form conscious memories. However, converging evidence from behavioral and neuroimaging studies have now established that indeed normal controls have both an MTL-dependent system for forming conscious memories as well as non-MTL-dependent systems for learning that is generally available to conscious recall (Destrebecqz & Cleeremans, 2001; Schendan, Searl, Melrose, & Stern, 2003; Stark, Reber, & Squire, 1998; Willingham, Salidis, & Gabrieli, 2002). This distinction forms the basis for the modern theory of multiple memory systems, which is now widely accepted (Gabrieli, 1998; Reber, Beeman, & Paller, 2013; Squire, 2004). Figure 1 displays a schematic of the multiple memory systems model (Squire & Zola-Morgan, 1988, 1991; Squire, 2004). MTL-dependent memory is referred to as either "explicit" (Cohen & Squire, 1980) or "declarative" (Graf & Schacter, 1985) memory, while non-MTL forms of memory are correspondingly labeled either "implicit" or "non-declarative." Within each of these major branches (explicit/declarative and implicit/non-declarative), there are further distinctions. For example, declarative memory can be further divided into memory for events (episodic memory) and memory for facts (semantic memory) (Tulving, 1983; Shimamura & Squire, 1987; Tulving, 1989; Tulving & Thompson, 1973; Jacoby 1991).

Based on Reber's definition, implicit learning must be located in the implicit/non-declarative side of the diagram, since he specifies "independently of conscious attempts to learn and largely in the absence of explicit knowledge about what was acquired." (Reber, 1993, p. 5) However, because of the inherent difficulty in operationalizing "conscious" vs. "unconscious," some researchers have advocated the preservation of an ability in amnesics as a necessary criterion (as opposed to phenomenological evidence alone) (Seger, 1994). According to this perspective, lack of awareness is a necessary but not sufficient criterion for implicit learning.

The category "non-declarative memory" can be further subdivided based on the behavior and neuroanatomy. In the next section, I discuss where implicit learning fits within the subdivisions of non-declarative memory.

Implicit learning differentiated from other types of non-declarative memory: The importance of abstraction. The term "implicit learning," as used by cognitive psychologists and neuroscientists, does not include all forms of non-declarative memory formation. Rather, it is one particular type of non-declarative memory phenomenon. As seen in Figure 1, other types of non-declarative memory include classical conditioning, priming, associative learning (emotional and motor), and non-associative learning (Gabrieli, 1998; Squire & Zola-Morgan, 1991). "Implicit learning" does not appear overtly in this taxonomy, but the first part of Reber's definition is instructive here: "Implicit learning is the acquisition of *knowledge* that takes place largely independently of conscious attempts to learn...[emphasis added]" as well as Seger's second criterion that implicit learning must be learning of "information that is more complex than a single simple association or frequency count" (Seger 1994, p. 164). Neither associative learning (which includes classical and operant conditioning) nor non-associative learning satisfy this criterion.

Non-associative learning refers to habituation and sensitization effects, i.e. when an organism becomes more or less responsive to a given stimulus. No new association is formed between stimulus and response, but rather the "weight" or tuning of the stimulus sensation-perception relationship is altered (Kandel, Schwartz, Jessel, 2000; Byrne, 2009). Many forms of non-associative learning take place at the level of peripheral neurons and do not require central nervous system involvement at all.

The term "associative learning" is used in psychology and neuroscience to refer to classical conditioning and operant conditioning. In classical conditioning, a neutral stimulus (the conditioned stimulus, CS) is associated with a stimulus (the unconditioned stimulus, US) that elicits an existing (involuntary) response in the organism's behavioral repertoire until the CS alone can elicit the response; this can be seen as an association between a stimulus and a response (CS and response) or between a stimulus and a stimulus (US and CS). In both humans and other mammals, specific cerebellar lesions can impair the learning of these associations (Gabrieli, 1998). In operant conditioning, an organism's behavior is shaped through reward and punishment; the organism learns to associate some behavior with some consequence and as a result exhibits the behavior more or less frequently. Central nervous system involvement is necessary for operant conditioning (Byrne, 2009).

Implicit learning is defined by Seger and Reber to exclude simple associative learning. In addition to the definition in Reber (1993), a statement in an earlier paper stresses this point: "Implicit knowledge results from the induction of an abstract representation of the structure that the stimulus environment displays, and this knowledge is acquired in the absence of conscious, reflective strategies to learn" (Reber 1989, p. 219). Seger paraphrases this as “Implicit learning is an unconscious learning process that yields abstract knowledge” (Seger 1994, p.162).

In this context, “abstract representation” refers to a general law or principle on the basis of observation of particular instances. As Goodman (1955/1983) pointed out, without the “blessing of abstraction,” we would have to individually represent every object and event we encounter, but with abstraction we can represent categories of objects and events and make inferences (and select responses) based on category membership. Abstract representations could consist of category knowledge, general rules, or more complex relational structures (Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

The significance of this emphasis on abstract representation has its roots in the “cognitive revolution” of the mid-20th century (Reber, 1993). Before that time, the reigning paradigm was Behaviorism and its axiom that all behavior could be explained by simple associative learning. A corollary was the non-existence of internal representations: if all behavior could be explained by S-S or S-R associations, then abstract mental representations of the world would be unnecessary and at best epiphenomenal.

What the cognitive revolution brought was examples of behavior that could not be explained by simple associative learning alone (Thagard, 1996; Baars, 1986). Famously, Chomsky used the abstract structure of language (syntax) as an example—simple associative learning cannot account for the learning of syntax (Chomsky, 1980). Memory for particular instances is not sufficient; induction of some abstract representation of rules and categories is required.

Similarly, implicit learning is concerned with learning that is unconscious, but nevertheless cannot be explained simply by associative learning. Some abstract representation of the structure of the stimuli is necessary (Seger, 1994; Reber, 1993). While Behaviorist models of instruction exist, I am vehemently not advocating them in my attempt to link implicit learning and education. Rather, implicit learning could potentially relate to instruction because transfer across contexts is a goal in education and requires induction of general principles (rather than memorization of specific instances). While transfer is desirable in education, it is also difficult to achieve (Bransford, 1999). If implicit learning—a means of acquiring abstract representations—can be exploited in instructional strategy and design, the problem of transfer may become more tractable. Chapter 3 will focus on the potential relationships between implicit learning and education.

While the learning of abstract rules or categories cannot be explained by simple association learning, nevertheless reasonable mechanisms have been proposed. The abstraction of general rules or categories from (often sparse) examples is referred to as induction and is well known in philosophy and artificial intelligence studies as a difficult problem (Goodman, 1955/1983; Hume, 1748/1993; Quine, 1960). Formal models of induction range from production system models (Holland, Holyoak, Nisbett & Thagard, 1986) to modern hierarchical Bayesian models (HBMs) (Tenenbaum et al., 2011).

Hierarchical Bayesian Models are a type of statistical learning model. Although a wide range of statistical learning models are used in artificial intelligence and machine learning studies, in cognitive psychology—especially developmental cognitive psychology—the most influential statistical learning models have been those focusing on transitional probabilities (a type of conditional probability). Several authors (including Amso & Davidow, 2012; Perruchet & Pacton, 2006) have pointed out similarities between implicit learning research and statistical learning research in cognitive psychology. Transitional probability models have mostly been applied to problems of language acquisition, for example how a continuous stream of sounds can be segmented into words. However, since (first) language acquisition (by infants/children) is often considered an example of induction, abstraction or implicit rule-learning, any potential mechanisms for learning that come from this research tradition may be applicable more broadly in implicit learning paradigms. A transitional probability is a type of conditional probability—given syllable X, what is the probability that syllable Y will follow? A low transitional probability between two elements signals a boundary, such as the boundary between two words. Whether implicit learning can be explained by conditional probability models alone or whether a computational model of implicit learning requires more sophisticated models is an open question. Therefore, throughout this paper we will critically address the proposal (by Amso & Davidow, 2012 and Perruchet & Pacton, 2006) that implicit learning could be reduced to conditional probability learning.

Next we will distinguish between implicit memory and implicit learning, and then discuss how implicit learning fits in the remaining branch of the multiple memory systems taxonomy, procedural memory.

Implicit learning is not implicit memory: Memory stages and intentional vs. incidental acquisition . "Implicit memory" is a term sometimes confused with "implicit learning." Rarely, "implicit memory" is used to refer to all non-declarative memory processes or phenomena; in this sense, it would encompass implicit learning. However, usually "implicit memory" is used in a narrower sense to refer to a particular non-declarative memory phenomenon; when used this way, "implicit learning" and "implicit memory" do not overlap (Reber, 2013; Seger, 1994). To delineate the difference between the typical use of "implicit memory" on the one hand and "implicit learning" on the other requires a brief explanation of some additional memory research concepts: (temporal) stages of memory and "incidental" vs. "intentional" learning.

Memory researchers distinguish between three "stages" of memory: encoding (which includes acquisition and consolidation), maintenance (or storage), and retrieval. Encoding refers to the formation of a memory from percepts and sensory input. Maintenance refers to the process that preserves memories that are not being actively utilized at a given time. Retrieval is the process by which previously stored memories are used to create conscious representations or to perform learned behavior (Tulving & Thompson, 1973; Tulving & Craik, 2000). Memory disorders can affect any or all of these three stages of memory, and the neural correlates of these stages are not identical (Squire, 1987). When memory researchers speak of learning, they are primarily concerned with the encoding stage; however, to demonstrate that learning has taken place, some later performance requiring memory retrieval is required. As Squire so clearly states: "Learning is the process of acquiring new information, while memory refers to the persistence of learning in a state that can be revealed at a later time" (Squire, 1987, p. 3)

Another distinction drawn by memory researchers is between "incidental" and "intentional" learning. Intentional learning refers to learning of material (e.g. a list of words or a motor sequence) that the experimental subject is directed or instructed to learn. In contrast, incidental learning takes place when the experimental subject learns material that he or she is exposed to in some way, but not directed to remember or learn. For example, a subject might be exposed to a list of words and asked to do some sort of non-memory task (e.g. generate rhyming words or make semantic decisions); the subject might later be tested for his or her memory of the words. This memory for the words in the list would be considered the result of incidental learning, since the subject was not instructed to remember the words and therefore is presumed to not have made an effort to remember the words. Depending on the experiment design and stimuli, incidental and intentional learning can take place simultaneously (McLaughlin, 1965; Nissen & Bullemer, 1987; Song, Howard, & Howard, 2007). For example, researchers may contrast memory for words learned intentionally with memory for words learned incidentally within one experiment.

Both of these distinctions are part of the difference between the concepts of *implicit memory* and *implicit learning* in memory research. Here, the latter part of Reber's definition—learning that takes place "independently of conscious attempts to learn"—is relevant. "Implicit memory" is the term assigned to the phenomenon of involuntary, effortless, incidental *retrieval* of previously learned or known information. The way the previously learned information was acquired is not a defining criterion for implicit memory: it could be either incidental or intentional. Generally, implicit memory studies use priming tasks and are interested in the retrieval phase. As Schacter (1987, p. 501) put it, implicit memory is “when previous experiences facilitate performance on a task that does not require conscious or intentional recollection of those experiences.” In contrast, implicit learning studies are concerned with *acquisition* of knowledge or skills that takes place without awareness; in this case the *acquisition must be incidental* and the acquisition stage is the stage of interest. Retrieval could potentially be either incidental or intentional, though if the learned material remains below the subject’s threshold of awareness, intentional retrieval may not be possible[4]. That is, the experimental subject could demonstrate implicit learning either in ways he is aware of (e.g. answering questions) or ways he is not aware of (e.g. performance measures such as accuracy or reaction time) (Buchner & Wippich, 1997; Seger, 1994). See Figure 2 for a tabular comparison of implicit memory and implicit learning.

Procedural memory and implicit learning. In addition to conditioning and priming, amnesic patients demonstrate the ability to learn complex new tasks. Importantly, these tasks cannot be successfully performed by simple stimulus-response association: instead, some inductive process must take place to create an abstract representation of a sequence, category, or other data structure. This preserved ability to learn complex tasks is now referred to as procedural memory. While amnesics, who lack MTL structures, are able to learn these tasks at rates similar to neurotypicals, patients

[4] Studies that use debriefing after implicit tasks often find explicit knowledge has been learned in SRT (Knee, Thomason, Ashe, & Willingham, 2007; Song et al., 2007), and sometimes PCT (Gluck et al., 2002; A. L. Price, 2009), but not usually AGL (Reber, 1993; Knowlton et al. 1994, 1996).

with disorders of the basal ganglia (such as Parkinson's disease (PD) or Huntington's Disease (HD)) generally cannot learn these tasks as well as neurotypicals (J. V Filoteo, Maddox, & Davis, 1998; Foerde & Shohamy, 2011; B J Knowlton, Mangels, & Squire, 1996). These findings suggest that while MTL structures may not be necessary for procedural memory, the basal ganglia may be critical to procedural memory. Neuroimaging studies and animal studies (Packard & Knowlton, 2002) support this conclusion (Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Willingham, Salidis, & Gabrieli, 2002). Neurotypicals who learn these tasks often are unable to report or explain how they are performing the task, despite their high performance measures (Sanchez, Gobel, & Reber, 2010).

Since procedural memory is preserved in amnesics, generally inaccessible to consciousness in neurotypicals, and cannot be explained by simple learning mechanisms such as conditioning, it seems to fit well with Reber's definition of implicit learning. Seger (1994) agrees that—depending on the definition of procedural memory used—implicit learning and procedural memory may be more or less coextensive. However, she also points out an important exception: In Anderson's (1987) conception of skill learning, a deliberate, effortful process is transformed into a fluent, automatic process through repeated practice; in this model, the resulting knowledge-how is referred to as "procedural memory." This definition of procedural memory is not compatible with Reber and Seger's definitions because it requires the initial declarative approach. In contrast, we now know that skills can be learned implicitly without a declarative phase (Song et al., 2007; Willingham & Goedert-Eschmann, 1999).

Furthermore, since “procedural memory” could refer to the acquisition, consolidation, storage, or retrieval of procedural knowledge, it might be more precise to say that implicit learning is more or less coextensive with procedural memory acquisition[5].

Procedural memory can be further divided into “habit learning” and “skill learning.” These two types of learning are both spared in amnesic patients (Seger & Spiering, 2011), but differ from each other qualitatively. The tasks that were first reported to be preserved in amnesic patients—rotary pursuit (Corkin, 1968), and mirror tracing (Milner, 1962)—are examples of skill learning. Although the original examples were sometimes explained away as “merely” motor learning, subsequent studies established that not only motor skills, but also perceptual skills such as mirror reading (Cohen & Squire, 1980) and cognitive skills such as probabilistic classification (Knowlton, Squire, & Gluck, 1994) could be learned by amnesics at levels similar to neurotypical adults. Neuroimaging studies established that neurotypical adults show similar patterns of brain activity for both motor skill learning and cognitive skill learning (Poldrack et al., 1999), suggesting that the mechanisms of skill learning are mostly common across cognitive and motor skill learning.

However, neither Seger’s nor Reber’s definition of implicit learning provide criteria that could distinguish between habit learning and skill learning, and in fact, not only tasks that are considered habit learning tasks, but also tasks that are considered skill learning tasks have been used in studies of implicit learning. As Seger says, “Implicit learning includes both habits and skills. Within the domain of skills, implicit learning

---

[5] This distinction is increasingly important as a rich vein of research on procedural memory consolidation is currently being tapped (e.g. Brown & Robertson, 2007; Janacsek & Nemeth, 2012).

cuts across the distinctions among perceptual, motor, and cognitive skills, including tasks from each that involve learning patterns or rules that exist in the stimuli and excluding tasks that do not.”

Nevertheless, it may be helpful to keep in mind the potential distinction between skill learning and habit learning. They differ not only in task demands, but also in neuropsychological dissociations. Patients with HD and PD demonstrate impaired performance in habit learning tasks such as probabilistic category learning (Knowlton, Mangels, et al., 1996), implicit visual category learning (Ashby, Noble, Filoteo, Waldron, & Ell, 2003; Maddox, Aparicio, Marchant, & Ivry, 2005; Price, Filoteo, & Maddox, 2009), and artificial grammar learning (Smith, Siegert, & McDowall, 2001) but demonstrate some preservation of function in skill learning tasks such as rotary pursuit, mirror tracing, and serial response task (Gabrieli, Stebbins, Singh, Willingham, & Goetz, 1997; Smith et al., 2001). The possibility that skill learning and habit learning have different neural correlates suggests that they could potentially also vary in their developmental trajectories and degree of inter-individual variation. These issues are discussed in Chapter 3 and Chapter 4.

Operationalizing Implicit Learning: The Tasks

Implicit learning has been proposed as an explanation for many real-world situations, such as riding a bike or learning unspoken social conventions (Reber, 1993; Evans, 2008). However, to study implicit learning in a controlled environment, several experimental tasks are conventionally used: artificial grammar learning (AGL), serial response tasks (SRT), and probabilistic classification tasks (PCT). In a sense, these tasks represent an operationalized definition of implicit learning. It is based on these tasks in particular that researchers make inferences about implicit learning in general. Therefore, a careful explanation of these tasks is necessary to completely explain what is meant by implicit learning in the context of cognitive psychology and cognitive neuroscience research.

Artificial grammar learning (AGL). Reber's work on implicit learning used the Artificial Grammar Learning (AGL)[6] task extensively and almost exclusively. In AGL paradigms, a Markov chain finite-state model (FSM) (see Figure 3) is used to generate unpronounceable, meaningless letter strings. Letter strings generated by the FSM[7] are considered "grammatical." This underlying structure addresses the "abstract" criterion. Working initially with healthy young participants, Reber found that after exposure to example letter strings created by the FSM, participants were able to distinguish novel grammatical letters strings from novel non-grammatical[8] letter strings at rates better than chance (Reber, 1993, Chapter 3). However, participants were unable to explain their performance or describe any rules governing the creation of letter strings. Thus AGL meets criteria for implicit learning as incidental learning or learning without awareness. Later, Knowlton and colleagues demonstrated that amnesics could also perform the AGL task (Knowlton, Ramus, & Squire, 1992). Thus AGL also meets criteria for "implicit" in the sense of non-declarative i.e. not MTL-dependent memory. However, Parkinson's and Huntington's patients have also demonstrated intact AGL, suggesting that it may not share basal ganglia dependence with the other measures of implicit learning.

---

[6] The idea of using finite-state grammars as material to be learned in a laboratory task actually originated with George Miller's "Project Grammarama" in 1957. Miller was focused on explicit learning but found that the task was too difficult for students to master in a single laboratory session and therefore did not meet his needs (Miller, 1958; Chomsky & Miller, 1958; Mathews & Cochran, 1997).

[7] i.e. following the rules of the FSM, could have been created by the Markov chain

[8] i.e. letter strings that could not have been created by the Markov chain—violation of its "rules"

How do participants perform the AGL task i.e. what is learned or abstracted? Many theories (see Pothos, 2007) have been proposed as attempts to explain this question. Are participants able to unconsciously abstract the rules of the Markov chain based only on examples? Most authors agree that participants are most likely not able to abstract the exact rules of the original Markov chain, but perhaps to form a set of rules that are compatible with it, given the stimuli they have seen. The major competing explanation is that participants discriminate between grammatical and non-grammatical test stimuli by using "chunk strength[9]." The chunk strength of an AGL letter string is a measure of the frequency with which its components occurred in the training (exposure) stimuli. Each bigram or trigram[10] is considered a "chunk" and its frequency in the training set can be counted. The chunk strength of a string is the average of its bigram- and trigrams' frequencies. Thus a string made up of frequently occurring bigrams and trigrams has a high chunk strength and a string made up of rarely occurring chunks has a low chunk strength.

Perruchet and Pacton (2006) point out that although some researchers explain the mechanism underlying chunk learning as based on raw frequencies or simple associations, nevertheless conditional frequencies (as used in statistical learning) could play a role (Perruchet & Pacton, 2006).

[9] Meulemans and van der Linden (2003) suggested that chunk learning may represent a form of explicit learning that takes place in tandem with abstract implicit learning, but Knowlton and colleagues (B J Knowlton & Squire, 1996) demonstrated that amnesics use both chunk strength and abstract knowledge to perform AGL.

[10] Unit of two or three (respectively) consecutive letters. For example, the letter string "XJVT" contains the bigrams XJ, JV, and VT and the trigrams XJV and JVT. Each bigram or trigram is a "chunk."

Serial response task (SRT). First described by Nissen & Bullemer (1987), a serial response task requires participants to respond to sequential stimuli. In many cases participants view (usually) four positions on a computer screen (linear or quadrants); a target stimulus can appear in any of the four positions. Participants are instructed to press the key or button corresponding to the position when it contains a target. Unknown to participants, sometimes the target moves in a repeating pattern (sequence), so if the positions are labeled [1 2 3 4] the target might move repeatedly in the pattern [121432134]. Reaction times to sequence and non-sequence ("random") trials/blocks are compared and decreased reaction time to sequence trials (relative to random trials) is seen as evidence of sequence learning.

To exclude the possibility that the sequence is learned simply by associating each position with the following position, sequences used must be second-order conditional (Reed & Johnson, 1994), meaning that a given position cannot be predicted from the preceding position alone, but can be from the two preceding positions together. Concerns have been raised that participants may be learning the sequence explicitly. For this reason, most implementations of the SRT include a debrief/post-test in which the participant is asked whether they noticed a pattern, and if so whether they can state or recreate it. Under some conditions, some participants do display explicit sequence knowledge (Foerde, Poldrack, & Knowlton, 2007; Song et al., 2007); this can be taken into account when calculating implicit learning scores. To further protect against explicit sequence learning, some studies use an Alternating Serial Response Task (ASRT). In this modified version of the SRT, random trials are interspersed with the sequence trials, so that if the sequence was originally [1 2 3 4], in the ASRT the participant would see [1 r

2 r 3 r 4 r], where r represents a randomly inserted position. This insertion results in high frequency triplets and low frequency triplets (e.g. sequence-random-sequence forms a high frequency triplet and random-sequence-random forms a low frequency triplet). In ASRT studies, differences in RT between high frequency triplet trials and low frequency triplet trials are taken as evidence of implicit sequence learning.

Neuropsychological evidence supports the non-MTL-dependence of SRT. Amnesics demonstrate intact SRT learning (Nissen & Bullemer, 1987), but HD patients (Willingham & Koroshetz, 1993) and PD patients (Ferraro et al., 1993) do not.

Can transitional probabilities explain sequence learning in SRT and ASRT tasks? Given the second-order conditional nature of the sequences in SRT and the non-adjacent dependencies in ASRT, conditional probabilities based on proximity are probably not sufficient to explain implicit sequence learning in SRT. What is actually learned in SRT and how it is represented remains a debated question and may be beyond the scope of this thesis (Forkstam & Petersson, 2005; Reed & Johnson, 1994; Ruenger & Frensch, 2008; Stefaniak, Willems, Adam, & Meulemans, 2008).

Probabilistic classification task. In the probabilistic classification task, participants learn to match cue stimuli to targets based on trial-and-error. That is, given a cue stimulus, they guess which target will be correct, and then they receive feedback (correct/incorrect) after each guess. Importantly, the cues are only probabilistically (and arbitrarily) related to the targets; any particular cue only yields “correct” feedback with its target some of the time. This makes PCT unlike other category-learning tasks in which category membership is deterministic. One reason for the probabilistic relationships is to prevent explicit learning of cue-target associations; another may be to simulate the noisy inputs that *in vivo* category learning is subject to. The original and perhaps most frequently used implementation of PCT is the "weather prediction task," so called because the targets are either "sun" or "rain" (Knowlton et al., 1996; Knowlton, Squire, & Gluck, 1994).

Amnesics similarly show normal learning early, but impairments late in the task (relative to normal controls) (Knowlton, Squire, & Gluck, 1994); this is in keeping with the finding that in normal controls, early learning on the task is implicit, but late in the task explicit learning may play a role. On the other hand, Parkinson's and Huntington's patients are impaired at the task from the beginning (Knowlton et al., 1996). Thus at least early learning in PCT appears to be non-MTL-dependent, and all PCT learning seems to require intact basal ganglia.

The question of how people solve the weather prediction task has been addressed most notably by Gluck and colleagues (Gluck, Shohamy, & Myers, 2002). However, these accounts have focused on possible strategies rather than computational mechanisms for solving the task. Probabilistic classification involves many-to-one mapping and therefore is probably executed by processes similar to those used in other types of categorization tasks. Although the feedback for cue-target associations is probabilistic, nevertheless it is assumed that participants begin to associate one target preferentially with each of the cues (i.e. even though Cue X-Target Y gives "correct" feedback only 75% of the time, participants adopting an "optimal" strategy should choose Target Y any time they see Cue X). That is, an optimal strategy includes abstracting a deterministic relationship from probabilistic data. While this process may be explainable in terms of existing models of induction, transitional probabilities alone cannot account for it.

In summary, implicit learning: a) is usually unconscious; b) does not depend on the MTL-system; c) refers to memory acquisition rather than retrieval; d) is specific to abstract representations of complex stimulus structures; e) overlaps with both habit learning and skill learning within the broader category of procedural learning; f) is experimentally investigated with tasks such as AGL, SRT, and PCT, none of which can be completely explained by transitional probability learning alone. With these facts in mind, we can now turn to the question of how implicit learning may be useful in educational contexts.

Chapter Three: Implicit Learning and Education

The relationship between the implicit/explicit distinction in cognitive psychology and neuroscience research and the procedural/conceptual distinction in math education research is mapped. Suggestions are made for relating implicit learning research to instructional strategies. Different approaches are suggested for habit learning and skill learning paradigms.

Educational researchers, and in particular math education researchers, have long recognized that multiple types of knowledge contribute to academic learning and performance. However, that recognition has not always corresponded to the different types of knowledge or memory identified by cognitive psychologists and neuroscientists. In this chapter I attempt to find correspondences between these two understandings as well as proposing additional ways basic cognitive research on implicit learning could be applied to educational practice.

The Representation of Procedural Memory

In Chapter 1 we made the point that “implicit learning” refers to memory acquisition (in contrast to “implicit memory” which can be more focused on retrieval i.e. in priming paradigms). However, when we speak of “knowledge representation” or “procedural knowledge,” we are no longer referring to acquisition, but rather to the storage phase of memory. This is not necessarily a contradiction or violation of the definition of “implicit learning” laid out in Chapter 1 because researchers acknowledge that the manner in which knowledge is acquired can affect the form of its representation (Stadler, 1989, 1992; Seger, 1994; Morris, Bransford, & Franks, 1977). For this reason, Reber (1993) discusses “implicit learning” hand-in-hand with the idea of “tacit knowledge," i.e. knowledge that is generally unavailable to consciousness but nevertheless able to influence behavior[11].

The representation of implicitly learned knowledge has been characterized as "implicit (versus explicit) access, abstract (versus concrete), structural (versus surface-based), complex (versus simple)" (parentheses added) and the use of implicitly learned knowledge has been characterized as "automatic (versus controlled) processing" (Forkstam & Petersson, 2005). Since most psychology/neuroscience researchers accept a substantial overlap between implicit learning and procedural memory acquisition (as well the overlap between implicit learning and the subdivisions of procedural memory, skill learning and habit learning), therefore the properties attributed to procedural memory (storage) can be applied to implicitly learned knowledge with relatively little controversy[12]. Procedural memory representations have been characterized as complex and not available to consciousness (Lewicki, Hill, & Bizot, 1988; Willingham, Nissen, & Bullemer, 1989). Similarly, the habit learning has often (but not universally) been defined as "inflexible (Hirsh 1974; Mishkin 1984), slow or incremental (Mishkin, 1984), unconscious (Squire & Zola-Morgan, 1988, 1991), [and] automatic (Shiffrin and Schneider, 1977)" (Seger & Spiering, 2011, p. 2). Since skill learning research tends to focus on the acquisition, discussion about the representation of skill knowledge is scarce, but what can be found is highly consistent (if not overtly identified) with the descriptions

---

[11] Although implicitly learned knowledge can be tacit, it is not necessarily limited to pre-conscious retrieval. Some evidence supports the idea that under certain circumstances, implicitly learned information or knowledge can yield information that is accessible to consciousness (Ruenger & Frensch, 2008; Ruenger, 2012; Seger, 1994).

[12] Note that I am deliberately avoiding the term "procedural knowledge" to refer to the product of procedural memory acquisition, habit learning, skill learning, or implicit learning. This is because the term "procedural memory" is no longer generally used by cognitive psychologists and cognitive neuroscientists. Furthermore, since "procedural knowledge" IS used by education researchers and in a way that may NOT be coextensive with procedural memory, I am not using "procedural knowledge" in the context of "procedural memory" to avoid confusion with the construct from education research.

of procedural memory, implicitly acquired knowledge, and habit learning knowledge above (Doyon, 1997; Ouellet, Beauchamp, Owen, & Doyon, 2004).

Before moving on to explain how implicit learning and procedural memory could be useful constructs in education research, a discussion of the related distinction between procedural knowledge and conceptual knowledge in education research is necessary.

Procedural Knowledge and Conceptual Knowledge in Education Research

Math education research has its own deep history of distinguishing between knowledge types, again congruent with—if not influenced by—Ryle's distinction between "knowledge-that" and "knowledge-how" (Brownell, 1945; Hiebert & Lefevre, 1986; Skemp, 1976 cited in Star, 2005) In this research tradition, the opposing constructs are "conceptual knowledge" and "procedural knowledge," defined by Hiebert and Lefevre (1986) as follows: conceptual knowledge [is]

> knowledge that is rich in relationships. It can be thought of as a connected web of knowledge, a network in which the linking relationships are as prominent as the discrete pieces of information. Relationships pervade the individual facts and propositions so that all pieces of information are linked to some network. (pp. 3-4)

In contrast, procedural knowledge is described in these terms:

> One kind of procedural knowledge is a familiarity with the individual symbols of the system and with the syntactic conventions for acceptable configurations of symbols. The second kind of procedural knowledge consists of rules or procedures for solving mathematical problems. Many of the procedures that

> students possess probably are chains of prescriptions for manipulating symbols. (pp. 7-8)

A complicating factor pointed out by Star (2000, 2005, 2007) and others (De Jong & Ferguson-Hessler, 1996) is that frequently in math education research, the distinction between type of knowledge (conceptual vs. procedural) is conflated with a distinction between quality of knowledge (deep vs. superficial). Star (2005, 2007) has lucidly argued that this is not a necessary conflation and that in fact a case can be made for both superficial conceptual knowledge as well as deep procedural knowledge. For example, Star (2005) argues that while Hiebert and Lefevre's (1996) discussion of procedural knowledge characterizes it primarily in terms of algorithms (sets of sequential steps), nevertheless heuristics ("rules of thumb" or "abstract procedures for problem solving" (Star, 2005, p. 407))could also be examples of (non-algorithmic) procedural knowledge.

Although the mapping between the conceptual/procedural distinction in math education research and the declarative/procedural distinction in cognitive psychology/neuroscience is not strictly bijective, nevertheless, constructive parallels can be drawn.

Since Hiebert and Lefevre's (1996) definition of conceptual knowledge mentions "facts and propositions," the best match within the multiple memory taxonomy may be with semantic memory (within declarative memory). Semantic memory is usually defined as memory for facts (Tulving, 1983; Squire & Zola-Moran, 1989, 1991), is accessible to consciousness, and its acquisition depends on medial temporal lobe (MTL) structures. (Within declarative memory, it is contrasted with episodic memory, autobiographical memory for events). However, Hiebert and Lefevre (1996) emphasize that the facts are embedded in a network of relationships among facts (as opposed to isolated facts). This is not unlike Ma's (1999) concept of "knowledge packets" for mathematics education—sets of related concepts whose knowledge and knowledge of relations is necessary for deep mathematical understanding. Star (2005, 2007) questions the necessity of these interrelationships and postulates that superficial conceptual knowledge may consist of isolated facts. This nuance is not captured by the multiple memory systems taxonomy; definitions of semantic memory make no distinction between isolated facts and those well-embedded in semantic networks[13]. Schema theory in cognitive psychology also stresses the importance of interrelated concepts (Schank & Abelson, 1977; Rumelhart, 1980; Mandler, 1984), but research connecting schema theory to the multiple memory systems perspective is all but non-existent. Both rote-memorized information and information richly connected to other concepts are considered equally valid examples of semantic memory.

Although the relationship between "procedural knowledge" in math education research and "procedural memory" in cognitive neuroscience research is more complicated, it may also be more productive. Procedural memory (skill learning and habit learning) and implicit learning have been characterized as unconscious, gradually acquired, complex, abstract, and automatic. Perhaps the most striking difference between the constructs of "procedural knowledge" and "procedural memory" is the criterion of unconscious learning, or knowledge that is unavailable to consciousness. Math education research definitions of "procedural knowledge" do not require this criterion, and in fact in

[13] Such a difference could be explored in patients with semantic dementia. When diagnosis is made, especially if patients are diagnosed early, a reaction time/priming paradigm could be used to map the relationships among concepts and identify a bank of "richly embedded" and "isolated" concepts. One possibility is that as the disease processes, knowledge of richly embedded concepts may be affected later than isolated concepts (Bright, Moss, Stamatakis, & Tyler, 2008; Nishio & Mori, 2009).

the case of "deep procedural knowledge," lack of access to consciousness may be a criterion for exclusion (Star, 2005, 2007).

However, other characteristics of skill learning and habit learning align well with descriptions of procedural knowledge in math education research. Algorithms require or consist of a set of ordered steps; this is almost identical to many definitions of skill in skill learning. Both algorithms and skills require practice (gradual learning) and can become automatic. Similarly, the definition of heuristics as "procedures for abstract problem solving" is not unlike the view of habit learning as acquisition of complex, abstract representations that are the applied to make decisions (Seger & Peterson, 2013; Seger & Spiering, 2011).

Other types or aspects of "deep procedural knowledge" as described by Star and others may require cognitive functions outside the realm of procedural learning. For example, Star (2005) states that "Deep procedural knowledge would be knowledge of procedures that is associated with comprehension, flexibility, and critical judgment.." and cites VanLehn's proposal of teleological knowledge of a procedure, "meaning knowledge of its design or justification for its use" (VanLehn 1990, cited in Star, 2005). Under the current state of understanding in cognitive neuroscience, comprehension, critical judgment, and justification for the use of a procedure are functions or types of knowledge that are not included in implicit learning or procedural knowledge. Comprehension and knowledge of a procedure's design and meaning almost certainly require declarative memory. Furthermore, critical judgment and flexibility require executive functions (such as attention, working memory, problem solving, and task flexibility and planning). Although cognitive functions are not identical to their neural correlates, nevertheless if two functions have different (necessary) neural correlates, the case for their non-identity is strengthened (Poldrack, 2006; Price & Friston, 2002). Executive functions rely on the frontal lobes, while procedural learning requires the basal ganglia (Alvarez, Emory & Emory, 2006). Patients with damage to the frontal lobes (such as Alzheimer's patients) demonstrate normal-like procedural learning abilities but impaired executive functions, while patients with basal ganglia disorders (such as Parkinson's disease and Huntington's disease) have generally intact executive functions but largely impaired procedural learning abilities (Arroyo-Anllo, Ingrand, & Gil, 2012; Filoteo, Maddox, Ing, & Song, 2007; Gabrieli et al., 1997; Knowlton, Squire, et al., 1996; Reber, Martinez, & Weintraub, 2003; Smith & McDowall, 2006).

Some aspects of the math education conception of "procedural knowledge" match well with the existing "branches" of the multiple memory systems taxonomy. Specifically, algorithms and heuristics may be highly coextensitve with skill learning and habit learning, respectively. In contrast, deep procedural knowledge may require executive functions, declarative knowledge and possibly coordination or connections between procedural and declarative knowledge. A similar relationship exists for the conception of fluency and procedural learning.

Fluency. Influential reports on both reading and mathematics instruction have stressed the importance of fluency. The National Reading Panel (2000) considered fluency "an essential aspect of reading" and defined it as **"**the ability to recognize words easily, read with greater speed, accuracy, and expression, and to better understand what is read. Children gain fluency by practicing reading until the process becomes automatic; guided oral repeated reading is one approach to helping children become fluent readers"

(p. 3-1). The National Research Council (2001) identified procedural fluency as one of its five strands of mathematical proficiency, and defined it as “skill in carrying out procedures flexibly, accurately, efficiently, and appropriately[14]” (p 8-23).

What these definitions share is a belief in the importance of skill learning. Laboratory measures of skill learning include accuracy and greater speed—this can be seen clearly in SRT and PCT. Arguably, participants performing these tasks develop fluency in them, just as educators want students to develop fluency in decoding and in mathematical procedures.

Importantly, both the NRC and NRP definitions of fluency include qualifications that implicate explicit mechanisms: the calls for flexibility, comprehension, and appropriate deployment of skills. However, (just as with some aspects of deep procedural knowledge) through the lens of cognitive neuroscience, flexibility, reasoning, and comprehension invoke executive functions. These abilities must “wrap around” skill learning in order to achieve the desired outcomes for students, but they are not inherently part of skill learning.

Once those qualifiers are stripped away, what’s left is an almost perfect match for procedural learning as defined in the laboratory: not dependent on the frontal lobes or MTL, but requiring practice and repetition. If the desired outcome is fluency, then the thing-to-be-learned is probably a skill, and therefore the appropriate instructional method can probably be informed by research on skill learning. While all of reading is not necessarily a skill amenable to procedural memory, decoding is. Similarly, until math facts are memorized, performing arithmetic operations is a skill. Any set of steps that

---

[14] Conceptual understanding and procedural fluency were two of five “strands” highlighted in the report as aspects of mathematical proficiency. The others are strategic competence, adaptive reasoning, and productive disposition (NRC, 2001).

students are expected to learn and apply later could potentially be treated as a cognitive skill: long division, re-writing sentences into the active voice, solving oxidation-reduction reactions. If fluency is an educational desideratum, then procedural memory is relevant to instructional practice. Even the best teaching for understanding or conceptual knowledge—while critical for deep understanding and higher-order reasoning--can produce only the most limited benefits for fluency. Conceptual knowledge is governed by the declarative memory system and can require as little as one trial or example; fluency is the realm of procedural memory and requires practice.

Relevance to Education

Why should educators be interested in unconscious, automatic, fast processing? Beyond the correspondences to math education research above, there are at least two compelling reasons. First, implicit learning may already be affecting formal education via interference from intuitive theories. Second, research findings on skill learning and habit learning may be applicable to a wide range of content in the K-12 curriculum. I address each of these points below.

Insidious influence of implicit learning. Implicit learning and procedural memory were not originally proposed as constructs to explain learning in carefully controlled laboratory tasks, but rather to explain learning that takes place in vivo. In particular, many researchers have speculated that intuitive theories are acquired implicitly (Lewicki et al., 1988; Reber, 1989). Intuitive theories in many domains have been investigated, including physics, biology, and social relations (Carey, 1985; Rhodes, 2011). While the acquisition and development of intuitive theories represents a remarkable feat of induction from scarce data (Tenenbaum et al., 2011), nevertheless intuitive theories can be an obstacle to formal education.

In science education research, conceptual change models of pedagogy emphasize engaging students' intuitive theories and then trying to present formal theories in a way that supplants the intuitive theories (Strike & Posner, 1985). However, in practice intuitive theories can be highly resistant and difficult to disrupt (Cheng & Brown, 2010; Hatano & Inagaki, 1997; Taber & Tan, 2011).

Furthermore, even if students do not enter with problematic intuitive theories in place, they may develop representations—such as heuristics—that capture only some aspects of the formal theory. In the laboratory, implicit learning has regularly been observed taking place in parallel—perhaps unavoidably—with explicit learning (Song et al., 2007; Willingham & Goedert-Eschmann, 1999; Willingham et al., 2002). Thus when students are explicitly instructed, they may develop and attempt to apply implicitly-learned rules (which may not be as specific or accurate as the explicitly instructed rules). In this way, information that is learned simultaneously implicitly and explicitly can be problematic in a "test phase" (for example, an educational assessment) if students do not have a way to either connect the two representations or to appropriately inhibit one of them (Cheng & Brown, 2010; Ohst, Fondu, Glogger, Nückles, & Renkl, 2014; Sanborn, Mansinghka, & Griffiths, 2013).

For example, one study (McClary & Talanquer, 2011) examined students' use of heuristics rather than explicit rules when predicting the acid strength of a molecule, given its structural formula. The heuristics used by students did not yield as accurate predictions as explicit rule use. These students were (presumably) given explicit instruction in the rules for predicting acid strength from structural formulas: why, then, did they resort to heuristics? One possible explanation is that they were attempting to use what they had learned implicitly, perhaps to conserve cognitive resources, i.e. reduce cognitive load. Based on the information given in the study, it is not clear how the students acquired their idiosyncratic heuristics—they could have developed the heuristics through (implicit) habit learning mechanisms, or through (explicit) hypothesis testing processes. However, since they were given explicit instructions, their motivation for independently developing heuristics through hypothesis testing may have been low. Whether these particular heuristics were implicitly learned or not, this study highlights how students' use of heuristics can be both problematic and difficult to inhibit or repress.

Perhaps conceptual change approaches could be more effective if instead of *only* engaging students' intuitive theories via declarative knowledge, formal theories could be presented to the procedural learning system. To replace naïve intuitive theories with formal theories in a way that is equally robust and automatic requires engaging the system that developed the intuitive theories. Presenting abstract information to the procedural learning system requires neither subliminal messages nor Skinnerian conditioning, but rather, carefully structured repeated experiences with feedback. In the next section I will elaborate on how models of implicit learning from the laboratory could potentially inform instructional strategies.

Opportunities for skill learning and habit learning in the K-12 curriculum. Recall from Chapter 2 that skill learning and habit learning are both types of procedural memory/learning and/or implicit learning. Neuropsychological findings (Filoteo et al., 1998; Knowlton, Squire, et al., 1996) suggest that they are distinguishable functions[15], so we will distinguish between them here. Furthermore, skill learning seems to map well onto algorithmic learning, while habit learning may be more applicable to the internalization of heuristics. Therefore, we will discuss first how skill learning, then habit learning may be applicable to conventional K-12 curricula.

As described above, skill learning research findings may be readily applicable to any algorithmic aspect of a curriculum, such as long division, orthographic decoding, solving systems of equations, solving oxidation-reduction reactions, or converting sentences from passive to active voice. Examples of algorithmic procedures in K-12 math, science, and even language arts abound. Teaching and learning of any of these procedures could benefit from research on skill learning and from a fundamental orientation that procedural fluency requires engaging the procedural memory system through practice. Examples of research findings on skill learning that could be applicable to these types of algorithms include findings on memory consolidation (Brown & Robertson, 2007; Janacsek & Nemeth, 2012), feedback (Maddox & Ing, 2005; Rendell, Farrow, Masters, & Plummer, 2011; Seger & Miller, 2010; Wulf, Shea, & Lewthwaite, 2010), and blocked practice (Merbah & Meulemans, 2010), among others. The connection between skill learning research and algorithm teaching and learning is straightforward.

Habit learning, on the other hand, may not appear as readily applicable to K-12 content. In the previous section, a connection was drawn between habit learning and heuristic learning; because of the connation of "heuristic" as "shortcut" is not completely

[15] Though studies with neurotypicals often appear unconcerned about the difference.

accurate, this may not necessarily sound like a good thing. However, habit learning may apply more broadly than this narrow construal of "heuristic."

Intuitive theories are the result of abstracting general rules or properties from repeated experiences or examples. Although intuitive theories may fall short of formal theories (for example in physics) in their accuracy and predictive power, there is no *a priori* reason that the same mechanisms that create intuitive theories could not be used to internalize formal theories; all that is required is the correct input to the system. Learning of complex, structured domains—including but not limited to category membership—is required in many areas, across content areas in K-12 education. Habit learning is a mechanism for induction—for extracting abstract structure from examples. Any type of taxonomy or classification structure is theoretically amenable to habit learning approaches: classifying types of [math] problems to select the best strategy for solving them, recognizing patterns in data (such as strength of correlations), or classifying organisms into their phyla and subcategories. In any situation where this is the case, research on habit learning could provide useful guidance.

Although the information could be learned by the declarative memory system, when the desired outcome is for students to demonstrate fluent, automatic use of such information, and particularly when the structure must be induced from examples, habit learning could provide an advantage. Furthermore, it appears to be the case that sometimes implicit learning mechanisms may be better at learning complexly structured domains than explicit learning systems. For example, a medical education study trained novices (students or laypeople without medical knowledge) to recognize and classify (diagnose) skin lesions without any recourse to explicit rules or explicit knowledge of skin lesions (Aldridge, Glodzik, Ballerini, Fisher, & Rees, 2011). This study found that the novices trained this way reached a diagnostic accuracy of 98%, compared to a control group which reached only 33% accuracy in diagnosing the skin lesions. The authors state, “novices achieved this high degree of accuracy without any use of explicit definitions of likeness or rule-based strategic information” (p. 279). It may be the case that when “professional expertise” in applied areas (such as medicine) relies on fluent categorization, such expertise can be fostered by structured exposures to exemplars in implicit learning paradigms.

Effective implementation of habit learning paradigms in the classrooms necessitates attention to what is known about implicit learning from the lab. A study on implicit learning of atomic bonding rules with 11-12 year olds demonstrates the pitfalls of ignoring this point (Sætrevik, Reber, & Sannum, 2006). This study used a paradigm very much like an artificial grammar learning (AGL) study: participants (students) were exposed to correct molecule diagrams and given instructions to memorize them in a training phase, then in a testing phase they were shown new molecule diagrams (both correct and incorrect) and asked to mark each as “right” or “wrong.” The students in this implicit condition did not do better than students whose training phase was irrelevant stimuli, and did worse than students who received explicit instruction. The authors therefore concluded that AGL-style implicit learning was not as effective as explicit instruction.

However, it should be noted that AGL studies show the lowest levels of accuracy of all the standard laboratory implicit learning tasks (generally about 60-65%, compared to 90% in probabilistic classification and serial response tasks). This may be due to the lack of feedback in the training phase. We know from cognitive neuroscience that feedback is critically important for habit learning (Foerde & Shohamy, 2011; Packard & Knowlton, 2002; Seger, 2006). If the students had been presented the training molecule diagrams in a feedback learning paradigm, the results might have been different. Furthermore, the training set consisted of only 20 examples; a more extensive training set, or repetition of the stimuli could further have increased learning in the implicit learning group. While explicit knowledge of the rules for atomic bonding is best learned explicitly, fluency and automaticity in recognizing valid vs. invalid molecular diagrams could potentially be established via an implicit learning paradigm.

Research on optimizing inductive learning and category learning is quickly moving beyond general recommendations for feedback and many trials. Recent findings on the effects of the order of examples presented (Sandhofer & Doumas, 2008), the timing and spacing of trials (Birnbaum, Kornell, Bjork, & Bjork, 2013; Vlach, Sandhofer, & Bjork, 2014), salience of certain stimulus features (Noh, Yan, Vendetti, Castel, & Bjork, 2014) and other aspects of training session construction (Lindsey, Mozer, Huggins, & Pashler, 2013) have all recently emerged. These findings could easily be adapted to develop “best practices” for applied implicit learning paradigms.

Some properties of habit learning may be problematic for application to educational settings. In particular, the idea that that habit learning is unconscious and inflexibly automatic may be problematic. Furthermore, some questions have been raised about the extent to which implicitly learned information can transfer across contexts (Bitan & Karni, 2003; Gomez, Gerken, & Schvaneveldt, 2000; Gomez & Schvaneveldt, 1994; Gomez, 1997). Regarding conscious access and inflexibility, Seger and colleagues

(Seger & Spiering, 2011) have recently challenged these assumptions about habit learning in general and suggested that motor (but not executive) corticostriatal loops are limited this way.

Future research can address and clarify these issues, as well as questions of how and whether procedural knowledge and declarative knowledge can be combined or coordinated. However, some fundamental questions about implicit learning have not yet been empirically verified. One of these is the question of developmental trajectory: most studies of implicit learning have used adult research participants; will these results generalize to children? How does implicit learning change during development from infancy to adulthood? This question must be addressed before habit learning and skill learning best practices are applied to K-12 education take place. Similarly, whether stable individual differences in implicit learning exist, or whether implicit learning ability can be reliably measured in individuals, remains an unanswered question. Chapters 4 and 5 address these questions empirically with a set of experiments.

Chapter Four: Developmental Differences in Implicit Learning

This study addresses the question of developmental invariance in implicit learning by comparing performance on implicit learning tasks between 10 year old children and adults. Unlike previous developmental studies of implicit learning or procedural learning, which used only a single task to operationalize implicit learning, the current study used several of the most widely used implicit learning tasks: Artificial Grammar Learning (AGL), Serial Response Task/Serial Reaction Time Task (SRT/SRTT), Probabilistic Classification Task (PCT), rotary pursuit, and mirror learning. While the children and adults did not differ on gross performance of the tasks, item analysis revealed that adults were more sensitive to stimulus frequency than children were.

Implicit processes in general have long been assumed to be developmentally invariant (Lloyd & Newcombe, 2009). Implicit learning is no exception to this trend. However, empirical support for this assumption is sparse and the results are contradictory. In the current study we attempt to address the issue of developmental invariance in implicit learning using a combination of implicit learning tasks.

Background

Converging evidence from neuropsychology, neuroimaging, human behavioral studies, and animal studies support the multiple memory systems understanding of human memory (Gabrieli, 1998; Squire, 2004). As described at length in Chapter 2, declarative memory is memory for facts and events, is accessible to consciousness, and is dependent on the medial temporal lobe (MTL). While there are several kinds of non-declarative memory (not accessible to consciousness and not dependent on MTL), procedural memory is memory for skills and habits and depends largely on the basal ganglia and cerebellum (Reber, 2008, 2013; Seger, 1994, 2006).

Declarative memory ability and the MTL develop gradually from childhood to adulthood; both behavioral evidence (Kail, 1990; Ghetti, Lyons, & DeMaster, 2012; Ofen et al., 2007) and neuroimaging evidence support this (Gogtay et al., 2004; Gogtay et al., 2006; Raznahan et al., 2011; Shaw et al., 2008; Thompson et al., 2005). However, the extent to which procedural memory develops during childhood has been questioned by researchers and theorists, some of whom have proposed that procedural memory is developmentally invariant (Reber, 1993; Parkins, 1998).

One reason for this may be the assumption in dual-systems theories that automatic, "reflexive" systems are "early evolving" and "shared with animals" (Evans, p. 258. Tables 1 and 2). Another reason may be reports of sophisticated learning mechanisms in infants could potentially be explained by implicit learning (Lloyd & Newcombe, 2009). (As discussed in Chapter Two, implicit learning and procedural learning are potentially co-extensive terms).

Evidence for developmental invariance of implicit memory is mixed. Neuroanatomically, the basal ganglia and cerebellum mature gradually from childhood to adulthood, in a pattern that does not suggest developmental invariance (Lenroot & Giedd, 2006; Sowell, Trauner, Gamst, & Jernigan, 2002; Toga, Thompson, & Sowell, 2006). The few behavioral studies of implicit learning in childhood that have been conducted to date reach conflicting conclusions.

As described in Chapter 1, typical laboratory tasks for studying implicit memory include artificial grammar learning (AGL), serial response tasks (SRT), and probabilistic classification tasks (PCT). Reber reported adult-like performance in AGL by 4-, 8-, and 14-year olds (Roter, 1985, cited in Reber, 1993) as well as improvements in sequence vs. random trials in an SRT-like task by 3-5 year olds. Meulemans (1998) cites another unpublished study, this one by Lewicki's group (Czyzewska, Hill, & Lewicki, 1991, cited in Lewicki et al., 1992) which indicates that preschool children (4- to 5-year-olds) could implicitly acquire knowledge about a rule-based visual pattern in a search task. Below I summarize and discuss the body of published studies on implicit learning in children using AGL, SRT, and PCT.

AGL studies with children. Two studies have examined whether children can successfully perform AGL, including using modifications to the paradigm to make it more applicable to children[16]. The first study (Fischer, 1997) tested children 9-11 years old in a conventional AGL task. The results suggested that children in this age range are able to complete the task by abstracting some version of the underlying rules rather than relying on surface features such as chunk strength[17]. More recently, Witt and Vinter (Witt & Vinter, 2012) substituted patterns of colored flags for the letters originally used in AGL and used a generation task in the test phase (rather than the conventional forced-choice test). The subjects in this study were children 5-8 years old (younger than the 9-11 year old children in the Fischer (1997) study). While the children overall were able to perform the task, younger children seemed more attuned to surface features of the stimuli rather than underlying rules, although the eight-year-old (older) group showed some evidence of using rules. Neurotypical adults generally reach accuracy levels in the range of 55-65%, close to chance (50%), but statistically significantly different. The children in

[16] Because depending on their age, children may have less letter knowledge than adults or may find a letter-based AGL difficult or boring.
[17] As discussed in Chapter 1, chunk strength can be used as a measure of surface similarity in AGL tasks.

these studies reached similar levels of accuracy in the test phase. Together, these two studies suggest that while children are able to perform the task at adult-like levels, nevertheless, a transition from dependence on surface features to use of abstract rules may take place with maturation.

Because of theories suggesting that implicit learning deficits may underlie dyslexia (Fawcett & Nicolson, 1990; Nicolson, Fawcett, Brookes, & Needle, 2010), AGL studies have been conducted comparing children with learning disabilities to typically developing (TD) children. Pavlidou and colleagues (Pavlidou, Kelly, & Williams, 2010; Pavlidou, Williams, & Kelly, 2009) and compared dyslexic and typically developing[18] children (mean age 10 years old). Consistent with the Fischer study, the results show that typically developing 10 year olds are able to perform the task and respond based on abstract structure (grammaticality) more than surface features (chunk strength). Reber and colleagues compared AGL in adults and children with Williams Syndrome and typically developing/ed children and adults; they reported no differences in AGL performance between adults and children (lowest age 9 years old) and no main effects of chunk strength (Don, Schellenberg, Reber, DiGirolamo, & Wang, 2003). A study comparing good and poor spellers (Ise, Arnoldi, Bartling, & Schulte-Koerne, 2012) (ages 9.4-9.8 years) found that both groups were able to perform the AGL task, but reported a main effect of chunk strength (but no main effect of grammaticality). Finally a study comparing TD and ADHD children (Rosas et al., 2010) reported that the ADHD children performed better than TD children on the task, but this study did not consider the effect of chunk strength.

[18] Presumed-to-be typically developing by ruling out e.g. dyslexia

The consensus these studies might point to is that while children as young as 5 years old can perform AGL to adult-like levels of accuracy, surface features (chunk strength) dominate the judgments made by children younger than 9 or 10 years old. This finding does not necessarily imply that younger children are performing in a way completely unlike adults, since adult amnesics and normals use both abstract (grammaticality) and surface feature (chunk strength) information to perform the AGL task (Knowlton & Squire, 1996). Rather than younger children performing more like adults, it may be that the older children (9 years old and older) are less like adults in their insensitivity to chunk strength, suggesting a complex, possibly parabolic trajectory. However, no studies have tested the directly compared AGL in typically developing children and adults, so this proposition cannot be confirmed by the existing literature.

SRT studies with children. Several studies have directly compared SRT performance in children and adults, but have produced conflicting results. Implicit learning in SRT tasks is usually measured by comparing the difference in reaction time between sequence trials and random trials; if sequence-specific learning has taken place, the reaction time will be shorter (on average) for pattern trials. Sequence-specific learning indicates that a specific pattern has been abstracted and learned, whereas general decreases in RT across all trial types only indicates increasing familiarity with the task and fluency with the motor components required by the task. Meulemans and colleagues reported no difference in the ratio of random-to-sequence reaction times SRT performance between children (6 years old and 10 years old) and adults (Meulemans, der Linden, & Perruchet, 1998). Similarly, De Guise and Lassonde tested children and adolescents in four different age groups (6-8 years, 9-11 years, 12-14 years, 15-16 years) and found no difference in implicit sequence learning across groups. At first, Thomas and colleagues (Thomas & Nelson, 2001) found a similar lack of developmental differences in sequence-specific learning across 3 age groups (4 year olds, 7 year olds, and 10 year olds). However, a second study comparing adults and children (7 to 11 years old) revealed that the adults showed greater levels sequence learning and learned the sequence sooner than the children (Thomas et al., 2004). In another study comparing performance on an SRT-like task between pre-adolescents (10- 13 years old) and young adults (18-29 years old), both groups showed similar sequence-specific reductions in RT (Ruitenberg, Abrahamse, & Verwey, 2013).

One study (Lum, Kidd, Davis, & Conti-Ramsden, 2010) tested the same children at 5 years old and then again a year later. Although the children (and an age-matched control group) were faster at Time 2, nevertheless the degree of sequence learning (change in RT between sequence and random blocks) was not different[19]. Finally, a study by Janacsek and colleagues (Janacsek, Fiser, & Nemeth, 2012) compared performance in a sequence learning task across a wide range of ages (4 years old to 68 years old) with a substantial number of participants in each age group (N~30 to 60). The highest sequence learning scores[20] were found in the four children's age groups, with a sharp decline between the 11-12 year old and 14-17 year old age groups, followed by a gentler decreasing trend into old age.

---

[19] An age-matched control roup was included at Time 2 to control for practice effects. The difference between roups at Time 2 was not statistically significant.

[20] Janacsek et al.(2012) used the Alternating Serial Response Task (ASRT) introduced by Howard and Howard (1997, 2004). In this task, the sequence items alternate with random items, resulting in high frequency triplets (HFT) and low frequency triplets (LFT). The measure of sequence learning is calculated from the difference in RT to HFT vs. LFT (in contrast to the difference used in traditional SRT: random-sequence blocks).

Other implicit learning tasks with children. Incidental covariation tasks (Lewicki, Hill & Czyzewska, 1992) have also been used to test developmental invariance of implicit learning in children. One study (Fletcher, Maybery, & Bennett, 2000) compared performance in an incidental covariation task between two age groups (5-8 year old children and 10-12 year old children) and found that the older children demonstrated greater levels of implicit learning. This suggests that as measured by incidental covariation tasks, implicit learning may not be developmentally invariant.

Finally, probabilistic classification tasks (PCT) (Knowlton et al., 1994) and dynamic control/process control tasks (Berry & Broadbent, 1984) have also been used as implicit learning tasks in experimental studies. PCT has been used recently and with a wide variety of participant types (Horan et al., 2008; Knowlton, Squire, et al., 1996; Meeter, Radicsa, Myers, Gluck, & Hopkins, 2008; Speekenbrink, Lagnado, Wilkinson, Jahanshahi, & Shanks, 2010), but dynamic control tasks are no longer appear to be widely used. However, in our search of the existing body of research, we found no studies using either probabilistic classification or dynamic control tasks/process-control tasks with children.

From the AGL studies with children, it appears that young children, like adults, may be swayed by simple surface features, while children older than 9 years seem more sensitive to the abstract grammar or rules consistent with the grammar[21]. The message from SRT studies is less clear: In some studies, children of different ages demonstrate as much sequence learning as others, but in other studies, adults and older children demonstrate more sequence learning. The most age-comprehensive study paints a more

---

[21] Many AGL researchers, including Reber (1993), allow that participants who successfully discriminate between grammatical and non-grammatical test items may be doing so on the basis of rules that are consistent with but not identical to those governing the original Markov chain.

complex picture, suggesting that younger children show more sequence learning than adolescents or adults; this places it at odds with the other studies rather than accommodating their results.

Because of the conflicting results from studies using different age groups and different tasks, a first step towards clearly establishing whether and how implicit learning develops would be a study using multiple tasks in both adults and children. The current study attempts to meet this need by comparing performance across a battery implicit learning tasks between 10 year old children and healthy adults.

Methods

Participants. Participants were 32 children (mean age 10 years, 5 months; 16 male, 16 female) and 29 adults (mean age 23.17 years; 13 male; 16 female) recruited in accordance with MIT Committee on the Use of Human Experimental Subjects (COUHES) regulations. Following both Harvard CUHS and MIT COUHES policies, the Harvard CUHS was notified of my involvement in the project and sent a letter to the MIT COUHES signaling that they had been notified[22]. Participants received Amazon gift cards for $20 for each hour of testing; for most participants this was three hours, and therefore sixty dollars. In addition, child participants were given stickers and at the conclusion of testing were allowed to choose a toy from a box of small, inexpensive toys. All adult participants gave written consent. Parents of child participants gave written consent and the child participants gave written assent after having the experiment explained to them and being given the opportunity to ask questions.

[22] MIT COUHES protocols including a researcher from another institution cannot be approved unless that researcher's home institution sends such a letter to the MIT COUHES. For this study, this requirement was satisfied.

Tasks and materials. Each participant performed a battery of experimental and standardized tasks as detailed below.

Implicit learning tasks. Several implicit learning tasks which feature prominently in the body of research on implicit learning were used as described below.

Artificial grammar learning. The AGL task used the stimuli and procedure described by Knowlton and Squire (1996), and also found in Knowlton, Mangels and Squire (1992) and Abrams and Reber (1989) (Appendix I).

The AGL task consisted of a training phase and a testing phase. In the training phase, participants viewed a series of letter strings on a computer screen. Participants were instructed to briefly view the string on the screen, then to write the string on a piece of paper and cover it before viewing the next string. Each of 23 training stimuli were presented twice, for a total of 46 training trials.

Between the training and testing phase, participants were told "The letter strings you just wrote down were actually created by a set of rules that specified things like what order the letters could go in." Subsequently, in the testing phase participants viewed a series of previously unseen (novel) letter strings on a computer screen again, and for each letter string they were instructed to make a forced choice decision: "Press 1 if you think this letter string came from the same set of rules as the words you wrote down. Press 0 if you think it did not."

The training phase letter strings were all legal or "grammatical" letter strings created by a Markov chain. The testing phase letter strings included 16 grammatical and 16 non-grammatical letter strings (based on the same Markov chain as the training stimuli); grammatical and nongrammatical strings were matched for length and chunk strength[23]

From the responses in the testing phase (forced-choice), we calculated the following three measures for each participant: First, percentage of hits (correct endorsement of grammatical items). Second, percentage of high chunk strength items endorsed (half the high chunk strength items were grammatical; half were not); this provides some insight into whether the participants were responding to test items based more on abstract or concrete implicit learning (see Knowlton & Squire, 1996)). Third, discriminability (d', d prime) score for grammatical items: z-score of the percentage false positives subtracted from the z-score of the percentage of hits; d' provides a measure of sensitivity since the scores are adjusted for response bias (MacMillan & Creelman, 1991).

Probabilistic classification task. In this task, participants saw on each trial some combination of four possible cue stimuli on a computer screen and were asked to make a forced-choice decision. After each decision, participants received feedback indicating whether their decision was correct or incorrect. Each cue stimulus appeared in one of four possible cue locations and could be either present or absent on each trial, so in total 15 possible cue combinations (with at least one cue present) were possible. Unknown to the participants, each cue combination was associated with each decision outcome some percentage of the time between 0 and 100%, so the trial-by-trial feedback was not entirely consistent for each cue-decision combination.

---

[23] Several researchers have suggested that rather than learning an entirely abstract set of rules, participants may be performing the forced-choice decision in AGL tasks based on frequently co-occuring letters, or chunks (bigrams and trigrams). For example, if the letter combination "XYZ" (trigram) occurred frequently in the training stimuli, then a testing letter string containing "XYZ" would likely have a high chunk strength. Each training bigram's or trigram's "associative strength" is the number of times it appears in the training stimuli. Each testing letter string's the chunk strength is the mean of the associative strengths of the chunks in that string. (see Knowlton & Squire, 1996; Meulemans et al. 1997; Pothos 2007).

Since the cue combinations ("patterns") were probabilistically associated with each outcome, to successfully perform the task the participants had to learn which outcome was *more probable* given a particular cue combination. This outcome is referred to as the "optimal response" for that cue combination. While the trial-by-trial feedback to the participant varied probabilistically, nevertheless the participants demonstrated learning by choosing the optimal response more often than the alternative. All measures of PCT performance used in this study were based on the participants' optimal responses rather than their matches to trial-by-trial feedback.
The order of stimulus presentations was random but constant across participants.

Serial response task. Participants viewed a computer screen divided into four equal quadrants and were directed to a set of four buttons on a Cedrus response box and told that each button corresponded to one of the target positions (quadrants)[24]. In each trial a target stimulus ("Ozzie the Octopus") appeared in one of the quadrants. Participants were instructed to press the corresponding key when this happened. Each participant completed a practice block of 8 trials before beginning the 9 experimental blocks.

Of the experimental blocks, 5 were "random" blocks (blocks 1, 3 , 5, 7, and 9) and 4 were "sequence" blocks (blocks 2, 4, 6, and 8). Sequence blocks consisted of a repeating 10-item second-order conditional sequence. Random blocks consisted of 60 trials in which the target quadrant order was randomized. Sequence blocks were 160 trials long (16 repetitions of the 10-item sequence). Several sequences were used, and were counterbalanced across participants and conditions.

---

[24] While many SRT tasks use a linear set of positions, some studies with children use quadrant-based target positions (Lum et al., 2010; K M Thomas & Nelson, 2001).

In addition to the traditional implicit SRT, an explicit SRT task was embedded in the stimuli. In two of the sequence blocks, participants were cued to attend to the order of target quadrants and told that the target (octopus) moved in a pattern. The sequence used in the implicit blocks was different from the sequence used in the explicit blocks. Two different colored targets (blue octopus and purple octopus) were used to differentiate the explicit sequence blocks from the implicit sequence blocks. Octopus color, which sequence was explicit, and whether implicit or explicit sequence blocks were presented first were all counterbalanced across participants.

Response time and accuracy for each trial were recorded. No feedback was provided and the sequence advanced regardless of whether the participant responded correctly or incorrectly. However, mean accuracy was at no lower than 77% for all but one child participant (62%), and no lower than 92% for adult participants.

In SRT tasks the measure of interest is usually derived from the difference in reaction time between random and sequence trials. Often, these metrics compare the reaction time in sequence blocks to reaction time in the random blocks both preceding and following ("flanking") the sequence block in question. In this case, first the median[25] reaction times for the random blocks flanking each sequence block (e.g. random blocks 1 and 3 flank sequence block 2) are averaged. The difference between this average and the median RT of the sequence is then divided by the flanking random block average. The resulting percentage is referred to here as the Skill Score (i.e. difference between RT in random vs. sequence trials expressed as a percent of random block RT). Four different

[25] RT medians are often preferable to means in SRT tasks because of high intra- and inter-individual variability, especially in early trials.

skill scores were obtained this way, one for each sequence block (implicit sequence 1, implicit sequence 2; explicit 1, explicit 2).

Rotary pursuit. Participants were asked to use a stylus to maintain contact with a rotating target on a flat surface. Participants first completed a practice trial to familiarize them with the task and to establish baseline speed (15, 30, 45, or 60 RPM). The speed at which the participant's time on target was closest to 5 seconds (or 25% of 20 seconds) was selected and used for all subsequent trials. Participants' time on and off target within each trial were recorded; four trials were performed, then a break of 1 minutes, then four more trials. The shape in which the target rotated was a modified rectangle (truncated corners). The measure of interest for this task is Time on Target (seconds or % of 20 seconds that the participant's stylus is in contact with the target per trial).

Mirror tracing. Participants traced the outline of a six-sided star while watching their hands in a mirror rather than watching their hands directly. They were instructed to stay inside the outline of the star while tracing, to look in the mirror rather than at their hands, and to trace as quickly as they could without sacrificing accuracy. Time to complete an outline and number of errors (stylus moves outside outline of star) per trial were recorded. Each participant completed five trials of mirror tracing. The measures of interest for this task are Time to Completion (the time it takes for the participant to complete one full tracing) and errors per trial (the number of times the participant's stylus goes outside the target area during one tracing).

Rotary pursuit and mirror tracing were included as measures of motor skill learning in the current study to examine the relationship between motor skill procedural learning and other types of implicit learning. In our study, each of these tasks was administered via a specialized apparatus built by the Lafayette Instruments company.

Comparison and control tasks. Kaufman Brief Intelligence Test, Second Edition (KBIT-2). The KBIT-2 provides measures of verbal, non-verbal, and composite IQ (Kaufman & Kaufman, 2004). The test itself consists of a multiple-choice picture vocabulary test, a matrix completion test, and an open-ended verbal riddles test. Visual stimuli for the picture and matrix tests are included in a test booklet provided by the publisher of the test.

California Verbal Learning Test (CVLT—II). The CVLT—II provides measures of declarative memory that can be expressed as raw, standardized, or scaled scores. In this assessment, participants are read a list of 16 words and must repeat as many as possible after the entire list is read; in addition to five trials of immediate recall, short delay and long delay trials are provided. However, in this study only the immediate (no delay) free-recall trials were used; our primary measure of interest was the scaled T score based on the total words recalled in the five immediate free-recall trials (maximum possible = 80). (T-scaled scores are adjusted for age.) The Standard Version of the CVLT was used for adults, and the child version (C-CVLT) for children; these differ only in the content of the word lists (they are identical in format, procedure, etc.). Both versions were created and published by Delis et al. (2000)

Processing speed. To evaluate participants' processing speed, we used the Coding and Symbol Search subtests from the fourth edition of the Wechsler Intelligence Scale for Children (Wechsler, 2003). On the Coding subtest, participants are asked to translate digits into symbols by referring to a corresponding digit-symbol key (nine novel symbols corresponded to the digits 1 through 9). On the Symbol Search subtest, participants indicate whether either of two symbols on the left side of a page matches any of five symbols on the right side of a page. Participants had 2 min to complete each task.

Working memory. Working memory was assessed with a count span task (Case, Kurland, & Goldberg, 1982; Cowan et al., 2005), in which participants viewed an array with blue circles, blue triangles, and red circles and were instructed to count only the blue circles (targets) within 4.5 s. After one or more arrays were presented, participants were prompted to write separately the number of targets presented in each display. Load ranged from 1 to 6 consecutive arrays and increased by one after three instances of a particular load.

Results

Table 1 displays demographic data for all participants.

One concern was how well the sample represented a wider population, since this would of course affect the generalizability of the findings. With regard to IQ, both groups (10-year olds and adults) were above the mean for the general population (100) for verbal, non-verbal, and composite IQ, and these differences were statistically significant (adults non- verbal $\bar{x}$ = 114.72, t = 5.64, p <0.001 ; adults verbal $\bar{x}$ = 117.97, t = 5.02, p <0.001; adults composite $\bar{x}$ = 119.00, t = 5.96, p<0.001; 10-year olds non-verbal $\bar{x}$ = 114.27, t = 17.01, p<0.001; 10-year olds verbal $\bar{x}$ = 115.26, t = 6.51, p <0.001; 10-year olds composite $\bar{x}$ = 116.73, t = 9.30, p <0.001). These differences probably resulted from self-selecting sampling methods which lead to greater inclusion of high-SES and high parental education individuals. Given these differences from general population norms, interpretation of results from this sample should be tempered by an understanding of the sample's slight atypicality. However, it should also be noted that the group means are within one standard deviation of the general population mean, so although the sample displays higher IQ (on average) than the general population, the magnitude of the difference is noteworthy but not necessarily overwhelming.

Similar differences from general population means were found in the total CVLT for the adult group, but not for the 10-year old group (Table 2a). No significant differences were found in standardized IQ between the groups (Table 2a). Since standardized IQ scores are normed by age, this result demonstrates that for their respective ages, the two groups were similar in intelligence. However, Table 2b displays unstandardized verbal and non-verbal KBIT scores for both groups. Similarly, we found that children had smaller WM spans than adults (t = 5.56, p<0.001) and slower processing speeds (t = 12.91, p<0.001) in keeping with previous studies.

Adults and children have also been shown to differ on measures of explicit learning. We used the California Verbal Learning Test to compare explicit learning between adults and children and found that adults scored higher on both recall and recognition measures when raw scores were compared (recall t = 3.16, p = 0.003; recognition t = 8.93, p<0.001), but that there were no significant differences when age-normed scores were compared (Table 2a).

Implicit learning tasks: overall. In the SRT, adult participants had higher skill scores for both implicit and explicit sequence learning (implicit t = -2.33, p=0.024; explicit t = -2.15 p=0.036). In the PCT, we compared optimal response by block and found no significant main effect of group and no evidence of a group-by-block interaction; however, there was a main effect of block such that optimal responses were more frequent in successive blocks ($F(49, 41) = 9.47$, $p<0.001$). In the AGL task we found no significant difference between adults' and children's discrimination (*d'*, d prime) of grammatical items ($t = 0.97$, $p = 0.83$) (Table 3).

In mirror tracing, there was a significant main effect of trial number (decrease in completion time over trials) ($F(17,38) = 8.54$, $p < .001$). There was also a significant main effect of group ($F(17, 38) = 29.16$, $p < .001$); adults' completion times were shorter than 10-year-olds' completion times. There was no significant interaction between group and trial number ($F(17,38) = .95$, $p >. 05$); adults' improvement over time was parallel to 10-year-olds'. A similar pattern was found for errors: a main effect of trial number ($F(17,38) = 9.03$, $p < .001$); a main effect of group ($F(17,38) = 23.83$, $p < .001$); but no interaction between trial number and group ($F(17,38) = 0.89$, $p > .05$).

In rotary pursuit, the measure of interest is time on target (TOT), i.e. the amount of time that the subject is able to maintain contact between the stylus and the rotating target. We found improvement over time (main effect of trial number) ($F(31,28) = 5.06$, $p < .001$), a main effect of group ($F(31,28) = 46.85$, $p < .001$), but no significant interaction($F(31,28) = 0.16$, $p > .05$).

For both mirror tracing and rotary pursuit measures, children were more variable (across individuals—as seen in higher standard deviations and more platykurtic histograms) than adults.

Discussion

The present study investigated whether implicit learning is developmentally invariant by comparing performance on traditional laboratory tasks of implicit learning by 10-year-old children and healthy young adults. We did not find a main effect of age group in any of the tasks used. However, item-level analyses of two tasks revealed that children's performance depended on different stimulus features than adults' performance did. These results support the position that the nature of implicit skill and habit learning (with regard to underlying statistical principles) changes from childhood to adulthood.

This study was cross-sectional rather than longitudinal, so it is possible that observed differences between the age groups were due to factors other than age alone. However, the groups were not different on measures of verbal or non-verbal IQ and were similar in demographic makeup. It is also possible that the lack of significant differences in overall performance between the two groups is due to low power (modest sample size). Finally, these results cannot necessarily be taken as evidence that the trajectory of implicit learning development is linear or monotonic since we only used one age group of children; a wider sampling of ages could reveal a non-linear (such as quadratic, "inverted-u") relationship.

The results of the present study suggest that implicit learning changes qualitatively between childhood and adulthood. These results are consistent with Janascek and colleagues' finding that "due to developmental changes in early adolescence, implicit skill learning processes undergo a marked shift in weighting raw probabilities vs. more complex interpretations of events" (Janacsek et al., 2012, p.496). Similar findings have been reported in other areas of inductive learning. For example, Markman noted that children initially use non-hierarchical ("flat") category structures and only over time develop the ability to use more complex category structures (Markman, 1991). Likewise, Kam and Newport (2005) reported that children were more likely to regularize inconsistent input, and Sloutsky and Fisher (2004, 2005) demonstrated that compared to adults, children show memory for specific exemplars rather than category-based memory. What these findings share with the current study is that they demonstrate children using less complex statistical learning algorithms than adults. However, with the exception of the Janascek et al. study, these examples did not investigate whether the learning that took place was accessible to consciousness or not (i.e. whether it met this criterion for implicit learning or not).

Thus, one direction for future research to bridge findings from implicit learning and statistical learning traditions would be for statistical learning studies to include a participant debrief and/or post-test to probe the extent to which participants have conscious access to what was learned[26]. Several authors (including Amso & Davidow, 2012; Perruchet & Pacton, 2006) have pointed out similarities between implicit learning research and statistical learning research. Conversely, implicit learning studies can be use tasks designed with more specific statistical properties than the tasks most commonly used now, or can be designed to exploit the specific statistical properties of the existing tasks. In particular, developmental studies of implicit learning may require this type of specificity. The discrepancies in previous studies (e.g. Thomas & Nelson (2001) compared to Thomas et al., (2004)) may be due to a lack of sensitivity (or differing sensitivity across tasks) in the tasks or measures used; although they gauged some overall implicit learning ability, they may not have been able to distinguish among different types of underlying algorithms or computations used to complete the tasks.

Both statistical learning studies and implicit learning studies may benefit from the use of simultaneous measures of voluntary (such as SRT, PCT, and AGL) and

[26] Such an approach would address the issue of information encapsulation raised by Fodor (1983). See Chapter 6 for a more detailed discussion.

involuntary (e.g. response time, pupillometry, eye-tracking, and ERP) responses. As Lloyd, Newcombe, and Keen (2003) pointed out, the apparent paradox of infants' superior inductive abilities (compared with young children) may be due in part to the fact that infant studies use looking time (an involuntary response) whereas studies with young children often require an explicit response.

A recent study by Amso and Davidow (2012) demonstrates how these strategies can be used in developmental implicit learning studies. In this study, the researchers placed two types of statistics in conflict: raw item frequencies (unconditional probabilities) and item relation probabilities (conditional probabilities). The participants were infants, children, and adults, and the measures included both saccade latencies as well as reaction time for an explicit response.

Another direction for future research is a systematic mapping of the development of inductive abilities (based on complexity/type of algorithm/computation). Currently, domain-specific studies in category learning, language learning, and causal learning and other domains rely largely on inconsistent age group contrasts. Strategic use of consistent multiple age groups and multiple domain testing could lead to a timeline of developmental milestones in inductive learning.

Furthermore, future research can address the extent to which experiences (such as instruction, training) can accelerate development along this timeline—can the requisite algorithms be taught? To the extent that such experience-based learning is limited, future research must investigate what aspects of neural development correspond to changes in inductive algorithms. For example, could cortical thinning or structural connectivity of the basal ganglia play a role?

To the extent that research on inductive processes can be applied to instructional strategies, we have to know and acknowledge that children's inductive processes may differ in important ways from adults', and that therefore findings from adults may not generalize to children. However, instructional strategies for adult or late-adolescent education may be able to benefit sooner from basic research on inductive processes. Future research detailing the developmental trajectory of inductive processes may lead to powerful tools for teaching.

Chapter Five: Individual Differences in Implicit Learning

This study addresses the questions of individual differences in implicit learning, the relationship between implicit learning and intelligence, and the extent to which implicit learning tasks are measuring a common construct using classical test theory, principal components analysis, and factor analysis. Several tasks were used to measure implicit learning: Artificial Grammar Learning (AGL), Serial Response Task/Serial Reaction Time Task (SRT/SRTT), Probabilistic Classification Task (PCT), implicit category learning, rotary pursuit, and mirror learning. Participants were healthy young adults from the Harvard University Psychology Study Pool (n=70). Moderate (r~0.3-0.6) correlations were found for test-retest reliability of separate tasks. Item analysis yielded reliability as high as 0.8 when tasks were combined. Intelligence and working memory were not correlated with implicit learning performance, and a unidimensional model could not accommodate both implicit learning and intelligence measures. However, a unidimensional factor model with loadings for SRT, PCT, and category learning could not be rejected. AGL, rotary pursuit, and mirror tracing did not fit well with the other tasks as shown by pairwise correlation, item analysis, PCA, and FA.

Much educational research focuses on individual differences in various aspects of learning, presumably with the aim of optimizing instruction based on individualization. However, if a learning ability was found that did not vary across individuals and was independent of IQ, potentially it could be exploited—for example, as a compensatory mechanism for students who have difficulty learning in more traditional ways. Implicit learning has been proposed as just such a universal learning mechanism. Furthermore, as I detailed in Chapter Three, some types of learning and instruction may be particularly suited to implicit learning paradigms. However, a basic question about the nature of implicit learning has not yet been answered: whether and to what extent individuals vary in their ability to learn in implicit learning paradigms, and whether these differences can be reliably measured with existing implicit learning tasks. Theoretical considerations may have contributed to the lack of research in this area.

Across various dual-system theories, the "implicit" system is often characterized as "universal" (as opposed to heritable), independent of general intelligence, and independent of working memory (Evans, 2008). This is one reason why implicit learning ability is assumed to be consistent across individuals and independent of IQ. More specifically, Arthur S. Reber, a prominent researcher on implicit learning, hypothesized several properties of implicit systems (including systems for implicit learning) that should differentiate them from explicit systems or forms of thinking (Reber, 1993). Among these hypothesized characteristics are: age independence, robustness to disease or injury, low variability across individuals (compared to variability across individuals on measures of explicit learning or reasoning), and IQ independence. The claim of robustness or resistance to damage in the face of injury or disease has been amply demonstrated (see Gabrieli, 1998, for a review). We have examined the relationship between age and implicit learning ability extensively elsewhere (Finn, Kalra, & Gabrieli, in preparation). However, the questions of inter-individual variability and IQ-independence have not been resolved by research to date. The current study addresses both these questions, as well as investigating cross-task reliability of implicit learning ability. (A fifth and final proposed property, regarding the conservation of implicit learning across phylogeny, will not be discussed here for reasons of relevance, but see e.g. Herbranson & Shimp (2008) if interested.)

Background

Currently, there is no standardized measure of implicit learning ability. Implicit learning is measured by scores on laboratory tasks, such as the serial response task (SRT), probabilistic classification task (PCT), and artificial grammar learning (AGL), which are conventionally used to look for differences across groups, not individuals. Furthermore, there are compelling theoretical reasons why individual differences in implicit learning ability may be unstable or of negligible magnitude when compared with random variation (i.e., low signal-to-noise ratio).

As Bucher and colleagues have pointed out (Buchner & Wippich, 2000; Buchner & Brandt, 2003), if it is the case that implicit learning tasks are less reliable than explicit memory tasks, then the apparent dissociations in performance (e.g. within amnesics between implicit and explicit tasks) may in fact reflect these differences in the reliability of the measures rather than differences in the underlying processes or abilities. In a series of experiments, this group has demonstrated how low reliability may be contributing to reported dissociations between explicit and *implicit memory*[27]. However, no such allegations have yet been brought against measures of *implicit learning*. Importantly, one of the principal reasons that Bucher and colleagues propose for the lack of reliability in implicit memory studies is the use of open-ended prompts (as opposed to forced-choice

---

[27] Buchner & Wippich (1997) distinguish between implicit learning and implicit memory as follows: implicit learning refers to situations in which the acquisition of knowledge is incidental (not intentional) but the retrieval may be incidental OR intentional. In contrast, implicit memory refers to the phenomenon of incidental or automatic retrieval; implicit memory is studied almost entirely via priming experiments. In the Multiple Memory Systems model, both implicit memory and implicit learning are considered types of nondeclarative memory because they both appear to be preserved in amnesic.

prompts) in the test phase. While the test phase of implicit learning tasks sometimes includes open-ended prompts, forced-choice prompts are almost always used. Thus this possible cause of low reliability in implicit memory may not be applicable to implicit learning tasks.

Even if inter-individual variation is observed in implicit learning task performance, it could be argued that this observed variation is due to error rather than variation in the underlying construct (implicit learning ability). From the perspective of classical test theory (CTT), the question is one of reliability: Can we reliably measure differences in implicit learning ability across individuals across time? What proportion of the observed variance is due to variance in implicit learning ("true score" variance)?

While we cannot directly measure the true score or its variance, we can estimate the ratio of true score variance to observed score variance. In CTT, the observed score can be expressed in terms of the true score and error (where T = true score, X = observed score, and e = error)[28]:

$$\text{(Eq. 1) } X = T + E$$

And the observed variance can likewise be expressed:

$$\text{(Eq. 2) } \sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

In CTT reliability can be defined as the correlation between two sets of observed scores from a replication of a measurement procedure ($\rho_{xx'}$), and equivalently as the proportion of observed score variance that is due to true score variance ($\frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2+\sigma_E^2}$) (Haertel, 2006).

Reliability across time. In practice, an exact replication of the same measurement is difficult because of possible carryover effects, i.e. individuals' responses

[28] CTT assumes that error is random and uncorrelated with true score.

during the second measurement may be influenced by their knowledge of the questions from their first exposure. For this reason, parallel forms are often used that consist of similar but non-identical items all drawn from a "population" or bank of items. When parallel forms are not available, researchers often attempt to estimate reliability through inter-item correlations (see Cronbach, 1951).

Thus the ideal approach to empirically addressing the question of whether implicit learning ability differs across individuals (and whether those differences can be measured in a replicable fashion with existing tasks) would be a test-retest study of implicit learning measurements with parallel forms. The current experiment attempts to answer the question in exactly this way.

Several thorough searches[29] of the published literature over the last year have failed to yield any publications that have used test-retest or parallel forms to investigate reliability of implicit learning tasks. Standard internal-consistency approaches to estimating reliability would not be appropriate for some types of implicit learning tasks, especially those in which the probability of a correct response is expected to increase with subsequent trials (i.e. PCT, SRT, forced-choice categorization with feedback). That is, in PCT and SRT, individual trials are not items; the measures of learning for these tasks are derived from differences (in accuracy or reaction time) within the training session. Hypothetically, internal consistency reliability could be estimated for the items in an artificial grammar learning task, but again no published studies can be found that attempt to do so.

---

[29] Using Web of Science/Citation Index, EBSCO, and every combination of the search terms "reliability," "test-retest," "parallel forms" and each of the following terms: "implicit learning," "artificial grammar," "probabilistic classification," "serial response." Searches conducted periodically between March 2013 and January 2014.

Reliability across tasks. If there are stable individual differences in implicit learning ability, then they should be stable across tasks. The results of studies using one type of implicit learning task are often assumed to generalize. However, there is little direct empirical evidence for agreement across tasks. Furthermore, some neuropsychological studies suggest some dissociations among the experimental tasks commonly used to study implicit learning. Patients with HD and PD demonstrate impaired performance in habit learning tasks such as probabilistic category learning (Knowlton, Mangels, et al., 1996), implicit visual category learning (Ashby et al., 2003; Maddox et al., 2005; fsPrice et al., 2009), and artificial grammar learning (Smith, Siegert, & McDowall, 2001) but demonstrate some preservation of function in skill learning tasks such as rotary pursuit, mirror tracing, and serial response task (Gabrieli, Stebbins, Singh, Willingham, & Goetz, 1997; Smith et al., 2001). The possibility that skill learning and habit learning have different neural correlates suggests that a) an individual's habit learning and skill learning abilities may differ and b) the reliability of habit learning and skill learning measures may differ from each other.

Moreover, behavioral studies using multiple measures of implicit learning have reported low correlations across tasks. Gebauer et al.(2007) conducted a study using a battery of tasks, including SRT, AGL and a process control task (a la Berry & Broadbent, 1986); Horan et al. (2008) compared implicit learning in schizophrenics and controls using PCT and AGL. Table 8 displays the correlations across tasks reported by these researchers (for neurotypicals only). In all cases the correlations observed were very low and non-significant.

In the current study I will revisit this issue in several ways. First, following the example of previous studies, I will examine pairwise correlations between scores on implicit learning tasks. Next, moving from pairwise to "group-wise" correlations, I will treat the scores from each task as an "item" and use item analysis to examine overall interitem correlation (as well as identifying items that do not fit well with the others). Based on the results of the item analysis, I will perform a principal components analysis to investigate the dimensionality of the data. Finally, I will test whether a unidimensional model can accommodate the implicit learning tasks using factor analysis.

Implicit learning and IQ. What about the hypothesized lack of relationship between intelligence and implicit learning? Knowing whether implicit learning ability is correlated with intelligence (either positively or negatively) would be theoretically useful to instructors and others attempting to incorporate implicit learning into education. If implicit learning ability is preserved at normal levels in students with low IQs, or if in fact a negative relationship exists between implicit learning and intelligence, then perhaps instructional strategies that take advantage of implicit learning could act as a compensatory mechanism; if, on the other hand, intelligence and implicit learning are positively related, then this strategy would be unsuccessful.

There is evidence that implicit learning may be intact in conditions in which IQ is usually depressed relative to neurotypicals, including Down's syndrome, mental retardation, and autism spectrum disorders (Atwell, Connors & Merrill, 2003; Brown, Aczel, Jimenez & Kaufman, 2010; Vicari, Verruci, & Carlesimo, 2007; Vinter & Detable, 2003, 2004, 2008; Bussy et al., 2011). These findings suggest that perhaps implicit learning is indeed IQ-independent, as predicted by Reber.

Furthermore, studies of neurotypical children and adults have suggested that there may be only a weak relationship between IQ and implicit learning, if any. In Table 7, I summarize the correlations between IQ and performance on implicit learning tasks reported in previous studies. Notably, the participants in many of these studies are children or adolescents; only Reber Walkenfeld & Hernstadt's (1991) study used adults exclusively. Several of the correlations from previous studies are low but statistically significant (at $p<0.05$): SRT and verbal (r = 0.22; Kaufman et al. 2010); AGL-CFT[30] (r = 0.10) and process control-CFT (r = 0.10; Gebauer et al., 2007). One of the highest correlations reported is between AGL performance and WAIS-R IQ (r = 0.25; Reber Walkenfeld & Hernstadt, 1991), but does not clear the p<0.05 threshold. The authors freely concede that this may be due to the low sample size (N=20) and that the correlation would likely be found statistically significant with a larger sample size. Here, as elsewhere (Reber, 1993), Reber clarifies that his claim about IQ and implicit learning is not that there is no relationship at all, but that the relationship will be small compared to the relationship seen between explicit tasks and IQ (he offers the same caveat regarding individual differences in implicit learning measures—that the measureable individual differences will be low compared to those observed for explicit learning).

In the current study we attempt to address all of the following questions about implicit learning: 1) Whether and to what extent reliable individual differences in implicit learning ability/performance exist and can be measured 2) Whether and to what extent individual differences in performance on implicit learning tasks can be explained by intelligence, working memory, or explicit learning 3) Whether and to what extent the

[30] CFT: Cattell, R. B., & Weiß, R. H. (1977) and (1980). Grundintelligenztest Skala 2 (CFT 2 & 3). Gottingen, Germany: Hogrefe Verlag fur Psychologie.

various laboratory tasks used to investigate implicit learning are in fact measuring a common construct.

To address the first question, we conducted a parallel forms test-retest reliability study. To address the second question, we collected intelligence, working memory, and explicit learning data on all the participants. We used the data collected to approach the third question using pair-wise correlation, a cross-task internal consistency reliability study, a principal components analysis, and an exploratory factor analysis.

Methods

Participants. Participants were 70 undergraduates (32 male, 38 female; mean age 17.7 years) recruited from the Harvard Psychology Study Pool in accordance with the Harvard University Committee on the Use of Human Subjects.

Materials. Experimental implicit learning tasks for study (AGL, PCT, SRT, category learning) two had two versions, referred to here as Form A and Form B. Order of forms was counterbalanced across participants (41 participants saw Form A at Time 1 and Form B at Time 2; 34 participants saw Form B first, then Form A).

Psychometric assessments. Several standardized, published psychometric instruments were used as described below.

Kaufman Brief Intelligence Test, second edition (KBIT-2). The KBIT provides measures of verbal, non-verbal, and composite IQ (Kaufman & Kaufman, 2004). The test itself consists of a multiple-choice picture vocabulary test, a matrix completion test, and an open-ended verbal riddles test. Visual stimuli for the picture and matrix tests are included in a test booklet provided by the publisher of the test.

California verbal learning test (CVLT—II). The CVLT—II provides measures of declarative memory which can be expressed as raw, standardized, or scaled scores. In this assessment, participants are read a list of 16 words and must repeat as many as possible after the entire list is read; in addition to five trials of immediate recall, short delay and long delay trials are provided. However, in this study only the immediate (no delay) free-recall trials were used; our primary measure of interest was the scaled T score based on the total words recalled in the five immediate free-recall trials (maximum possible = 80). The Standard Version of the CVLT-II was used as Form A and the Alternate Version as Form B; these differ only in the content of the word lists (they are identical in format, procedure, etc.). Both versions were created and published by Delis et al. (2000).

Experimental Tasks. Several experimental tasks taken from the literature on implicit learning were used as described below.

Artificial grammar learning. The artificial grammar learning task consisted of a training phase and a testing phase. In the training phase, participants viewed a series of letter strings on a computer screen. Participants were instructed to briefly view the string on the screen, then press a button on the keyboard to erase the string from the screen, then to write the string on a piece of paper and cover it before viewing the next string. Each of 23 training stimuli was presented twice, for a total of 46 training trials.

Between the training and testing phase, participants were told "The letter strings you just wrote down were actually created by a set of rules that specified things like what order the letters could go in." Subsequently, in the testing phase participants viewed a series of letter strings on a computer screen again, and for each letter string they were instructed to make a forced choice decision: “Press 1 if you think this letter string came from the same set of rules as the words you wrote down. Press 0 if you think it did not.”

The training phase letter strings were all legal or “grammatical” letter strings created by a Markov chain. The testing phase letter strings were completely novel (i.e. no overlap with training strings) and included 16 grammatical and 16 non-grammatical letter strings (based on the same Markov chain as the training stimuli); grammatical and nongrammatical strings were matched for length and chunk strength

Form A of the AGL task used the stimuli and procedure described by Knowlton and Squire (1996), and also found in Knowlton, Mangels and Squire (1992) and Abrams and Reber (1989) (Appendix I). Form B used strings generated by the Markov chain in Knowlton and Squire (1996) (See Figure1). Regardless of form, the procedure was the same.

From the responses in the testing phase (forced-choice), we calculated the following three measures for each participant: First, percentage of hits (correct endorsement of grammatical items). Second, percentage of high chunk strength items endorsed (half the high chunk strength items were grammatical; half were not); this provides some insight into whether the participants were responding to test items based more on abstract or concrete implicit learning (see Knowlton & Squire, 1996). Third, d’ (d prime) score for grammatical items: z-score of the percentage false positives subtracted from the z-score of the percentage of hits; d’ provides a measure of sensitivity since the scores are adjusted for response bias (MacMillan & Creelman, 1991).

Probabilistic Classification Task. In this task, participants are presented with a set of visual stimuli (consisting of some combination of four possible cue stimuli) on a computer screen and asked to make a forced-choice decision. After each decision, participants receive feedback indicating whether their decision was correct or incorrect. Each cue stimulus appears in one of four possible cue locations and can be either present or absent on each trial, so in total 15 possible cue combinations (with at least one cue present) are possible. Unknown to the participants, each cue combination is associated with each decision outcome some percentage of the time between 0 and 100%, so the trial-by-trial feedback is not entirely consistent for each cue-decision combination.

Form A used frequencies, conditional probabilities, and cue positions from Knowlton, Squire, and Gluck (1994). However, rather than using images of cards with shapes as cues, grey geometric shapes were presented in the cue positions (circle, square, diamond, inverted triangle). Form B used the same frequencies and conditional probabilities, but associated them with different cue combinations; furthermore, in Form B the cues were all hexagonal, but differed in color (blue, green, yellow, orange). (See Appendix 1, Figure 2.) In Form A the possible outcomes were “rain” or “sun,” indicated by clip art images of each, but in Form B the possible outcomes were “Team 1 Wins” or “Team 0 Wins,” indicated by images of a two baseball players with different color uniform stripes (Team 1 Blue, Team 0 Red). (See Appendix 1, Figure 2)

Since the cue combinations (“patterns”) are probabilistically associated with each outcome, to successfully perform the task the participants must learn which outcome is more probable given a particular cue combination. This outcome is referred to as the “optimal response” for that cue combination. While the trial-by-trial feedback to the participant varies probabilistically, nevertheless the participant demonstrates learning by consistently choosing the optimal response. All measures of PCT performance used in this study are based on the participants' optimal responses rather than their matches to trial-by-trial feedback. The order of stimulus presentations was random but constant across participants.

Serial response task. Participants viewed four white circles (1.5 in. diameter) with black outlines on a grey background and were directed to a set of four keys on a QWERTY keyboard ("1234" for Form A and "FGHJ" for Form B) and told that each key corresponded to one of the target positions (circles), preserving the left-to-right relationship. In each trial, one of the white circles "turned" yellow. Participants were instructed to press the corresponding key when this happened. Eight practice trials preceded the 8 experimental blocks. A response-stimulus interval of 250 ms separated each response from the following trial's stimulus.

Each block consisted of 96 trials. Blocks 1,3, 7, and 8 were random blocks; blocks 2,4,5 and 6 were sequence blocks. In sequence blocks, a 12-item second-order conditional sequence was presented 8 consecutive times. In random blocks, eight different 12-item second-order conditional sequences were presented . Each random block consisted of a different set of eight sequences. The two second-order conditional sequences (one for Form A and one for Form B) were taken from Reed and Johnson (1994).

Response time and accuracy for each trial were recorded. No feedback was provided and the sequence advanced regardless of whether the participant responded correctly or incorrectly. However, mean accuracy for each participant and for each blocks was at no lower than 88%, and most participants exceeded 95% accuracy throughout the task.

In SRT tasks the measure of interest is usually derived from the difference in reaction time between random and sequence trials. We computed the RT difference as follows: first, comparing mean reaction time in the last sequence block (Block 6) to the mean reaction time in the immediately following random block (Block 7); the raw difference is expressed in milliseconds. However, to provide context, that difference is then divided by the mean reaction time in the last sequence block (Block 6), expressing the difference in RT as a percentage of the sequence block R; this number is called the Skill Score.

Some researchers prefer to compare the reaction time in sequence blocks to reaction time in the random blocks both preceding and following. In this case, first the mean reaction times for the preceding and following random blocks (Block 3 and Block 7) are averaged. Then the mean reaction times across the three sequence blocks are averaged. The difference between these averages (random average (3,7) and sequence average (blocks 4,5,6)) is then divided by the random average. The resulting percentage is referred to here as the “Super” Skill Score.

We calculated Skill Scores and Super Skill Scores based on both mean and median reaction times for each participant in each block; medians are often preferable in SRT tasks because of high intra- and inter-individual variability, especially in early trials.

Category learning. Modeled after Seger et al. (2000) and Fried and Holyoak (1984), this task required participants to make a series of forced choice classifications of visual stimuli into one of two categories. Participants received feedback after each decision (trial) indicating whether their classification was correct or incorrect. As in the probabilistic classification task, the initial decisions are arbitrary and only with feedback are participants able to learn; thus the probability of a correct response increases with successive trials. The stimuli consisted of exemplars from two categories; each category was defined by a prototype stimulus. Participants classified 50 exemplars from each category for a total of 100 trials. The order of exemplar presentation was random but constant across participants; all participants saw the same set of stimuli.

Each stimulus (exemplar) consisted of a 10x10 grid of squares in two colors (Form A: red and blue; Form B: green and yellow). The exemplars of each category were formed by distorting category prototypes (never shown to the participants): each exemplar differed from its prototype in the color of 7 tiles randomly selected from the 100 tile grid. Each pair of prototypes shared 50 tiles (50%) in common with each other and consisted of 50 tiles of each color. Form A prototypes were recreated from Figure 1 in Seger et al. (2000) and Figure 1 in Fried and Holyoak(1984); Form B prototypes were created using MatLab (version 7.0; The Mathworks, 2011). (See Appendix 1, Figure 4 for examples).

All implicit learning tasks were administered with either Matlab (version 7.0, The Mathworks) and PsychoPy (Peirce, 2007, 2009)

Rotary pursuit. Participants were asked to use a stylus to maintain contact with a rotating spot on a flat surface. Participants first completed a practice trial to familiarize them with the task and to establish baseline speed (15, 30, 45, or 60 RPM). The speed at which the participant's time on target was closest to 5 seconds (or 25%) was selected and used for all subsequent trials. Participants' time on and off target within each trial were recorded; four trials were performed, then a break of 1 minutes, then four more trials. The shape in which the target rotated was a modified rectangle (truncated corners).

Mirror tracing. Participants traced the outline of a six-sided star while watching their hands in a mirror rather than watching their hands directly. They were instructed to stay inside the outline of the star while tracing, to look in the mirror rather than at their hands, and to trace as quickly as they could without sacrificing accuracy. Time to complete an outline and number of errors (stylus moves outside outline of star) per trial were recorded. Each participant completed five trials of mirror tracing.

Rotary pursuit and mirror tracing were included as measures of implicit motor learning in the current study. Each of these tasks was administered via a specialized apparatus built by the Lafayette Instruments company.

Other measures. In addition to the experimental tasks and standardized assessments, each participant also completed two personality tests and a multi-component survey at time 2; these were administered after all the experimental tasks and standardized assessments.

Personality tests. Previous research suggests that certain personality traits (openness to experience and intuition) may be associated with implicit learning (Kaufman, 2010). To further examine these relationships, we included two personality tests in our study. Participants' personality facets in the five-factor model were assessed using the short version (120 items) of the International Personality Item Pool Representation of the Revised Neuroticism-Extroversion-Openness Personality Inventory (NEO PI-R™) (Costa & McCrae, 1992), which is available on the Personality Test Center website (http://www.personalitytest.net/ipip/ipipneo1.htm). Second, Kiersey's temperament sorter was used to classify participants according to Myers-Briggs personality types (Kiersey, 1990). The Myers-Briggs test was administered as a Googledocs survey.

Self-report of state. The following question was included in the questionnaire that participants completed at Time 2: “Compared to your first session, how is your overall mental, physical, and emotional state today?” Answer choices were “pretty much the same” and “very different.” This question was included as an omnibus check for potential confounding factors that might create differences between T1 and T2 performance, such as intoxication, amount of sleep, and fatigue. Participants who responded “very different” were flagged but not removed from the data set.

Procedure. Testing was conducted in two sessions for each participant. In Session 1, participants completed one version of each of the implicit learning tasks (AGL, PCT, SRT, category learning) as well as rotary pursuit, mirror tracing, and one version of the CVLT, in that order. In Session 2, each participant completed the opposite version (Form A vs. Form B) for each of the implicit learning tasks, as well as the other version of the CVLT, the Kaufman Brief Intelligence Test (KBIT-2, Kaufman & Kaufman, 2004), a working memory task (reading span), a Five Factor personality inventory and a Myers-Briggs Type Indicator (Kiersey, 1990) and then the questionnaire, in that order.

For each participant, Session 1 and Session 2 were scheduled at least 7 days apart and no more than 21 days apart. Average time between sessions for all participants was 13.4 days. Forty one (41) participants received Form A of the experimental tasks and the standard version of CVLT first; 34 received Form B and the alternate version of the

CVLT first. All participants were debriefed and given the opportunity to ask questions after the conclusion of Session 2 testing.

Results

Results below attempt to answer the following questions: 1) Did implicit learning occur in the experimental tasks? 2) Was there agreement across forms of the tasks? 3) Were stable individual differences in implicit learning observed? 4) Was implicit learning associated with IQ? 5) Could observed differences in implicit learning task performance be explained by differences in explicit processes such as IQ and working memory?

Overall performance on experimental tasks. As a group, the participants demonstrated implicit learning via performance on the experimental tasks, in keeping with previous studies. Table 9 displays the mean accuracy for all participants on the forced choice tasks (AGL, PCT, Category learning) as well as the skill scores for the SRT. In all tasks, participants performed statistically above chance, demonstrating that learning did occur within each task. For the forced-choice tasks, scores above 50% (chance) indicate learning; in each case, we are also able to reject the null hypothesis that the difference from chance is within the range of participant performance variation. For the SRT, t-tests confirm that we are able to reject the null hypothesis that the skill score (any formulation of the skill score) is zero at $p<0.001$. Furthermore, SRT studies often report mean trial-type RT differences (random trials mean RT minus sequence trials mean RT) on the order of tens of milliseconds; here we have RT differences of 70-80ms, consistent with previous studies. By any conventional measure, implicit learning was demonstrated on the implicit learning tasks.

Agreement between parallel forms of experimental tasks. Roughly half the participants received Form A of each experimental task at Time 1 and Form B at Time 2; the remaining participants received the task versions in the alternate order. Thus for each task, four possible mean scores exist: Form A administered at Time 1 (A1); Form A administered at Time 2 (A2); Form B at Time 1(B1) and Form B at Time 2 (B2). Each participant completed either A1 and B2 or B1 and A2. Thus, the order of form presentation was counterbalanced across participants. Before looking at the reliability between Time 1 and Time 2 measures, several tests must be conducted. For the T1-T2 scores to be comparable, we need to confirm that order of presentation (A1-B2 vs. B1-A2) did not affect performance on each task, and also that within-task, across-session (and therefore across form) learning did not take place.

Table 9 displays the mean and standard deviation at A1, A2, B1, and B2 for all measures of all implicit learning tasks used in this study, as well as the corresponding scores for the CVLT, as a comparison. Generally A1-A2 scores appear to be in a similar range, as do B1-B2 scores; in some cases (particularly PCT scores), A scores are systematically different from B scores[31]. This does not violate our necessary criteria (defined above) but does demonstrate the need for score equating (explained below). Correlations across time for each pair of forms was compared (A1-B2, A2-B1). Across forms, low-to-moderate statistically significant correlations were achieved for all tasks except AGL; despite the very similar mean scores for Form A and Form B of the AGL task, the correlation between individual scores on Form A and Form B was low and not

[31] While the particular details are too lengthy to include here, complete information can be found in Tables 9a and 9b.

significant. This finding was the first of several suggesting that AGL may differ from the other tasks in important ways.

Differences between Time 1 and Time 2 scores were calculated for each set of Forms (i.e. A1-B2; A2-B1); t-tests were then conducted on these difference by order of presentation (H_0: A1-B2 = A2-B1). In most cases, the null hypothesis could not be rejected. However, this was not the case for some PCT scores.

Noting the discrepancy for PCT scores, we nevertheless continued and transformed Form B scores onto the scale of Form A scores (for all tasks, for all measures). This linear transformation was accomplished by z-transforming Form B scores ($z_B = \frac{B-\mu_B}{\sigma_B}$) then adding the mean of Form A scores (μ_A) , then multiplying by the standard deviation of Form A (σ_A):

$$\text{Transformed B score } = \sigma_A\left(\frac{B-\mu_B}{\sigma_B}\right) + \mu_A$$

In this way, each Form B score was transformed onto the scale of Form A, rendering them comparable (Kolen, 2004)[32]. These transformed B scores (for all subjects, all measures, and all tasks) were then used to investigate the reliability from Time 1 to Time 2.

Test-retest reliability of implicit memory measures. Correlation across test-retest/parallel forms can provide an approximation of the reliability of a measurement. Table 10 displays each of the measures and the correlation between the T1 and T2 value for each measure. The measures are explained below by task.

AGL measures. Total accuracy is equivalent to the "hits" (correct endorsements of grammatical test items). Endorsement rate of high chunk strength items is the percent of

---

[32] This process is akin to equating scores across two versions of a test where one version is more difficult than the other.

high chunk strength test items endorsed, regardless of grammaticality. Each of these measures is a percentage, with a possible range from 0 to 100%. To account for participants' response bias, we also calculated d' (d prime) score for endorsement of grammatical items: z-score of the percentage false positives subtracted from the z-score of the percentage of hits (MacMillan & Creelman, 1991). We also examined the raw (untransformed) difference between percent hits and percent false positive responses. The highest correlation found was between dprime scores (T1-T2 r=-.156; A-B r= .250), but even these fell short of statistical significance. Visual inspection of the T1-T2 and Form A-Form scatterplots did not suggest a non-linear relationship.

Category learning measures. Total accuracy is equivalent to the percentage of correct classifications across all trials. We also examined accuracy within each quarter of the trials (chronologically). An accurate trial is one in which the participant correctly categorizes the stimulus. Since learning takes place throughout the task, accuracy is expected to increase over time. The mean accuracy increases with each successive block, illustrating this trend. For this reason, the low correlations found between T1-T2 and Form A-Form B for the early quarters are understandable: in the early blocks, participants are largely guessing, but based on feedback their accuracy improves across blocks (quarters). We found moderately strong agreement between Form A and Form B for total accuracy (across all trials r = 0.4637, p<0.001), quarter 3 accuracy (r = 0.4413, p<0.001) and quarter 4 accuracy (r = 0.220, p<0.05). Visual inspection of the bivariate distribution of Form B q4 accuracy on Form A q4 accuracy is similar to that in q3, but here a ceiling effect may be depressing the correlation (scores for both forms are quite high) via restriction of range.

Similar relationships were observed for T1-T2 correlation in this task; indeed, these were the highest correlations found in the study (total: $r = 0.487$, $p<0.001$; q3: $r=0.4562$, $p<0.001$; q4 $r=0.2069$, *n.s.*). Visual inspection of the scatterplots suggests some ceiling effect, but no egregious outliers or suggestions of non-linear relationships. It is possible that a less restricted range of scores would reveal a stronger correlation between T1 and T2 scores, but a wider range of accuracy scores would also mean that more participants failed to master the task. Here alternative metrics may be of use in future studies (see Appendix 3).

PCT measures. The measures for the probabilistic classification task are calculated in a similar manner to the measures for the categorization task. One property specific to the PCT is the nature of a "correct" response. Since cues are probabilistically related to targets, trial-by-trial accuracy is not an appropriate measure; instead, following the example of Gluck, Shohamy and Myers (2002), we considered the target most frequently related to each stimulus to be the optimal response and based accuracy on each subject's optimal responses. As in the category-learning task, accuracy is expected to increase over time since learning takes place throughout the task. However, drop-offs in accuracy are sometimes seen in the final block or quarter of trials and are usually attributed to fatigue or frustration effects. Conventionally, accuracy in the third quarter of trials is used as the measure of implicit learning in probabilistic classification tasks, perhaps because this is often where a peak in accuracy is seen (the fourth-quarter decline may be due to subject fatigue).

As seen in Table 4, we found T1-T2 correlations for this task were about the same were weaker in early quarters and stronger in later quarters (Q1 r=0.024, Q2 r=0.249, Q3 r=0.361, Q4 r=0.359).

SRT measures. Before analyzing the SRT data, we calculated the mean and standard deviation of reaction time for each participant; we then discarded any trials in which the RT was more than two SDs away from the mean. This elimination prevents the inclusion of spurious trials in the analysis (RT too short: participant probably hit the key by mistake; RT too long: participant stretched, yawned, got distracted, etc. instead of reacting immediately). We computed the RT difference as follows: first, comparing mean reaction time in the last sequence block (Block 6) to the mean reaction time in the immediately following random block (Block 7); the raw difference is expressed in milliseconds. However, to provide context, that difference is then divided by the mean reaction time in the last sequence block (Block 6), expressing the difference in RT as a percentage of the sequence block R; this number is called the Skill Score.

We calculated Skill Scores and Super Skill Scores based on both mean and median reaction times for each participant in each block; medians are often preferable in SRT tasks because of high intra- and inter-individual variability, especially in early trials. Regardless of how the Skill Score was calculated, agreement between forms and across testing sessions was moderate and statistically significant. The median-based Skill Score had the highest agreement (T1-T2 r=0.3677, p<0.01; Form A-Form *B* r=0.4131, p<0.001). Visual inspection of the T1-T2 and Form A-Form scatterplots did not suggest a non-linear relationship.

Test-retest reliability of an explicit memory measure. Tables 9 and 10 also display the same metrics for the two versions of the CVLT. Agreement between forms ($r = 0.76$, $p<0.001$) and across testing sessions ($r = 0.77$, $p<0.001$) is high, as would be expected for a published, standardized assessment. However, note that the correlation between forms falls just short of 0.80, which is often considered a benchmark for published assessments. The scatterplots for the CVLT score are highly linear even those participants who rated their state at T2 as "very different" from T1 are not out of trend. Figure 5 displays these scatterplots.

Cross-task correlations for explicit process measures. In addition to the CVLT, our assessment battery included a working memory test (Reading Span: RSPAN) and verbal, non-verbal, and composite intelligence. Given the well-established relationship between working memory and intelligence, and that RSPAN and CVLT both require the use of verbal short-term memory, and that RSPAN and CVLT are verbally intensive, it seems reasonable to expect considerable correlations among these measures. However, we found only moderate correlations (Table 11a). Visual inspection of the scatterplots did not suggest any unusual observations or non-linear relationships.

Implicit Learning and Intelligence. Table 12 displays the correlations between verbal, non-verbal, and composite measures of intelligence based on the KBIT and measures of implicit learning from experimental tasks. Although the correlations are not always consistent from T1 to T2, there do appear to be moderate correlations between intelligence measures and implicit learning measures. At Time 1, both verbal, non-verbal intelligence and composite IQ are negatively correlated with AGL grammatical sensitivity(verbal $r = -0.2669$, $p<0.05$; non-verbal $r = -0.2229$, n.s.; composite: $r =$ -

0.3652, p=0.013) and high chunk strength endorsement (verbal r = -0.3564, p<0.001; non-verbal r = -0.1827, n.s.; composite r = -0.3046, p<0.05); the negative relationship is stronger for verbal intelligence. At Time 2, both verbal and non-verbal intelligence are weakly correlated with category learning performance but seem to have an additive effect as the relationship between composite IQ and category learning is moderately strong (r = 0.332, p<0.05). A similar pattern is present for IQ and PCT performance trending toward statistical significance (r = 0.229, p = 0.06)

Cross-task Relationships. To investigate whether the various experimental tasks were indeed measuring a common underlying capacity, and whether this capacity was associated with IQ and working memory, we used a series of quantitative tools: First, we examined pairwise correlations between tasks; next we used Cronbach's methods for internal reliability analysis treating each task as an "item" in an item analysis; next we used PCA to establish the dimensionality of the data. Finally, we used exploratory factor analysis (EFA) to test whether a unidimensional model could accommodate both the implicit learning tasks as well as explicit processes.

Correlations. Beyond establishing the reliability of implicit learning measures, we were also interested in determining whether these disparate tasks were in fact measuring a common underlying ability. The first step toward answering this question was to examine cross-task correlations. For each task, the measure with the highest T1-T2 correlation (see Table 10) was used; for AGL, grammaticality and chunk strength endorsements were treated separately since they potentially indicate different types of learning. Table 11b displays these correlations. At Time 1, we found noteworthy correlations between category learning and endorsement of High Chunk Strength items in

AGL($r = 0.317$, $p<0.05$); and between SRT and PCT performance ($r = 0.342$, $p<0.01$). At Time 2, PCT scores were correlated with AGL grammatical sensitivity (dprime) ($r= 0.3245$, $p<0.01$) category learning ($r = 0.4249$, $p<0.001$). Examination of the pairwise scatterplots suggested a non-linear relationship between some of the measures and the SRT median skill score, so we log-transformed the SRT median skill score and looked again. While the transformation did improve the appearance of linearity in the scatterplots, it did not substantially alter the correlations. No other non-linear relationships were immediately apparent.

Tables 11 and 12 also include measures from the Mirror Tracing (percent change in time and errors) and Rotary Pursuit tasks (percent change in time on target), classic indicators of procedural learning and as such a type of non-declarative learning. Interestingly, the motor learning tasks correlate only weakly (n.s.) with the implicit learning tasks.

Preliminary item analysis—Cronbach's Alpha. If we consider implicit learning ability the single construct that we are interested in and we treat the key measure from each of the tasks (at both Time 1 and Time 2) as items, we can examine the extent to which the items intercorrelate and therefore can be considered measures of that single construct. Cronbach's alpha provides an estimate of internal consistency reliability for these "items." Again, reliability estimated this way can be construed as the ratio of "true score" variance over total variance. If "implicit learning" is a relatively stable trait and if the laboratory tasks traditionally used to measure implicit learning do in fact all reflect this trait, then the estimate of internal consistency reliability should be high.

Furthermore, the addition of “items” that do not measure the same construct should decrease the estimate of internal consistency reliability.

We first estimated internal consistency reliability using multiple measures from the category-learning and probabilistic classification tasks, treating the average accuracy across each quarter of trials as a separate item. Based on the test-retest reliability, we included only the third and fourth blocks of probabilistic classification but all four blocks of category-learning. For AGL and SRT, the most reliable measure for each was used (d’ and median-based skill score, respectively). Using the standardized versions of these measures, the estimated internal consistency reliability was $\alpha = 0.8098$ (Table 13a). Notably, the excluded-item alpha for AGL dprime at Time 1 was higher than the overall alpha for all items.

Next we included High Chunk Strength endorsement measures from the AGL task to see whether HCS endorsement was also tapping into the same ability that d’ was measuring. However, we found that HCS endorsement did not seem to fit well with the other measures, as evidenced by the overall $\alpha = 0.8010$, excluded-item α for T1 HCS= 0.8151 and excluded-item α for T2 HCS= 0.8010.

Next, we included measures for mirror learning (error and time improvement) and rotary pursuit to see if these too were measuring some common element. However, the overall internal consistency reliability estimate was lower than when these items were excluded and likewise their excluded-item alphas were high (Table 13b).

Next we included measures for explicit learning (CVLT), verbal and non-verbal IQ, and working memory span. Again, overall alpha decreased (Table 13c).

As a result of these estimates of internal consistency reliability, we moved forward with the hypothesis that some common construct was being measured by AGL, PCT, SRT, and category-learning; that the motor learning tasks were not measuring this construct; and that the construct is not correlated with intelligence or working memory span.

Principal components analysis. Given the results of the item analysis and pairwise cross-task correlations, a principal components analysis seemed well-motivated. Tables 8 display the results. Each resulting vector (component) from a PCA can be expressed as a linear combination of the "predicting" variables:

(Eq 3) $PC1_i = w_1X_{1i} + w_2X_{2i} + w_3X_{3i} + w_4X_{4i}$

Where PC1 is the first principal component, each X_i is the value of that variable for a given participant(*i*), and each w is the weight (value of the coefficient) for that variable in the eigenvector for the first principal component.

The purpose of the PCA was to see whether time-point (T1 vs. T2) emerged as a component, whether skill learning tasks (SRT, motor learning tasks) and habit learning (PCT, category-learning) tasks loaded preferentially on different factors, and whether AGL separated from the other tasks.

Table 14 displays the eigenvalues and eigenvectors for the analysis described above. Given the eigenvalues and the related screeplots, we restricted our interpretation to the first five principal components.

Time-point did not appear to drive any of the components. However, the first principal component has moderately strong positive coefficients for all but one of the tasks (AGL at T1). Perhaps this component could be interpreted as "implicit learning ability." In this first component, oefficients for the motor learning tasks were small in magnitude and in one case negative. The strongest coefficients in the second principal component come from the probabilistic category task, suggesting this component is dominated by that task. Beyond this, we did believe substantive interpretations for the other components were warranted[33].

Exploratory factor analysis. Given the results of the item analysis and principal components analysis, we performed an exploratory factor analysis to compare nested factor models. Based on the results of the internal consistency reliability estimates and the PCA, we excluded the Time 1 measure for AGL dprime; at this time we cannot propose an explanation for its anomalous behavior.[34] We were seeking to answer the following questions with LR chi-square hypothesis testing: Could a unidimensional (one-factor) model for implicit learning measures be supported? Can we show that explicit abilities are not driving implicit learning?

Table 15a shows the results of a factor model using maximum likelihood estimation and contraining the number of factors to one. The LR chi-square test for this model against the saturated model did *not* lead us to reject the model ($\chi^2(14) = 20.07$, $p = 0.128$). A one-factor model to explain the implicit learning tasks is not implausible.

Tables 15b and 15c show the results of factor analyses for the above-listed variables as well as T1 CVLT, T2 CVLT, verbal IQ, non-verbal IQ, and working memory span. The first model does not constrain the number of factors, while the second model

[33] We were also interested in the possibility that High Chunk Strength endorsement was a measure of implicit learning. In models not shown here, it became clear that HCS did not fit together with the other measures.

[34] One possible explanation could have been that at T1 paricipants were too unfamiliar with the task and/or that at T2 participants were successfully using explicit strategies to complete the task. However, no measures of explicit learning or intelligence correlated strongly with T2 AGL dprime (or the other implicit learning measures).

constrains the number of factors to one (i.e. unidimensional model). In the multi-factor model, the new variables have small or negative loadings on the first factor. The unidimensional model is rejected compared to the saturated model ($\chi^2(14) = 20.07$, $p = 0.128$). A single factor cannot explain both implicit learning and explicit abilities.

Discussion

We began this study with the goal of answering three questions: 1) Whether and to what extent reliable individual differences in implicit learning ability/performance exist and can be measured. 2) Whether and to what extent individual differences in performance on implicit learning tasks can be explained by intelligence, working memory, or explicit learning. 3) Whether and to what extent the various laboratory tasks used to investigate implicit learning are in fact measuring a common construct.

The results of the test-retest reliability study address the first question. For three of the tasks, we found moderate correlations between test sessions on PCT, SRT, and category learning. This finding suggests that (contrary to Reber's (1993) prediction as well as general expectations of dual-process theories), implicit learning (at least as measured by these tasks) does in fact vary meaningfully across individuals and that individual differences in implicit learning can be reliably measured. If this is the case, then it is possible that a standardized assessment of implicit learning could be developed as a step towards individualizing instruction with implicit instructional methods. Currently the measures needed to attain an internal consistency reliability α of at least 0.8 (Table 13) require a great deal of time and effort, but we now know that this criterion is attainable, and the measures used can potentially be refined into a less time-consuming instrument.

However, alternative explanations of the within-task correlations between Time 1 and Time 2 are possible. One possibility is that the common variation across individuals from T1 to T2 reflects explicit learning, working memory span, or intelligence. Based on post-task (T2) interviews as well as observations, many participants did acquire some explicit knowledge of the sequence in the SRT task and the regularities in the PCT task[35]. This fact reflects the difficulty, even in the laboratory, of finding a "process pure" task.

I attempted to address the plausible alternative hypothesis that the correlation between T1 and T2 scores is being driven by an explicit process taking place in parallel with implicit learning in several ways. First, correlations between each of the task scores and the explicit process measures (IQ, WM, CVLT) were examined (Table 12). These correlations were generally low, and non-significant, which did not suggest that performance on implicit learning tasks was driven by IQ. Furthermore, when we attempted T2-T1 regressions with IQ or CVLT scores as covariates, IQ did not explain any additional variance; this finding casts doubt on the explanation that the observed individual differences were caused by explicit processes. Likewise, in the principal components analysis study, the explicit process measures did not tend to group with the implicit learning task scores. Finally, hypothesis testing using nested factor models allowed us to reject the possibility that performance on the explicit process tasks and performance on the implicit learning tasks could be explained by a single common underlying factor. The implicit learning tasks could plausibly be explained by a common factor, but when explicit process measures were included, the unidimensional model was rejected.

[35] Anecdotally, for PCT, and in particular Form B of PCT, participants seemed to be explicitly hypothesis-testing.

Notably, several tasks traditionally used as measures of implicit or procedural learning did not follow this pattern. Specifically, artificial grammar learning, rotary pursuit, and mirror tracing did not seem to be tapping the same construct as PCT, SRT, and category learning. Scores on the artificial grammar learning task (AGL) did not demonstrate test-retest reliability (i.e. correlations between T1 and T2 scores were low and non-significant. This finding is particularly notable because most of Reber's studies, on which he based his ideas about implicit learning more generally, used AGL as the primary or only measure of implicit learning. With regards to AGL, his prediction of a lack of stable individual differences appears to have been correct.

However, another explanation is possible: when participants returned at T2, their previous exposure to the task made them approach AGL at T2 in a different manner from their T1 approach. Specifically, since the testing phase of AGL requires disclosing to participants that a pattern was present in the training stimuli, participants may have anticipated that a pattern was present during the training phase of AGL at T2 and attempted to explicitly learn frequently co-ocurring letter combinations. This explicit effort may explain the lack of correlation between AGL scores from T1 to T2.

To establish the reliability of individual differences in AGL performance then, a test-retest paradigm may need to disguise the T2 AGL task more than was done here, for example presenting it in a different modality or with numbers or nonsense symbols rather than letters in the second version. Otherwise, an internal consistency reliability could potentially be calculated from among the test items (since learning is not expected to occur during the test phase) although the possible scores for each item are only 0 or 1.

AGL scores also did not behave like PCT, SRT, and category learning with regard to explicit process measures. Correlations between AGL and verbal and non-verbal IQ were not only significant, but negative (verbal r = 0.26, p =*0.05*; non-verbal r = 0.25, p= 0.04). The fact that the observed relationship between verbal IQ and AGL was negative, suggests perhaps that verbal knowledge/ability hampers implicit learning on this task. This is not unlike the finding of Finn et al. (in press) that adults' explicit processing impedes statistical learning of an artificial language. The finding of a negative correlation with non-verbal intelligence in more difficult to interpret. One might venture that perhaps participants with higher IQs are making more deliberate effort on the task, which could backfire. However, two pieces of data argue against this interpretation: the correlations between AGL performance[36] and Implicit Theory of Intelligence and Self-efficacy (r = -0.29, p = 0.02; r = 0.25, p = 0.06, respectively). On the one hand, participants who believe that intelligence is malleable and those who have high ratings of self-efficacy perform better, but on the other hand participants with high IQs perform worse. Further research is needed to verify and explain these apparently contradictory results. At this point, we recommend that conclusions from studies using AGL tasks not be assumed to generalize to implicit learning in general.

Another surprising finding was that performance on the rotary pursuit and mirror tracing tasks did not correlate with the implicit learning tasks; the PCA and EFA also suggested that rotary pursuit and mirror tracing did not share a common source of variance with SRT, PCT, or category learning. This finding is particularly surprising for SRT, since SRT has been considered a procedural learning task akin to rotary pursuit and

[36] Note that these correlations only apply to T1 AGL dprime. Correlations between explicit process measures and AGL dprime scores at T2 did not exceed r = 0.15 and did not approach significance).

mirror tracing. It may be the case that differences in cerebellar-mediated motor learning are overpowering variance due to (basal ganglia-mediated) implicit learning, or it may be the case that in fact the cerebellum and not the basal ganglia is primarily driving performance on this task. Furthermore, performance on the two tasks did not correlate with each other. One possibility reason for this is that they differ in their feedback structure ("open-loop" vs. "closed-loop").

Conclusion. Based on the results of this study, we now know there is a reliably measureable cognitive capacity that varies across individuals, but that is uncorrelated with IQ and working memory. This independence from IQ and WM is remarkable because few if any commonly investigated cognitive abilities fail to correlate with general intelligence (g) on some level. Throughout his exposition on implicit learning, Reber (1993) maintained that *in vivo* implicit and explicit learning processes probably operated in parallel or even in a complementary manner. By including implicit learning measures, we may now be able to account for additional, previously unaccounted for variance in performance on *in vivo* learning tasks. Now that we know that stable individual differences in implicit learning ability exist and can be measured, we can use these to predict and explain differences in learning across a wide variety of domains. For example, given the importance of automatic processes in reading and arithmetic, we can investigate whether students with greater (general) implicit learning ability on average are better at math fact recall or reading fluency. Furthermore, we now have a foothold from which to begin exploring potential interactions between implicit and explicit processes and how such potential interactions could potentially be harnessed to maximize performance in learning tasks.

Chapter Six: Conclusion and Directions for Future Study

We started with the postulated features of implicit learning specified by Reber:

1. Robustness to injury or disease
2. Age independence
3. IQ independence
4. Low individual variability
5. Conservation across phylogeny

At the outset of this dissertation, there was abundant evidence for selective sparing of implicit learning in various disease and injury conditions, but only sparse and mixed evidence for the age-independence or IQ-dependence; furthermore, there had been no empirical attempts to support or contradict the fourth. Based on the results of the studies described in Chapter 4 and Chapter 5, those postulated features must be re-evaluated. In particular, age-independence and low variability across individuals may no longer be defensible. These findings have implications for both basic research on implicit learning as well as for the possibility of applying research on implicit learning to instructional strategies.

Review of Findings

Regarding developmental non-invariance (age dependence), while we did not find gross differences in overall performance on implicit learning tasks between children and adults, nevertheless we found strong evidence that children's computational strategies differ from adults' across several tasks. In particular, children seemed to disregard item frequency, in contrast to adults who seemed to use item frequency as an important computational feature of the stimuli. Based on previous developmental studies as well as language acquisition research, it is not unexpected that children would be less sensitive to item frequency (Amso & Davidow, 2012; Finn, Lee, Kraus, & Hudson Kam, 2014; Gómez & Gerken, 2000; Marcovitch & Lewkowicz, 2009; Seidenberg, Macdonald, & Saffran, 2002); indeed, some developmental computational models (Frank & Tenenbaum, 2011; Tenenbaum et al., 2011) suggest such frequency insensitivity can be an asset in the difficult task of inducing data structures from sparse data.

With regard to individual differences in implicit learning performance, we found stable inter-individual differences across time and across tasks. Importantly, performance on implicit learning tasks was independent of verbal intelligence, non-verbal intelligence, and working memory. Thus we tentatively conclude that reliably measurable differences in implicit learning exist and cannot be attributed to differences in intelligence, working memory, or other explicit processes or capacities.

Implications for Future Research

The finding that implicit learning changes with age is in keeping with anatomical data about the neural correlates of implicit learning. We know that anatomically the basal ganglia mature throughout childhood into adulthood, so it makes sense that the functions of the basal ganglia would take time to mature as well. The development of skill and habit learning ability may depend on not only the maturation of the basal ganglia, but also the growth of its connections with other structures, for example via cortico-striatal white matter tracts. Future studies can trace the relationship between changes in the computations underlying implicit learning and changes in structural and functional connectivity using DTI, resting state functional connectivity studies, and high resolution structural MRI.

Similarly, imaging research with adults could establish whether anatomical properties of the basal ganglia and/or its connections to other cortical areas correlate with differences in implicit learning performance (across individuals, within age group). Behavioral research with adults could also establish whether differences in implicit learning performance are qualitative or quantitative—that is, whether low-performing adults are using the same computational strategies as high-performing adults, but less effectively, or if in fact they are using different computational strategies.

An important question for further research is the precise developmental course of implicit inductive learning and to what extent it parallels what we know about the development of inductive inference from studies that did not specify that the learning was implicit. In particular, if it is found that certain information structures (or the computations necessary to induce them) become available to children for both implicit learning and explicit in tandem, then a more parsimonious explanation requires some common substrate for both implicit and explicit use of inductive learning.

Along the same lines, the finding of stable individual differences can be further investigated. Other than IQ and age, what characteristics determine or influence an individual's habit learning and skill learning ability? One possibility is the individual's repertoire of abstract relations/information structures. For example, if the individual does not have a stable, abstract representation of the relation "is-a," then perhaps he would be unable to induce this relation in a habit-learning paradigm. This idea is analogous to my proposal above that children's ability to induce particular information structures is limited by their ability to (explicitly) understand such structures. Notably, this proposal has yet to be tested either developmentally or within adults, but opens an avenue for future research.

Finally, recent work that could lead to best practices for applied habit learning paradigms may need to be reconsidered in developmental terms; specifically, findings related to effects of the order of examples presented (Sandhofer & Doumas, 2008), the timing and spacing of trials (Birnbaum et al., 2013; Vlach et al., 2014), salience of certain stimulus features (Noh et al., 2014) and other aspects of training session construction (Lindsey et al., 2013) will need to be replicated or investigated with children of different ages, since the existing work has been conducted almost exclusively with adults.

Theoretical implications

Taken together, these findings paint a different picture of implicit learning than that which has dominated the scientific landscape for the past few decades. Rather than being a primitive, developmentally-invariant, individual-invariant, IQ-independent capacity, implicit learning varies across individuals with development, IQ, and possibly other individual-level variables. In particular, the findings of developmental non-invariance and IQ dependence raise questions about the extent to which implicit learning is "information encapsulated" in a Fodorian sense. (Indeed, Fodor's (1983) proposed characteristics for modules[37] are compatible with other sets of proposed characteristics of implicit processes or systems (Evans, 2008)).

[37]From Fodor (1983):
1. Domain specificity
2. Mandatory operation
3. Limited central accessibility
4. Fast processing
5. Informational encapsulation
6. 'Shallow' outputs
7. Fixed neural architecture

In the time since Fodor's characterization of mental modules, many previously "hard modular" models have been "softened" to include provisions for cascading (see e.g. Coltheart et al., 2001). This "softening" may be now be necessary for implicit learning. Furthermore, other assumed characteristics of implicit learning have been brought into question. Specifically, the idea that basal ganglia-dependent learning is always and necessarily inaccessible to awareness, while hippocampal learning is necessarily accessible has been contradicted (Ortu & Vaidya, 2013; Seger & Spiering, 2011).

Perhaps rather than being isolated, low-level, modular capacities, instead habit learning and skill learning are more centrally located and the representations generated by habit learning could become accessible to awareness under certain circumstances.

Implications for Education Research and Practice

In Chapters 2 and 3, I suggested ways that implicit learning research might be relevant to education. Specifically, I suggested that basic research on implicit learning could be applied to improve skill learning paradigms; to bypass immature explicit systems in young children and/or individuals with impaired explicit processing ability, such as low-IQ individuals; to counteract lingering effects of intuitive theories in conceptual change teaching; and to combine fluency and flexibility in skill use. Given the results of the studies conducted, we can now reevaluate these recommendations.

Regarding skill learning and fluency, nothing in the current studies suggests that findings from basic research on skill learning cannot be adapted into best practices and recommendations for instructional practice. Wherever a cognitive skill can be identified in the K-12 curriculum, practice with feedback can lead to fluency.

---

8. Characteristic and specific breakdown patterns
9. Characteristic ontogenetic pace and sequencing

On the other hand, the idea of using implicit learning to bypass immature explicit learning systems in young children or for example, low-IQ individuals may not be viable. We now have evidence that although children can and do learn implicitly, the computational strategies they use and the stimulus features they attend to may differ significantly from those used by adults. Presenting the same stimuli with the same feedback structure to children and adults will not result in the same knowledge structure being induced. Possibly, modifications could be made to optimize such paradigms for children after basic research on the development of children's computational strategies has established a developmental timeline. Similarly, basic research on the computational strategies available to MR individuals would be necessary before findings from the existing implicit learning literature could be applied to those individuals.

It may still be possible and advantageous to pair explicit conceptual change instruction with habit learning style examples-with-feedback to counteract or replace incorrect intuitive theories. However, before findings from research with adults can be applied to children, again detailed basic research on how children's habit learning develops and how it differs from adults will be necessary.

Finally, although again more research is needed, the tantalizing possibility of coordinating across memory systems to optimize performance still exists. For example, in a series of experiments with adults, Ashby and Crossley (2010) determined that declarative and procedural systems can interact (in a manner other than all-or-nothing competition) in category learning and that this interaction may be mediated by the hyperdirect pathway from the frontal lobe to the basal ganglia via the subthalamaic nucleus. Perhaps ways could be found to coordinate implicitly and explicitly learned information structures within the same domain, providing a means for fluency with flexibility.

For any basic research on implicit learning to become applicable to K-12 instructional practice, one key task must be completed. Learning objectives that require skill learning or that can be reconceptualized as category-learning or habit-learning in the K-12 curriculum must be identified. Current statements of learning objectives, for example in the Common Core, are not readily translatable into skill learning and category learning, but an individual with a background in both education and cognitive science could easily carry out this translation given a year or so of postdoctoral funding and no concurrent teaching requirements.

Appendix 1: Tables and Figures

Tables

Table 1.

Demographic Characteristics of Participants

	10 year olds (n =32)		Adults (n=29)	
Characteristic	*n*	*%*	*n*	*%*
Sex				
Male	16	50	13	45
Female	16	50	16	55
Race				
Asian	5	15	4	14
Black	1	3	2	7
White	26	81	23	79

Table 2a
Means and Standard Deviations, Differences Between Sample Values and Population Norms

Measure	Population Norm M	Population Norm SD	Sample Value M	Sample Value SD	*t (1-sample to pop norm)*	*t (2-sample by group)*
Intelligence						
Standard Verbal IQ						
Adults	100	20	117.97	17.18	5.02***	
10-year-olds	100	20	115.27	11.96	6.51***	0.54
Standard Non-verbal IQ						
Adults	100	20	114.34	14.06	5.64***	
10-year-olds	100	20	114.17	9.97	17.01***	0.06
Standard Composite IQ						
Adults	100	20	119	17.18	5.96***	
10-year-olds	100	20	116.73	9.28	9.30***	0.46
Explicit Learning						
CVLT Recall Total (t scale score)						
Adults	50	10	59.75	8.45	5.03***	
10-year olds	50	10	52.5	7.29	1.14	2.56*

*p<0.05, **p<0.01,
***p<0.001

Table 2b.

Means, Standard Deviations, and Group Differences Between Adults and 10-year-olds

	10 year olds		Adults			
Measure	M	SD	M	SD	*t(60)*	*p*
Intelligence						
Raw Verbal IQ	72.81	8.47	96.45	9.89	9.58***	<0.001
Raw Non-verbal IQ	35.59	3.76	41.17	4.44	5.05***	<0.001
Processing Speed and WM						-
Processing Speed	36.10	13.73	63.02	14.68	12.91***	<0.001
WM Span	3.57	1.43	5.41	1.65	5.56***	<0.001
Explicit Learning						-
CVLT Recall Total	52.50	7.29	59.75	8.45	3.16**	0.003
CVLT Recognition	0.79	0.04	0.99	0.03	8.93***	<0.001

*p<0.05, **p<0.01, ***p<0.001

Table 3.

Implicit Learning Tasks: Overall Performance Measures by Group

	10 year olds		Adults			
	M	SD	M	SD	*t(60)*	*p*
SRT Skill Score	0.08	0.09	0.13	0.05	1.39	0.17
PCT %Optimal	0.62	0.12	0.63	0.13	0.77	0.93
AGL d'	0.36	0.49	0.32	0.58	0.97	0.83
Mirror delta time	51.23	65.17	27.12	32.96	1.84	0.07
Mirror delta errors	29.96	43.85	18.26	26.69	1.26	0.21
Rotary delta TOT	2.92	2.17	2.81	2.29	0.19	0.85

*p<0.05, **p<0.01, ***p<0.001

Table 4.

Means, Standard Deviations, and Group Differences in Endorsement of AGL Test Items Between Adults and 10-year-olds

	10 year olds		Adults			
Item Type	M	SD	M	SD	*t(60)*	*p*
Grammatical HCS	0.63	0.21	0.66	0.22	0.42	0.67
Grammatical LCS	0.70	0.22	0.59	0.23	1.45	0.16
Non-grammatical HCS	0.59	0.26	0.62	0.25	0.44	0.66
Non-grammatical LCS	0.50	0.25	0.39	0.24	1.70	0.09

Table 7.

Correlations between implicit learning task performance and intelligence reported in previous studies

Authors	Task(s)	IQ measure	IQ-IL Correlation	Subjects
Reber, Walkenfeld & Hernstadt (1991)	AGL	WAIS-R	0.25	college students (N=20)
Kaufman et al. (2010)	SRT	Raven's	0.13	Students aged 16-18 (N=153)
	SRT	DAT Verbal	.22*	
Mayberry et al. (1995)	Classification	Raven's + PPVT	.02-.04	6 year old children (N=62) and 11 year olds (N=52)
Gebauer et al. (2007)	AGL	Gf^	0.05	German school students aged 11-32 (N~300)
	AGL	Gc	0.00	
	AGL	Gy	.10*	
	Process Control	Gf	.10*	
	Process Control	Gc	0.01	
	Process Control	Gy	0.03	
	SRT	Gf	0.05	
	SRT	Gc	-0.04	
	SRT	Gy	0.04	

Table 8.

Correlations across tasks reported in previous studies

	SRT	AGL	Process Control	PCT (slope)
SRT	.	0.03	0.01	.
AGL		.	0.00	0.04
Process Control			.	.
PCT (slope)				.

None of the correlations reported here were statistically significant.
PCT (slope)-AGL provided by Horan (personal communication) based on Horan et al. (2008);
all other correlations from Gebauer et al. (2007). None of the correlations were statistically significant.

Table 9a.

Reliability (T1-T2 correlation) of conventional and novel measures of implicit learning

Task name	Abbreviation	Possible score range	Measure/ Operationalization	Mean Score T1	T1 SD	Mean Score T2	T2 SD	T1-T2 Correlation
Probabilistic Classification Task Knowlton, Squire, & Gluck (1994)	PCT	0-100%	Total Accuracy	0.872	0.090	0.814	0.113	0.3525*
		0-100%	Accuracy in first 25% of trials	0.785	0.105	0.743	0.116	0.024
		0-100%	Accuracy in second 25% of trials	0.902	0.112	0.839	0.123	0.249
		0-100%	Accuracy in third 25% of trials	0.906	0.104	0.845	0.124	0.3607*
		0-100%	Accuracy in fourth 25% of trials	0.893	0.112	0.834	0.143	0.3586*
Prototype-distortion category learning Holyoak & Fried (1984)	CAT	0-100%	Total Accuracy	0.848	0.094	0.856	0.108	0.5013*
		0-100%	Accuracy in first 25% of trials	0.734	0.162	0.791	0.163	0.109
		0-100%	Accuracy in second 25% of trials	0.884	0.114	0.857	0.144	0.3884*
		0-100%	Accuracy in third 25% of trials	0.886	0.115	0.882	0.138	0.4575*
		0-100%	Accuracy in fourth 25% of trials	0.889	0.104	0.896	0.095	0.2376

Table 9a. (continued)

Reliability (T1-T2 correlation) of conventional and novel measures of implicit learning

Task name	Abbreviation	Possible score range	Measure/ Operationalization	Mean Score T1	T1 SD	Mean Score T2	T2 SD	T1-T2 Correlation
Serial Reaction Time	SRT	0 and up	Mean Raw Reaction Time difference (sec)	0.096	0.056	0.088	0.070	0.3523*
Nissen & Bullemer (1987)		0 and up	Median Raw Reaction Time difference (sec)	0.101	0.064	0.097	0.080	0.4100*
		0-100%	Single block Skill Score	0.358	0.327	0.352	0.412	0.3257*
		0-100%	Single block Skill Score based on medians	0.422	0.427	0.442	0.590	0.2917*
Artificial Grammar Learning	AGL	0-100%	Total Accuracy =endorsement rate of grammatical items (abstract learning)	0.637	0.159	0.678	0.147	0.060
Reber (1962, 1989, 1993)		0-100%	Endorsement rate of high chunk strength items (concrete learning)	0.597	0.141 1684	0.609	0.129	-0.022
		-4.65 to +4.65; typical values up to 2.0	dprime score (z(P[hits]) - z(P[false positive]))	0.399	0.344	0.416	0.339	-0.157
California Verbal Learning Test	CVLT	0 to 95	Scaled T score for Sum of Trials 1-5	54.424	9.882	55.477	9.562	0.7685***
	Explicit Learning/Declarative Memory							

Table 9b.

Agreement across parallel forms of implicit learning tasks

Task name	Abbreviation	Possible score range	Measure/ Operationalization	Mean Score Form A	Form A SD	Mean Score Form B	Form B SD	Form A-Form B Correlation
Probabilistic Classification Task Knowlton, Squire, & Gluck (1994)	PCT	0-100%	Total Accuracy	0.845	0.104	0.616	0.094	0.2559*
		0-100%	Accuracy in first 25% of trials	0.766	0.112	0.551	0.087	-0.038
		0-100%	Accuracy in second 25% of trials	0.871	0.122	0.630	0.114	0.190
		0-100%	Accuracy in third 25% of trials	0.877	0.117	0.638	0.129	0.2933*
		0-100%	Accuracy in fourth 25% of trials	0.865	0.128	0.645	0.128	0.2845*
Prototype-distortion category learning Holyoak & Fried (1984)	CAT	0-100%	Total Accuracy	0.849	0.101	0.828	0.111	0.4637***
		0-100%	Accuracy in first 25% of trials	0.758	0.163	0.739	0.158	0.058
		0-100%	Accuracy in second 25% of trials	0.866	0.134	0.836	0.142	0.278*
		0-100%	Accuracy in third 25% of trials	0.882	0.125	0.871	0.123	0.4413***
		0-100%	Accuracy in fourth 25% of trials	0.891	0.101	0.867	0.141	0.220

Table 9b. (continued)

Agreement across parallel forms of implicit learning tasks

Task name	Abbreviation	Possible score range	Measure/ Operationalization	Mean Score Form A	Form A SD	Mean Score Form B	Form B SD	Form A-Form B Correlation
Serial Reaction Time	SRT	0 and up	Mean Raw Reaction Time difference (sec)	0.088	0.064	0.058	0.075	0.3997***
Nissen & Bullemer (1987)		0 and up	Median Raw Reaction Time difference (sec)	0.093	0.072	0.063	0.081	0.3439**
		0-100%	Single block Skill Score	0.209	0.372	0.144	0.523	0.3931**
		0-100%	Single block Skill Score based on medians	0.228	0.515	0.158	0.783	0.4131***
Artificial Grammar Learning	AGL	0-100%	Total Accuracy =endorsement rate of grammatical items (abstract learning)	0.656	0.155	0.673	0.119	0.039
Reber (1962, 1989, 1993)		0-100%	Endorsement rate of high chunk strength items (concrete learning)	0.603	0.136	0.570	0.123	-0.021
		-4.65 to +4.65; typical values up to 2.0	dprime score (z(P[hits]) - z(P[false positive]))	0.407	0.342	0.250	0.412	-0.156
California Verbal Learning Test	CVLT	0 to 95	Scaled T score for Sum of Trials 1-5	55.310	9.471	54.754	9.855	0.7763***
	Explicit Learning/Declarative		Memory					

Table 10.

Reliability (T1-T2 correlation) of conventional and novel measures of implicit learning

Task name	Abbreviation	Possible score range	Measure/ Operationalization	Mean Score T1	T1 SD	Mean Score T2	T2 SD	T1-T2 Correlation
Probabilistic Classification Task Knowlton, Squire, & Gluck (1994)	PCT	0-100%	Total Accuracy	0.872	0.090	0.814	0.113	0.3525*
		0-100%	Accuracy in first 25% of trials	0.785	0.105	0.743	0.116	0.024
		0-100%	Accuracy in second 25% of trials	0.902	0.112	0.839	0.123	0.249
		0-100%	Accuracy in third 25% of trials	0.906	0.104	0.845	0.124	0.3607*
		0-100%	Accuracy in fourth 25% of trials	0.893	0.112	0.834	0.143	0.3586*
Prototype-distortion category learning Holyoak & Fried (1984)	CAT	0-100%	Total Accuracy	0.848	0.094	0.856	0.108	0.5013*
		0-100%	Accuracy in first 25% of trials	0.734	0.162	0.791	0.163	0.109
		0-100%	Accuracy in second 25% of trials	0.884	0.114	0.857	0.144	0.3884*
		0-100%	Accuracy in third 25% of trials	0.886	0.115	0.882	0.138	0.4575*
		0-100%	Accuracy in fourth 25% of trials	0.889	0.104	0.896	0.095	0.2376

*p<0.05; **p<0.01; ***P<0.001

Table 10. (continued)

Reliability (T1-T2 correlation) of conventional and novel measures of implicit learning

Task name	Abbreviation	Possible score range	Measure/ Operationalization	Mean Score T1	T1 SD	Mean Score T2	T2 SD	T1-T2 Correlation
Serial Reaction Time	SRT	0 and up	Mean Raw Reaction Time difference (sec)	0.096	0.056	0.088	0.070	0.3523*
Nissen & Bullemer (1987)		0 and up	Median Raw Reaction Time difference (sec)	0.101	0.064	0.097	0.080	0.4100*
		0-100%	Single block Skill Score	0.358	0.327	0.352	0.412	0.3257*
		0-100%	Single block Skill Score based on medians	0.422	0.427	0.442	0.590	0.2917*
Artificial Grammar Learning	AGL	0-100%	Total Accuracy =endorsement rate of grammatical items (abstract learning)	0.637	0.159	0.678	0.147	0.060
Reber (1962, 1989, 1993)		0-100%	Endorsement rate of high chunk strength items (concrete learning)	0.597	0.1411684	0.609	0.129	-0.022
		-4.65 to +4.65; typical values up to 2.0	dprime score (z(P[hits]) - z(P[false positive]))	0.399	0.344	0.416	0.339	-0.157
California Verbal Learning Test	CVLT	0 to 95	Scaled T score for Sum of Trials 1-5	54.424	9.882	55.477	9.562	0.7685***
	Explicit Learning/Declarative Memory							

*p<0.05; **p<0.01; ***P<0.001

Table 11a.

Cross-task pairwise correlations for explicit tasks

	T1 CVLT	T2 CVLT	Verbal IQ	Non-Verbal IQ	Total IQ	Working Memory
T1 CVLT	1					
T2 CVLT	0.7618*	1				
Verbal IQ	0.3685*	0.3337*	1			
Non-Verbal IQ	0.0078	0.1014	0.4102*	1		
Total IQ	0.2553	0.2059	0.8651*	0.8063*	1	
Working Memory	-0.0134	-0.0627	0.1509	0.1503	0.3327*	1

Table 11b.

Cross-task pairwise correlations for implicit learning tasks

	AGL dprime	PCT	CAT	SRT	Rotary	Mirror
AGL dprime	1					
PCT	0.2684*	1				
CAT	0.0141	0.2853*	1			
SRT	-0.0287	0.0354	0.2329	1		
Rotary	0.026	0.0275	0.0536	-0.1423	1	
Mirror	0.0217	0.3873*	-0.0523	0.0897	0.1334	1

Table 11c.

Pairwise correlations between implicit learning measures and key covariates used in PCA

	T1 AGL d'	T2 AGL d'	T1 PCT	T2 PCT	T1 Cat Learning	T2 Category Learning	T1 SRT Skill	T2 SRT Skill	T1 CVLT	T2 CVLT	stan verbal	stan non verbal	iq stan	WM Span	Mirror Tracing delta Error	Mirror Tracing delta time
T1 dprime	1.00															
T2 dprime	-0.17	1.00														
T1 PCT	0.07	0.21	1.00													
T2 PCT	-0.11	0.23	0.35*	1.00												
T1 Category Learning	-0.06	0.16	0.21	0.28*	1.00											
T2 Category Learning	-0.24	0.09	-0.06	0.33*	0.50*	1.00										
T1 SRT Skill	-0.08	0.05	-0.07	0.06	0.08	0.15	1.00									
T2 SRT Skill	0.01	0.01	-0.04	-0.05	0.06	0.17	0.29*	1.00								
T1 CVLT	0.16	0.07	0.27	0.14	0.18	0.02	-0.11	0.00	1.00							
T2 CVLT	0.13	0.22	0.33*	0.29*	0.04	0.16	-0.01	0.08	0.76*	1.00						
stan verbal	-0.26*	0.03	0.02	0.02	-0.15	0.04	0.17	-0.07	0.37*	0.33*	1.00					
stan nonverbal	-0.25*	0.19	-0.03	0.10	0.14	0.23	0.16	0.04	0.01	0.10	0.41*	1.00				
iq standard	-0.35*	0.14	0.11	0.15	0.06	0.30	0.14	-0.05	0.26	0.21	0.87*	0.81*	1.00			
WM Span	-0.10	-0.21	-0.14	-0.01	0.08	0.11	0.03	0.22	-0.01	-0.06	0.15	0.15	0.33*	1.00		

Mirror Tracing delta Error	0.09	0.00	-0.16	0.00	0.10	-0.12	0.06	-0.21	0.15	0.13	0.24	0.15	0.24	0.14	1.00	
Mirror Tracing delta time	0.08	0.07	-0.19	-0.20	0.00	-0.13	0.03	0.10	0.07	0.02	-0.03	0.10	-0.09	0.05	0.34*	1.00
Rotary Pusuit %	0.13	-0.09	-0.06	0.16	-0.03	0.01	-0.11	0.03	0.08	0.10	0.15	0.12	0.24	0.14	0.14	-0.02

Table 12.
Correlations between implicit learning task measures and intelligence measures

	Verbal	Nonverbal	IQ
CATT1Correct	-0.079	0.202	0.057
CATT2Correct	0.161	0.375*	0.306
SRTlogskill1	0.244	0.082	0.198
SRTmedlogskill1	0.228	0.071	0.182
SRTlogskill2	-0.086	0.046	-0.034
SRTmedlogskill2	-0.075	0.103	0.002
PCTT1Block3	0.008	-0.042	-0.014
PCTT1Block4	0.037	0.005	0.029
PCTT2Block3	0.237	0.230	0.279
PCTT2Block4	0.126	0.099	0.139
AGL AccT1	-0.039	-0.069	-0.062
AGL AccT2	0.025	0.307*	0.191
AGL Chunk StrengthT1	-0.409*	-0.091	-0.305*
AGLChunk Strength T2	-0.090	-0.054	-0.079

Table 13a.

Item analysis - Implicit learning tasks

Item	Obs	item-test correlation	item-rest correlation	average interitem correlation	Excluded item alpha
T1 AGL d prime	78	0.4237	0.1674	0.2285	0.8163
T2 AGL d prime	67	0.3632	0.1815	0.2212	0.8099
T1 PCT Block 3 Acc	68	0.4608	0.3278	0.2157	0.8048
T1 PCT Block 4 Acc	65	0.6971	0.6211	0.1971	0.7864
T2 PCT Block 3 Acc	68	0.4643	0.3433	0.2148	0.8041
T2 PCT Block 4 Acc	65	0.5826	0.4839	0.2060	0.7955
T1 Category Learning Block 1 Acc	59	0.4424	0.3351	0.2153	0.8045
T2 Category Learning Block 1 Acc	59	0.5118	0.4123	0.2099	0.7994
T1 Category Learning Block 2 Acc	59	0.5179	0.4162	0.2092	0.7987
T2 Category Learning Block 2 Acc	59	0.6807	0.6058	0.1969	0.7862
T1 Category Learning Block 3 Acc	59	0.6601	0.5816	0.1984	0.7879
T2 Category Learning Block 3 Acc	59	0.7428	0.6791	0.1922	0.7811
T1 Category Learning Block 4 Acc	59	0.7254	0.6585	0.1935	0.7825
T2 Category Learning Block 4 Acc	59	0.5739	0.4822	0.2052	0.7947
T1 SRT Median Skill Score	65	0.2842	0.1554	0.2276	0.8155
T2 SRT Median Skill Score	65	0.2440	0.1085	0.2311	0.8185
Test scale				0.2101	0.8098

Table 13b.

Item analysis - Implicit learning items and motor skill learning items

Item	*N*	item-test correlation	item-rest correlation	average interitem correlation	Excluded item alpha
Rotary Pursuit Improvement	69	0.4178	0.2327	0.1803	0.7984
Mirror tracing time improvement	66	0.2836	0.1548	0.1850	0.8033
Mirror tracing error improvement	66	0.3555	0.2016	0.1804	0.7984
T1 AGL d prime	78	0.3818	0.1582	0.1829	0.8011
T2 AGL d prime	67	0.3501	0.2050	0.1787	0.7966
T1 PCT Block 3 Acc	68	0.4734	0.3519	0.1728	0.7899
T1 PCT Block 4 Acc	65	0.6695	0.5905	0.1605	0.7748
T2 PCT Block 3 Acc	68	0.5023	0.3973	0.1706	0.7873
T2 PCT Block 4 Acc	65	0.5282	0.4207	0.1677	0.7838
T1 Category Learning Block 1 Acc	59	0.4216	0.3181	0.1743	0.7916
T2 Category Learning Block 1 Acc	59	0.5079	0.4138	0.1694	0.7859
T1 Category Learning Block 2 Acc	59	0.4917	0.3963	0.1706	0.7873
T2 Category Learning Block 2 Acc	59	0.6714	0.6005	0.1605	0.7748
T1 Category Learning Block 3 Acc	59	0.6126	0.5346	0.1633	0.7785
T2 Category Learning Block 3 Acc	59	0.7273	0.6664	0.1557	0.7684
T1 Category Learning Block 4 Acc	59	0.6942	0.6266	0.1593	0.7732
T2 Category Learning Block 4 Acc	59	0.5359	0.4474	0.1667	0.7826
T1 SRT Median Skill Score	65	0.2748	0.1584	0.1825	0.8008
T2 SRT Median Skill Score	65	0.2388	0.1117	0.1852	0.8036
Test scale				0.1719	0.7977

Table 13c.

Item analysis - Implicit learning items and intelligence/explicit learning measures

Item	N	item-test correlation	item-rest correlation	average interitem correlation	Excluded item alpha
T1 CVLT	60	0.3412	0.1578	0.1528	0.7829
T2 CVLT	66	0.4566	0.3392	0.1487	0.7774
Verbal IQ	62	-0.0300	-0.1455	0.1685	0.8020
Non-verbal IQ	62	0.3414	0.2289	0.1508	0.7803
Working memory span	70	0.1987	0.0266	0.1587	0.7904
T1 AGL d prime	78	0.3346	0.1062	0.1559	0.7870
T2 AGL d prime	67	0.3654	0.2277	0.1495	0.7785
T1 PCT Block 3 Acc	68	0.4728	0.3650	0.1447	0.7718
T1 PCT Block 4 Acc	65	0.6662	0.5972	0.1342	0.7562
T2 PCT Block 3 Acc	68	0.4415	0.3299	0.1455	0.7731
T2 PCT Block 4 Acc	65	0.5618	0.4752	0.1393	0.7639
T1 Category Learning Block 1 Acc	59	0.4640	0.3679	0.1442	0.7712
T2 Category Learning Block 1 Acc	59	0.4720	0.3774	0.1438	0.7705
T1 Category Learning Block 2 Acc	59	0.5393	0.4505	0.1409	0.7663
T2 Category Learning Block 2 Acc	59	0.6884	0.6211	0.1340	0.7558
T1 Category Learning Block 3 Acc	59	0.6458	0.5731	0.1355	0.7582
T2 Category Learning Block 3 Acc	59	0.7200	0.6585	0.1314	0.7516
T1 Category Learning Block 4 Acc	59	0.6957	0.6302	0.1334	0.7548
T2 Category Learning Block 4 Acc	59	0.5736	0.4905	0.1384	0.7627
T1 SRT Median Skill Score	65	0.2404	0.1256	0.1546	0.7854
T2 SRT Median Skill Score	65	0.2649	0.1476	0.1537	0.7842
Test scale				0.1456	0.7817

Table 14.

Principal Components Analysis

Component	Comp1	Comp2	Comp3	Comp4	Comp5
Eigenvalue	2.192	1.619	1.328	1.189	1.136
Proportion	0.199	0.147	0.121	0.108	0.103
Coefficients	Eigenvector 1	Eigenvector 2	Eigenvector 3	Eigenvector 4	Eigenvector 5
T1 AGL d prime	-0.161	0.1426	0.443	0.3726	0.5478
T2 AGL d prime	0.3638	0.0712	-0.4424	-0.1414	-0.0851
T1 Category Learning Total Acc	0.4258	0.0297	-0.0023	0.5058	-0.0611
T2 Category Learning Total Acc	0.4603	-0.1479	0.1044	0.3594	-0.3042
T1 PCT Total Acc	0.2739	0.4598	0.1172	-0.2035	0.3692
T2 PCT Total Acc	0.3707	0.3999	0.0251	0.0837	0.112
T1 SRT Median Skill Score	0.2372	-0.4498	-0.0502	-0.1152	0.4207
T2 SRT Median Skill Score	0.2514	-0.524	0.2816	-0.0362	0.1466
Rotary Pursuit	-0.2533	0.0767	0.2248	0.3891	-0.3442
Mirror tracing time	0.1304	-0.1321	0.5655	-0.3644	-0.2588
Mirror tracing error	0.1885	0.2834	0.3613	-0.3301	-0.2523

Table 15a.

Exploratory Factor Analysis - Factor loadings for one-factor model of implicit learning.

Variable	Factor1	Uniqueness
T2 AGL d prime	0.2385	0.9431
T1 PCT Optimal	0.2041	0.9583
T2 PCT Optimal	0.469	0.7801
T1 Category Learning Total Acc	0.7258	0.4732
T2 Category Learning Total Acc	0.6779	0.5405
T1 SRT Median Skill Score	0.1767	0.9688
T2 SRT Median Skill Score	0.1764	0.9689

Table 15b.

Exploratory Factor Analysis - Factor loadings for multi-factor model of implicit and explicit processes.

-	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	
Eigenvalue	1.6030	1.6460	1.4237	1.0697	1.0737	0.7369	
Proportion	0.2122	0.2179	0.1885	0.1416	0.1422	0.0976	
Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Uniqueness
T2 AGL d prime	0.1942	0.2262	0.0819	0.0309	0.3934	0.0512	0.7461
stdT1Optimal	0.1617	0.0378	-0.1978	0.0772	0.7410	-0.0971	0.3689
stdT2Optimal	0.3432	0.0476	-0.0413	0.0561	0.4145	0.0808	0.6968
T1 Category Learning Total Acc	0.9912	0.1306	0.0141	-0.0164	0.0000	0.0000	0.0000
T2 Category Learning Total Acc	0.5303	0.1573	0.1475	-0.0295	0.0659	0.4961	0.4210
T1 SRT Median Skill Score	0.0039	0.1527	0.7913	0.5920	0.0000	0.0000	0.0000
T2 SRT Median Skill Score	0.1010	0.0243	0.2055	0.4165	-0.0774	0.4807	0.5365
T1 CVLT	0.1222	0.4927	-0.6742	0.5364	0.0000	0.0000	0.0000
T2 CVLT	-0.0186	0.4364	-0.3896	0.4401	0.3437	0.1902	0.3095
Verbal IQ	-0.3591	0.9292	0.0488	-0.0718	0.0000	0.0000	0.0000
Non-Verbal IQ	0.0014	0.4430	0.2775	-0.2189	0.0049	0.3302	0.5697
WM Span	0.0574	0.1802	-0.0045	-0.0096	-0.2638	0.3098	0.7986

Table 15c.

Exploratory Factor Analysis - Factor loadings for single-factor model of implicit and explicit processes

Variable	Factor1	Uniqueness
T2 AGL d prime	0.231	0.9466
T1 PCT Optimal	0.3653	0.8665
T2 PCT Optimal	0.1857	0.9655
T1 Category Learning Total Acc	0.048	0.9977
T2 Category Learning Total Acc	0.1451	0.9789
T1 SRT Median Skill Score	0.0136	0.9998
T2 SRT Median Skill Score	0.1494	0.9777
T1 CVLT	0.7298	0.4674
T2 CVLT	0.9725	0.0541
Verbal IQ	0.3727	0.8611
Non-Verbal IQ	0.0769	0.9941
WM Span	-0.0036	1

Figures

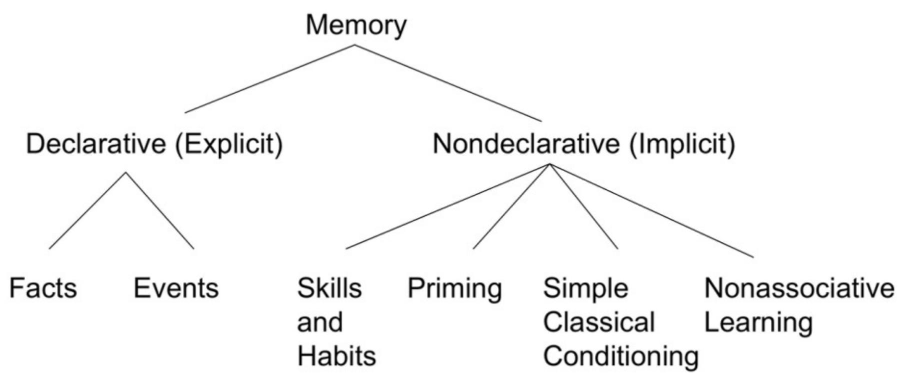

Figure 1. Taxonomy of memory systems from Squire & Zola-Morgan (1991).

Table 1
Comparison of Implicit Memory and Implicit Learning

Feature	Implicit memory (priming)	Implicit learning
Knowledge gained	Verbatim stimulus (sometimes associations)	Novel pattern or rule
Stimuli	Usually verbal	Usually visual or visuospatial
Role of awareness	Not required	Possibly required
Role of attention	Minimal	Important

Figure 2. Table comparing implicit memory and implicit learning from Seger (1994).

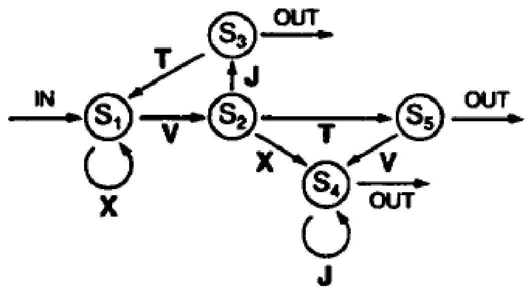

Figure 3. Example Markov chain finite state model used to generate stimuli for an AGL experiment (from Knowlton & Squire, 1996, based on Reber, 1989).

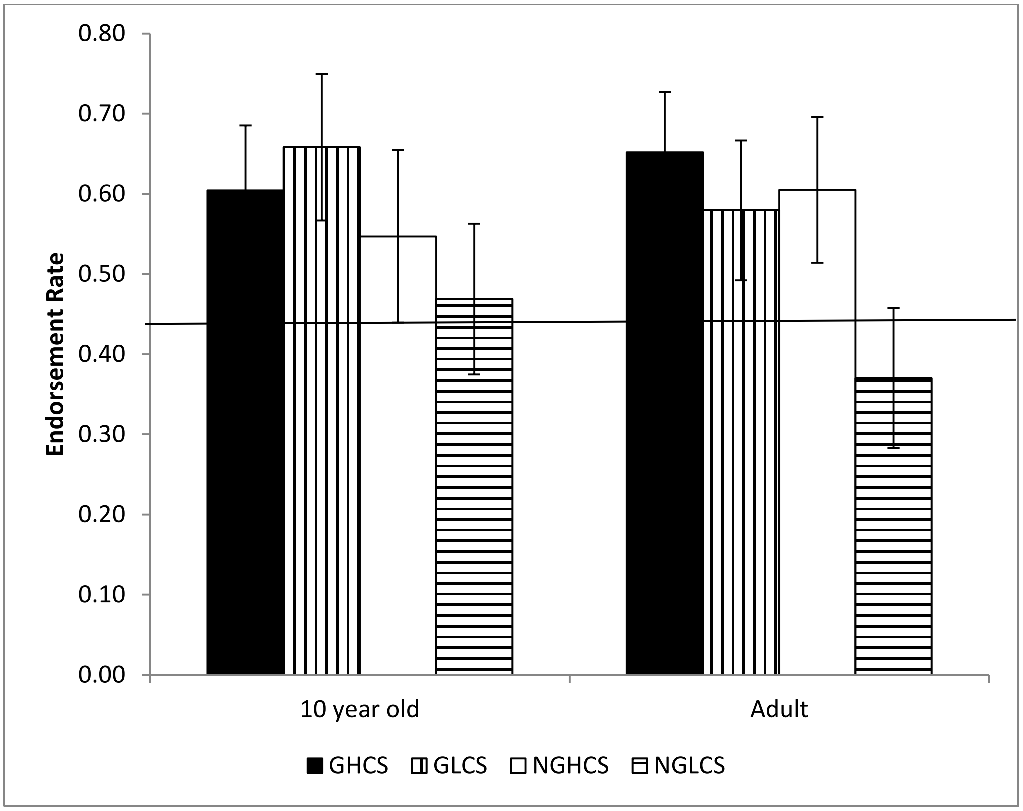

Figure 4. Endorsement Rates for AGL Test Items by Group and Item Type.

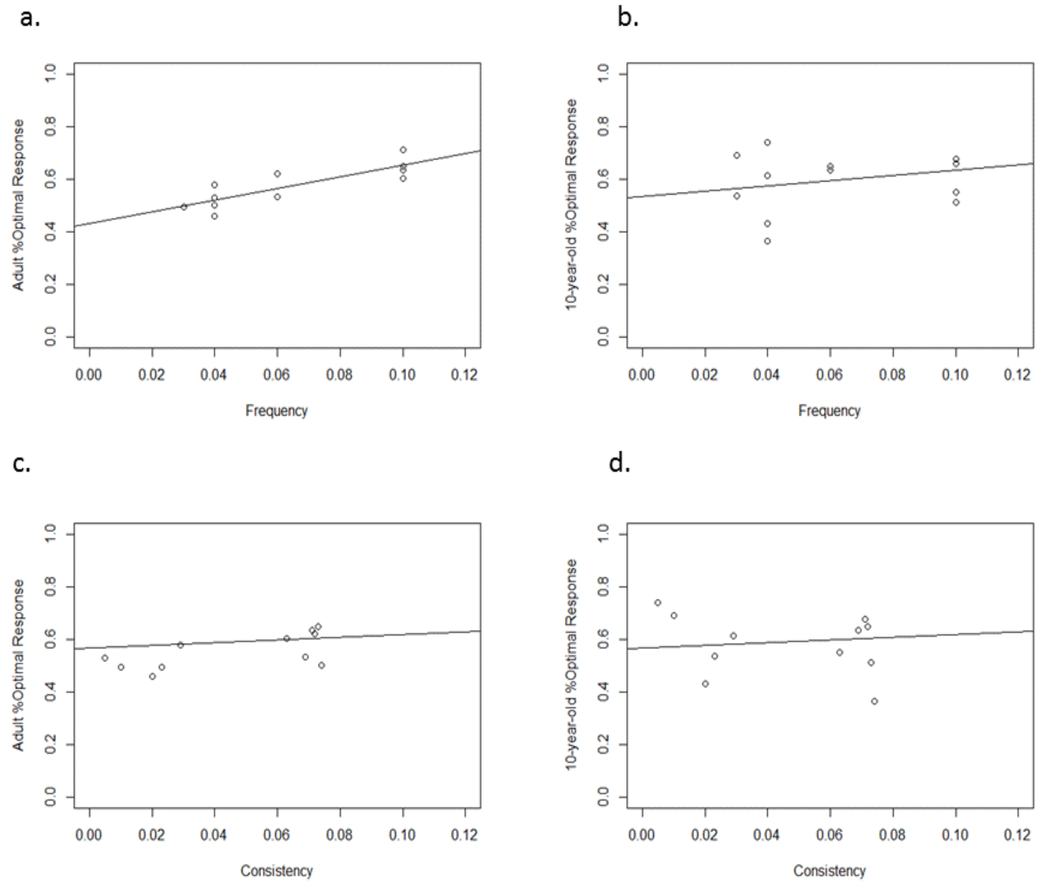

Figure 5. Scatterplots of Optimal Response Rate on PCT Stimulus Features by group. (a.) Adult optimal response rate on stimulus frequency (r = 0.91). (b.) 10-year old optimal response rate on stimulus frequency (r = 0.33). (c.) Adults optimal response rate on stimulus consistency(r = 0.85). (d.) 10-year-old optimal response rate on stimulus consistency (r = 0.26).

Appendix 2: Item Analyses for Developmental Study

Detailed analysis for PCT and AGL. The test stimuli for AGL contained four types of items: grammatical High Chunk Strength (HCS), grammatical Low Chunk Strength (LCS), Non-grammatical High Chunk Strength (NHCS), and Non-grammatical Low Chunk Strength (NLCS). If participants were endorsing test items based on abstract structure, they would have higher endorsement rates for grammatical items regardless of chunk strength. On the other hand, if they were basing their decisions on surface features (Chunk Strength) then endorsement rates would be high for both HCS and NHCS.

In a group-by-item (2x4) type ANOVA, we found a main effect of item type ($F(3, 59) = 6.69$, $p < 0.001$), but no significant main effect of group ($F(1, 59) = 1.40$, $p = 0.24$) or significant interaction($F(4, 59) = 1.4$, $p = 0.24$), although these might have been trending toward significance (Table 5a). Descriptively, the mean endorsement rate across both groups was highest for GHCS (0.66), followed by GLCS (0.65), NGHCS (0.60), then NGLCS (0.47). However, when the endorsement rates were disaggregated by group, two different trends emerged. Children were more likely to endorse grammatical items, regardless of chunk strength, whereas adults endorsed NHCS items almost as frequently as grammatical items (Table 3, Figure 4).

To further explore these differences, we fit several linear repression models to the disaggregated data (Table 5). While chunk strength and stimulus length remain important predictors of adult endorsement rate even in a model that contains them both (Model 8), grammaticality is never a statistically significant predictor of adult endorsement rate. On the other hand, grammaticality is the only statistically significant predictor of 10-year-old endorsement rate and remains so even when length and chunk strength are controlled for (Model 4).

For the PCT items, we analyzed optimal response by cue combination (pattern). We found differences in which items adults and children classified correctly (optimally). A group-by-pattern ANOVA of optimal responses yielded a main effect of pattern ($F(50,22) = 10.52$, $p<0.001$) and an interaction between pattern and group ($F(50, 22) = 3.32$, $p<0.001$), but no main effect of group ($F(50,22) = 1.10$, $p>0.05$). Table 6 displays the children's' and adults' percent optimal response for each pattern. In addition, Table 6 displays the frequency of each pattern (number or percent of appearances in the stimulus set) as well as the percent of that pattern's presentations for which feedback indicated that Outcome 1 was correct. While frequency was strongly correlated with adults' rates of optimal responding ($r = 0.90, p<0.001$; Figure 5a), it was not reliably correlated with children's rates of optimal responding ($r = 0.35, p>0.05$; Figure 5b). It appears that although they were able to produce similar overall performance on implicit learning tasks, children's and adults' responses were driven by different stimulus characteristics (e.g. frequency).

Table 5a.

Group by Item Type Analysis of Variance for AGL Test Items

	Partial SS	*df*	MS	*F*	*p*
Main Effect of Age Group	0.005	1	0.005	0.100	0.756
Main Effect of Item Type	1.243	3	0.414	8.050	0.000
Interaction: Age GroupX Item Type	0.202	3	0.067	1.310	0.274

Table 5b.

Regression Models for Endorsement of AGL Test Items disaggregated by group

	M1	M2	M3	M4	M5	M6	M7	M8
	Kid Endorse	Kid Endorse	Kid Endorse	Kid Endorse	Adult Endorse	Adult Endorse	Adult Endorse	Adult Endorse
Grammatical	0.110*			0.0991*	0.13			0.09
	(2.64)			(2.29)	(1.93)			(1.62)
Chunk Strength		0.03		0.02		0.186**		0.185**
		(0.62)		(0.57)		(2.94)		(3.41)
Length			0.03	0.02				0.0672*
			(1.49)	(1.09)			0.0732*	(2.66)
							(2.43)	
_cons	0.545***	0.582***	0.459***	0.439***	0.509***	0.466***	0.25	0.13
	(18.74)	(16.38)	(4.79)	(4.68)	(11.01)	(9.92)	(1.81)	(1.10)
N	29	29	29	29	29	29	29	29
R-sq	0.205	0.014	0.076	0.249	0.121	0.243	0.18	0.489
adj R-sq	0.175	-0.023	0.042	0.159	0.089	0.215	0.149	0.427

t statistics in parentheses

*p<0.05, **p<0.01, ***p<0.001

Table 5c.

Regression models for endorsement rate for combined groups

	M1	M2	M3	M4
Group	-0.010		-0.010	0.025
	(-0.29)		(-0.31)	-0.380
Item Type		-0.0658***		
		(-4.44)		
Item Type 2: GLCS			0.000	0.000
			(-0.00)	(-0.00)
Item Type 3: NGHCS			-0.048	-0.052
			(-1.04)	(-0.76)
Item Type 4:			-	-0.123

NGLCS			0.203***	
			(-4.37)	(-1.80)
2.group#2.it				0.000
				0.000
2.group#3.it				0.007
				-0.080
2.group#4.it				-0.148
				(-1.59)
Intercept	0.596***	0.755***	0.659***	0.640***
	-23.230	-18.640	-17.640	-13.230
N	192.000	192.000	192.000	192.000
R-sq	0.000	0.094	0.122	0.140

t statistics in parentheses

** p<0.05, **p<0.01, *** p<0.001*

Table 6.

PCT Pattern (cue combination) characteristics and responses by group

pattern	Raw Freq.	% Frequency	Optimal Response	Percent Sunny	Consistency	10 year old %optimal	Adult %optimal
1	510	0.1	0	0.573	0.073	0.512	0.648
2	204	0.04	0	0.495	0.005	0.740	0.530
3	204	0.04	1	0.426	0.074	0.365	0.500
4	510	0.1	1	0.651	0.151	0.658	0.712
5	663	0.13	0	0.659	0.159	0.669	0.720
6	306	0.06	0	0.572	0.072	0.647	0.620
7	153	0.03	1	0.510	0.010	0.692	0.493
8	153	0.03	0	0.523	0.023	0.538	0.493
9	306	0.06	1	0.569	0.069	0.635	0.533
10	663	0.13	1	0.685	0.185	0.737	0.714
11	204	0.04	0	0.480	0.020	0.433	0.460
12	204	0.04	1	0.471	0.029	0.615	0.580
13	510	0.1	1	0.563	0.063	0.550	0.604
14	510	0.1	0	0.571	0.071	0.677	0.636
Total	5100	1	7				
Mean			0.5	0.5	0.55	0.60	0.60

References

Alvarez, J. A. & Emory, E., Julie A.; Emory, Eugene (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review* **16** (1): 17–42.

Aldridge, R. B., Glodzik, D., Ballerini, L., Fisher, R. B., & Rees, J. L. (2011). Utility of Non-rule-based Visual Matching as a Strategy to Allow Novices to Achieve Skin Lesion Diagnosis. Acta Dermato-Venereologica, 91(3), 279–283. Retrieved from http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=63481764&site=ehost-live&scope=site

Amso, D., & Davidow, J. (2012). The development of implicit learning from infancy to adulthood: Item frequencies, relations, and cognitive flexibility. Developmental Psychobiology, 54(6), 664–673.

Arroyo-Anllo, E. M., Ingrand, P., & Gil, R. (2012). Improvement of Semantic Categorization through Procedural Learning in Alzheimer's Disease. JOURNAL OF ALZHEIMERS DISEASE, 30(1), 121–129. doi:10.3233/JAD-2012-111856

Ashby, F. G., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems. Neurobiology of Learning and Memory, 94(1), 1–12. doi:10.1016/j.nlm.2010.03.001

Ashby, F. G., Noble, S., Filoteo, J., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in Parkinson's disease. NEUROPSYCHOLOGY, 17(1), 115–124. doi:10.1037/0894-4105.17.1.115

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: the roles of discrimination and retrieval. Memory & Cognition, 41(3), 392–402. doi:10.3758/s13421-012-0272-7

Bitan, T., & Karni, a. (2003). Alphabetical knowledge from whole words training: effects of explicit instruction and implicit experience on learning script segmentation. Brain Research. Cognitive Brain Research, 16(3), 323–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12706213

Bright, P., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2008). Longitudinal studies of semantic dementia: the relationship between structural and functional changes over time. Neuropsychologia, 46(8), 2177–88. doi:10.1016/j.neuropsychologia.2008.02.019

Brown, R. M., & Robertson, E. M. (2007). Inducing motor skill improvements with a declarative task. Nature Neuroscience, 10(2), 148–9. doi:10.1038/nn1836

Buchner, A., & Wippich, W. (1997). Differences and commonalities between implicit learning and implicit memory. In M. A. Stadler & P. A. Frensch (Eds.), Handbook of Implicit Learning. Thousand Oaks, CA: Sage.

Cheng, M., & Brown, D. E. (2010). Conceptual Resources in Self-developed Explanatory Models: The importance of integrating conscious and intuitive knowledge. International Journal of Science Education, 32(17), 2367–2392. doi:10.1080/09500690903575755

Cheng, M.-F., & Brown, D. E. (2010). Conceptual Resources in Self-developed Explanatory Models: The importance of integrating conscious and intuitive knowledge. INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 32(17), 2367–2392. doi:10.1080/09500690903575755

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science, 210(4466), 207–210. doi:10.1126/science.7414331

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. PSYCHONOMIC BULLETIN & REVIEW, 8(2), 343–350. doi:10.3758/BF03196171

Don, A. J., Schellenberg, E. G., Reber, A. S., DiGirolamo, K. M., & Wang, P. P. (2003). Implicit learning in children and adults with Williams syndrome. DEVELOPMENTAL NEUROPSYCHOLOGY, 23(1-2), 201–225. doi:10.1207/S15326942DN231&2_9

Doyon, J. (1997). Skill Learning. In IThe Cerebellum and Cognition nternational Review of Neurobiology (Vol. 41, pp. 273–294). Elsevier. doi:10.1016/S0074-7742(08)60356-6

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. Annual Review of Psychology, 59(1), 255–278. Retrieved from http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=31159962&site=ehost-live&scope=site

Fawcett, A. J., & Nicolson, R. I. (1990). Automaticity: A new framework for dyslexia research? Cog, 35, 159–182.

Filoteo, J. V., Maddox, W. T., Ing, A. D., & Song, D. D. (2007). Characterizing rule-based category learning deficits in patients with Parkinson's disease. NEUROPSYCHOLOGIA, 45(2), 305–320. doi:10.1016/j.neuropsychologia.2006.06.034

Filoteo, J. V, Maddox, W. T., & Davis, J. (1998). Probabilistic category learning in patients with amnesia, Huntington's disease, or Parkinson's disease: The role of the hippocampus and basal ganglia. JOURNAL OF COGNITIVE NEUROSCIENCE, 10(S), 108.

Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). When It Hurts (and Helps) to Try: The Role of Effort in Language Learning. PloS One, 9(7), e101806. doi:10.1371/journal.pone.0101806

Fischer, J. P. (1997). Artificial grammar learning by 9- to 11-year-old children. Annee Psychologique, 97(2).

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. doi:10.1037//0278-7393.26.1.246

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

Foerde, K., & Shohamy, D. (2011). The role of the basal ganglia in learning and memory: Insight from Parkinson's disease. NEUROBIOLOGY OF LEARNING AND MEMORY, 96(4, SI), 624–636. doi:10.1016/j.nlm.2011.08.006

Forkstam, C., & Petersson, K. M. (2005). Towards an explicit account of implicit learning. Current Opinion in Neurology, 18(4), 435–41. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16003121

Frank, M. C., & Tenenbaum, J. B. (2011). Three ideal observer models for rule learning in simple languages. Cognition, 120(3), 360–71. doi:10.1016/j.cognition.2010.10.005

Gabrieli, J. (1998). Cognitive neuroscience of human memory. Annual Review of Psychology, 49, 87–115. Retrieved from http://www.annualreviews.org/doi/abs/10.1146/annurev.psych.49.1.87

Gabrieli, J. D. E., Stebbins, G. T., Singh, J., Willingham, D. B., & Goetz, C. G. (1997). Intact mirror-tracing and impaired rotary-pursuit skill learning in patients with Huntington's disease: Evidence for dissociable memory systems in skill learning. NEUROPSYCHOLOGY, 11(2), 272–281. doi:10.1037//0894-4105.11.2.272

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 (Cold Spring Harbor, N.Y.), 9(6), 408–18. doi:10.1101/lm.45202

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., … Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 101(21), 8174–8179. doi:10.1073/pnas.0402680101

Gogtay, N., Nugent III, T. F., Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., … Thompson, P. M. (2006). Dynamic mapping of normal human hippocampal development. HIPPOCAMPUS, 16(8), 664–672. doi:10.1002/hipo.20193

Gómez, R., & Gerken, L. (2000). Infant artificial language learning and language acquisition. Trends in Cognitive Sciences, 4(5), 178–186. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10782103

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. Cognitive Psychology, 33(2), 154–207.

Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. Memory & Cognition, 28(2), 253–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10790980

GOMEZ, R. L., & SCHVANEVELDT, R. W. (1994). What is Learned from Artificial Grammars - Transfer Tests of Simple Association. Journal of Experimental Psychology-Learning Memory and Cognition, 20(2), 396–410.

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), 501–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3160813

Hatano, G., & Inagaki, K. (1997). Qualitative changes in intuitive biology. EUROPEAN JOURNAL OF PSYCHOLOGY OF EDUCATION, 12(2), 111–130. Retrieved from http://apps.webofknowledge.com.ezp-prod1.hul.harvard.edu/full_record.do?product=UA&search_mode=MarkedList&qid=22&SID=4DL5y1zshL9SuwoLvJv&page=1&doc=5&colName=WOS

Horan, W. P., Green, M. F., Knowlton, B. J., Wynn, J. K., Mintz, J., & Nuechterlein, K. H. (2008). Impaired implicit learning in schizophrenia. Neuropsychology, 22(5), 606–17. doi:10.1037/a0012602

Ise, E., Arnoldi, C. J., Bartling, J., & Schulte-Koerne, G. (2012). Implicit learning in children with spelling disability: evidence from artificial grammar learning. Journal of Neural Transmission, 119(9).

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: age-related differences in implicit sequence learning across the human lifespan. Developmental Science, 15(4), 496–505.

Janacsek, K., & Nemeth, D. (2012). Predicting the future: From implicit learning to consolidation. INTERNATIONAL JOURNAL OF PSYCHOPHYSIOLOGY, 83(2, SI), 213–221. doi:10.1016/j.ijpsycho.2011.11.012

Knee, R., Thomason, S., Ashe, J., & Willingham, D. T. (2007). The representation of explicit motor sequence knowledge. MEMORY & COGNITION, 35(2), 326–333. doi:10.3758/BF03193453

Knowlton, B. J., Mangels, J. a, & Squire, L. R. (1996). A neostriatal habit learning system in humans. Science (New York, N.Y.), 273(5280), 1399–402. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8703077

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. (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 (Cold Spring Harbor, N.Y.), 1, 106–120. doi:10.1101/lm.1.2.106

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.

Kolen, M. J. (2004). *Test Equating, Scaling, and Linking : Methods and Practices*. (R. L. Brennan, Ed.) (Second Edition.). Springer New York : Imprint: Springer.

Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. Neuroscience and Biobehavioral Reviews, 30(6), 718–29. doi:10.1016/j.neubiorev.2006.06.001

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. Cognitive Psychology, 20(1), 24–37. doi:10.1016/0010-0285(88)90023-0

Lindsey, R. V, Mozer, M. C., Huggins, W. J., & Pashler, H. (2013). Optimizing instructional policies. In Advances in Neural Information Processing Systems (pp. 2778–2786).

Lloyd, M. E., & Newcombe, N. (2009). Implicit memory in childhood: reassessing developmental invariance. In M. Courage; & N. Cowan (Eds.), The development of memory in infancy and childhood (2nd ed). (2nd ed.). New York: Psychology Press.

Lum, J., Kidd, E., Davis, S., & Conti-Ramsden, G. (2010). Longitudinal study of declarative and procedural memory in primary school-aged children. Australian Journal of Psychology, 62(3), 139–148.

Maddox, W. T., Aparicio, P., Marchant, N. L., & Ivry, R. B. (2005). Rule-based category learning is impaired in patients with Parkinson's disease but not in patients with cerebellar disorders. JOURNAL OF COGNITIVE NEUROSCIENCE, 17(5), 707–723. doi:10.1162/0898929053747630

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. doi:10.1037/0278-7393.31.1.100

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: the independent contributions of conditional probability and pair frequency information. DEVELOPMENTAL SCIENCE, 12(6), 1020–1025. doi:10.1111/j.1467-7687.2009.00838.x

MAYBERY, M., TAYLOR, M., & OBRIENMALONE, A. (1995). Implicit Learning - Sensitive to Age but Not IQ. Australian Journal of Psychology, 47(1), 8–17.

McClary, L., & Talanquer, V. (2011). Heuristic Reasoning in Chemistry: Making decisions about acid strength. International Journal of Science Education, 33(10), 1433–1454. Retrieved from http://ezp-prod1.hul.harvard.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=60977198&site=ehost-live&scope=site

Meeter, M., Radicsa, 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. doi:10.1016/j.neubiorev.2007.11.001

Merbah, S., & Meulemans, T. (n.d.). LEARNING A MOTOR SKILL: EFFECTS OF BLOCKED VERSUS RANDOM PRACTICE A REVIEW. PSYCHOLOGICA BELGICA, 51(1), 15–48. Retrieved from http://apps.webofknowledge.com.ezp-prod1.hul.harvard.edu/full_record.do?product=UA&search_mode=Refine&qid=8&SID=2AEayGwHqRzw3IxDkRq&page=4&doc=38

Meulemans, T., der Linden, M. Van, & Perruchet, P. (1998). Implicit sequence learning in children. Journal of Experimental Child Psychology, 69(3), 199–221.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16(5), 519–533. doi:10.1016/S0022-5371(77)80016-9

Nicolson, R. I., Fawcett, A. J., Brookes, R. L., & Needle, J. (2010). Procedural learning and dyslexia. Dyslexia, 16(3), 194–212. doi:10.1002/dys.408

Nishio, Y., & Mori, E. (2009). [Semantic dementia--a multimodal disorder of conceptual knowledge]. *Brain and Nerve = Shinkei Kenkyū No Shinpo*, 61(11), 1236–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19938680

Nissen, M. J., & Bullemer, P. (1987). Attentional Requirements of Learning : Evidence from Performance Measures. COGNITIVE PSYCHOLOGY, 19.

Noh, S. M., Yan, V. X., Vendetti, M. S., Castel, A. D., & Bjork, R. A. (2014). Multilevel Induction of Categories: Venomous Snakes Hijack the Learning of Lower Category Levels. Psychological Science, 25(8), 1592–1599. doi:10.1177/0956797614535938

Ohst, A., Fondu, B. M. E., Glogger, I., Nückles, M., & Renkl, A. (2014). Preparing learners with partly incorrect intuitive prior knowledge for learning. Frontiers in Psychology, 5(July), 664. doi:10.3389/fpsyg.2014.00664

Ortu, D., & Vaidya, M. (2013). A neurobiology of learning beyond the declarative non-declarative distinction. Frontiers in Behavioral Neuroscience, 7, 161. doi:10.3389/fnbeh.2013.00161

Ouellet, M.-C., Beauchamp, M. H., Owen, A. M., & Doyon, J. (2004). Acquiring a cognitive skill with a new repeating version of the Tower of London task. Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale, 58(4), 272–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15648730

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. Annual Review of Neuroscience, 25, 563–93. doi:10.1146/annurev.neuro.25.112701.142937

Pavlidou, E. V, Kelly, M. L., & Williams, J. M. (2010). Do Children with Developmental Dyslexia Have Impairments in Implicit Learning? DYSLEXIA, 16(2), 143–161. doi:10.1002/dys.400

Pavlidou, E. V, Williams, J. M., & Kelly, L. M. (2009). Artificial grammar learning in primary school children with and without developmental dyslexia. Annals of Dyslexia, 59(1).

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. Trends in Cognitive Sciences, 10(5), 233–8. doi:10.1016/j.tics.2006.03.006

Poldrack, R. a. (2006). Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences, 10(2), 59–63. doi:10.1016/j.tics.2005.12.004

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. doi:10.1037//0894-4105.13.4.564

Pothos, E. M. (2007). Theories of artificial grammar learning. Psychological Bulletin, 133(2), 227–44. doi:10.1037/0033-2909.133.2.227

Price, A., Filoteo, J. V., & Maddox, W. T. (2009). Rule-based category learning in patients with Parkinson's disease. NEUROPSYCHOLOGIA, 47(5), 1213–1226. doi:10.1016/j.neuropsychologia.2009.01.031

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. doi:10.3758/MC.37.2.210

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. Trends in Cognitive Sciences, 6(10), 416–421. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12691761

Raznahan, A., Shaw, P., Lalonde, F., Stockman, M., Wallace, G. L., Greenstein, D., … Giedd, J. N. (2011). How Does Your Cortex Grow? JOURNAL OF NEUROSCIENCE, 31(19), 7174–7177. doi:10.1523/JNEUROSCI.0054-11.2011

REBER, A. S. (1989). IMPLICIT LEARNING AND TACIT KNOWLEDGE. JOURNAL OF EXPERIMENTAL PSYCHOLOGY-GENERAL, 118(3), 219–235. doi:10.1037//0096-3445.118.3.219

Reber, P. J. (2008). Cognitive Neuroscience of Declarative and Nondeclarative Memory. In T. Benjamin, AS and DeBelle, JS and Etnyre, B and Polk (Ed.), HUMAN LEARNING: BIOLOGY, BRAIN, AND NEUROSCIENCE (Vol. 139, pp. 113–123).

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

Reber, P. J., Beeman, M., & Paller, K. A. (2013). Human Memory Systems : A Framework for Understanding the Neurocognitive Foundations of Intuition, 474–483.

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

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(3), 585–594. doi:10.1037//0278-7393.20.3.585

Rendell, M. A., Farrow, D., Masters, R., & Plummer, N. (2011). Implicit Practice for Technique Adaptation in Expert Performers. INTERNATIONAL JOURNAL OF SPORTS SCIENCE & COACHING, 6(4), 553–566. doi:10.1260/1747-9541.6.4.553

Rosas, R., Ceric, F., Tenorio, M., Mourgues, C., Thibaut, C., Hurtado, E., & Aravena, M. T. (2010). ADHD children outperform normal children in an artificial grammar implicit learning task: ERP and RT evidence. Consciousness and Cognition, 19(1).

Ruenger, D. (2012). How sequence learning creates explicit knowledge: the role of response-stimulus interval. PSYCHOLOGICAL RESEARCH-PSYCHOLOGISCHE FORSCHUNG, 76(5), 579–590. doi:10.1007/s00426-011-0367-y

Ruenger, D., & Frensch, P. A. (2008). How incidental sequence learning creates reportable knowledge: The role of unexpected events. JOURNAL OF EXPERIMENTAL PSYCHOLOGY-LEARNING MEMORY AND COGNITION, 34(5), 1011–1026. doi:10.1037/a0012942

Ruitenberg, M. F. L., Abrahamse, E. L., & Verwey, W. B. (2013). Sequential motor skill in preadolescent children: the development of automaticity. Journal of Experimental Child Psychology, 115(4), 607–23. doi:10.1016/j.jecp.2013.04.005

Sætrevik, B., Reber, R., & Sannum, P. (2006). The utility of implicit learning in the teaching of rules. Learning and Instruction, 16(4), 363–373. doi:10.1016/j.learninstruc.2006.07.006

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. Psychological Review, 120(2), 411–37. doi:10.1037/a0031912

Sanchez, D. J., Gobel, E. W., & Reber, P. J. (2010). Performing the unexplainable: Implicit task performance reveals individually reliable sequence learning without explicit knowledge. PSYCHONOMIC BULLETIN & REVIEW, 17(6), 790–796. doi:10.3758/PBR.17.6.790

Sandhofer, C. M., & Doumas, L. A. A. (2008). Order of presentation effects in learning color categories. JOURNAL OF COGNITION AND DEVELOPMENT, 9(2), 194–221. doi:10.1080/15248370802022639

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. doi:10.1016/S0896-6273(03)00123-5

SEGER, C. A. (1994). CRITERIA FOR IMPLICIT LEARNING - DE-EMPHASIZE CONSCIOUS ACCESS, EMPHASIZE AMNESIA. BEHAVIORAL AND BRAIN SCIENCES, 17(3), 421–422.

Seger, C. A. (1994). Implicit learning. Psychological Bulletin, 115(2), 163–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19283555

Seger, C. A. (2006). The basal ganglia in human learning. *The Neuroscientist : A Review* Journal Bringing Neurobiology, Neurology and Psychiatry, 12(4), 285–90. doi:10.1177/1073858405285632

Seger, C. A., & Miller, E. K. (2010). Category Learning in the Brain. In Hyman, SE (Ed.), ANNUAL REVIEW OF NEUROSCIENCE, VOL 33 (Vol. 33, pp. 203–219). doi:10.1146/annurev.neuro.051508.135546

Seger, C. A., & Peterson, E. J. (2013). Categorization = decision making plus generalization. NEUROSCIENCE AND BIOBEHAVIORAL REVIEWS, 37(7), 1187–1200. doi:10.1016/j.neubiorev.2013.03.015

Seger, C. A., & Spiering, B. J. (2011). A critical review of habit learning and the Basal Ganglia. Frontiers in Systems Neuroscience, 5(August), 66. doi:10.3389/fnsys.2011.00066

Seidenberg, M. S., Macdonald, M. C., & Saffran, J. R. (2002). Does Grammar Start Where Statistics Stop ?, (October), 553–554.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., … Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. JOURNAL OF NEUROSCIENCE, 28(14), 3586–3594. doi:10.1523/JNEUROSCI.5309-07.2008

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. doi:10.1016/j.brainres.200S.10.025

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.

Song, S., Howard, J. H., & Howard, D. V. (2007). Implicit probabilistic sequence learning is independent of explicit awareness. Learning & Memory (Cold Spring Harbor, N.Y.), 14(3), 167–76. doi:10.1101/lm.437407

Sowell, E. R., Trauner, D. a, Gamst, A., & Jernigan, T. L. (2002). Development of cortical and subcortical brain structures in childhood and adolescence: a structural MRI study. Developmental Medicine and Child Neurology, 44(1), 4–16. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11811649

Speekenbrink, M., Lagnado, D. A., Wilkinson, L., Jahanshahi, M., & Shanks, D. R. (2010). Models of probabilistic category learning in Parkinson's disease: Strategy use and the effects of L-dopa. JOURNAL OF MATHEMATICAL PSYCHOLOGY, 54(1, SI), 123–136. doi:10.1016/j.jmp.2009.07.004

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. Neurobiology of Learning and Memory, 82(3), 171–7. doi:10.1016/j.nlm.2004.06.005

Squire, L. R., & Zola-Morgan, S. (1988). Memory: brain systems and behavior. Treands in Neuroscience, 11, 170–175.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. Science, 253, 1380–1386.

Star, J. R. (2005). Reconceptualizing Procedural Knowledge. Journal for Research in Mathematics Education, 36(5), 404–411.

Star, J. R. (2007). Research Commentary : Foregrounding Procedural Knowledge. Journal for Research in Mathematics Education, 38(2), 132–135.

Stark, C. E. L., Reber, P. J., & Squire, L. R. (1998). Contrasting signatures of declarative and nondeclarative memory: an fMRI study. JOURNAL OF COGNITIVE NEUROSCIENCE, 10(S), 47.

Stefaniak, N., Willems, S., Adam, S., & Meulemans, T. (2008). What is the impact of the explicit knowledge of sequence regularities on both deterministic and probabilistic serial reaction time task performance? Memory & Cognition, 36(7), 1283–1298.

Stickgold, R., Malia, A., Maguire, D., Roddenberry, D., & O'Connor, M. (2000). Replaying the game: Hypnagogic images in normals and amnesics. SCIENCE, 290(5490), 350–353. doi:10.1126/science.290.5490.350

Taber, K. S., & Tan, K. C. D. (2011). The Insidious Nature of "Hard-Core" Alternative Conceptions: Implications for the constructivist research programme of patterns in high school students' and pre-service teachers' thinking about ionisation energy. International Journal of Science Education, 33(2), 259–297. doi:10.1080/09500691003709880

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. Science (New York, N.Y.), 331(6022), 1279–85. doi:10.1126/science.1192788

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & 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), 1339–51. doi:10.1162/0898929042304688

Thomas, K. M., & Nelson, C. a. (2001). Serial reaction time learning in preschool- and school-age children. Journal of Experimental Child Psychology, 79(4), 364–87. doi:10.1006/jecp.2000.2613

Thompson, P. M., Sowell, E. R., Gogtay, N., Giedd, J. N., Vidal, C. N., Hayashi, K. M., … Toga, A. W. (2005). Structural MRI and brain development. In NEUROIMAGING, PT B (Vol. 67, p. 285+). doi:10.1016/S0074-7742(05)67009-2

Toga, A. W., Thompson, P. M., & Sowell, E. R. (2006). Mapping brain maturation. Trends in Neurosciences, 29(3), 148–59. doi:10.1016/j.tins.2006.01.007

Vlach, H. A., Sandhofer, C. M., & Bjork, R. A. (2014). Equal spacing and expanding schedules in children’s categorization and generalization. Journal of Experimental Child Psychology, 123, 129–37. doi:10.1016/j.jecp.2014.01.004

Willingham, D. B., & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. Psychological Science, 10(6), 531–536.

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. doi:10.1037//0278-7393.15.6.1047

Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. JOURNAL OF NEUROPHYSIOLOGY, 88(3), 1451–1460. doi:10.1152/jn.00461.2001

Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. Journal of Neurophysiology, 88(3), 1451–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12205165

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

Wulf, G., Shea, C., & Lewthwaite, R. (2010). Motor skill learning and performance: a review of influential factors. Medical Education, 44(1), 75–84. doi:10.1111/j.1365-2923.2009.03421.x

Vita

Priya Kalra

1999-2004	University of California, Berkeley Berkeley, CA	B.A. June 2004
2004-06	Research Assistant, Memory & Aging Center, UCSF San Francisco, CA	
2006-07	Brunel University London, United Kingdom	M.Sc. September 2007
2007-08	Graduate School of Education Harvard University	M.Ed. June 2008
2008-14	Doctor of Education Candidate Graduate School of Education Harvard University	
2008-13	Graduate Student Researcher Gabrieli Laboratory Massachusetts Institute of Technology	
2010-13	Teaching Fellow Graduate School of Education Harvard University	
	Teaching Fellow Faculty of Arts and Sciences Harvard University	
2014	Statistics Workshop Leader University of California, Berkeley	
2015	Visiting Lecturer University of California, Davis	