Developing Production Skills through Implicit Learning

Giulia Bovolenta

Jesus College

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

September 2019

Declaration

I hereby declare that this thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as stated in the Declaration and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This dissertation does not exceed the regulation length of 80,000 words including footnotes, references and appendices but excluding the bibliography.

Giulia Bovolenta
September 2019

Dedicated to Mira,

who once convinced me to do a Masters degree.

Acknowledgements

My time at Cambridge has been a very happy one, and this was in no small measure thanks to the people I met during my time in the Department (now Section) of Theoretical and Applied Linguistics. For four years, they have helped me in many ways great and small, and made me feel part of a community. I would like to express my gratitude to my supervisor, John N. Williams, for giving me the freedom to explore my ideas, but always being there to help me when I needed it. I would like to thank Dimitris Alikaniotis, Connor Quinn and Carla Pastorino Campos, who welcomed me as a friend on my very first day in the PhD room; Elaine Schmidt and Ana I. Peréz, for their constant help and friendship, and all the other fellow PhD students and postdocs with whom I shared life in the department: Andrew Caines, Cherry Lam, Chris Mengying Xia, Elspeth Wilson, Roberto Sileo, Alim Tusun, Julia Heine, Costanza Conforti, Xiaobin Chen. I would especially like to thank Jane Walsh, Napoleon Katsos, Prof. Ianthi Tsimpli, Dora Alexopoulou and my advisor, Brechtje Post, for their guidance, advice and friendship. I should also express my gratitude to E. Matthew Husband: the mentorship and teaching he provided during my time as his student continue to prove invaluable. I gratefully acknowledge financial support from the Economic and Social Sciences Research Council, which funded the first two years of this doctorate. Finally, I would never have been in the position to do any doctorate if it weren't for the loving support of my family over all these years, both moral and material: my husband Mira, my parents Mirella and Germano, my brother Ennio and his partner Valentina, to whom I give my heartfelt thanks.

Abstract

Language proficiency largely relies on implicit knowledge, which is unconscious and operates independently of voluntary control. Implicit learning is a process of incidental learning which results in the acquisition of implicit knowledge. We know that adult learners can acquire knowledge of novel L2 linguistic rules through implicit learning, as evidenced by their performance on receptive tasks. However, it is unclear whether implicit learning processes can also support the development of L2 production skills. The central question of this dissertation was whether it would be possible for learners to acquire implicit knowledge of a new rule through implicit learning and use it directly in spoken production. Our second question concerns the relationship between production and comprehension: we asked whether implicit knowledge acquired through a production task would also lead to improved performance in comprehension. To address these questions, we trained participants on a semiartificial language based on a rule naturally found in Czech: specifically, the usage rule for a pair of spatial prepositions (*v* and *na*) which alternate depending on the distinction between open and enclosed spaces. Training was carried out using a novel methodology based on elicited oral imitation, which was also used to test productive knowledge. Participants were also tested on comprehension, using both reaction time and recognition memory paradigms. Our findings suggest that it is possible to acquire implicit productive knowledge through a production-based task, and to generalise it to new instances in spoken production. The results of our experiments also show that learning outcomes were sensitive to the specific procedure used to train participants, which appeared to interact with individual differences in working memory. Finally, we found limited evidence that implicit knowledge acquired through production could be transferred to comprehension, supporting a skill-specific account of implicit knowledge.

Contents

List of Figures

Chapter 1

Introduction

When performing many daily tasks, we rely on finely tuned knowledge which we can use effectively, even though we may not able to talk about it. Knowing how to ride a bicycle, for instance, does not mean we can describe the laws of physics which allow the bike to remain upright when moving forwards, or state the amount of force we need to exert on the bike for that to happen – we just know how to pedal without falling. Our knowledge of it is implicit: we have it but we are not aware of it, and we cannot verbalise it. Language is another skill which largely relies on implicit knowledge: in our native language, we usually have strong intuitions about what is correct and what is not, even though we may not be able to explain why. First language (L1) acquisition – including the development of production skills – takes place in a largely implicit way: children acquire the structure of their L1 with no explicit instruction and little feedback (Ambridge et al., 2013; Jackendoff, 2002; Kulinich et al., 2019; Wonnacott et al., 2012). By contrast, adults who learn a second language (L2) often do so in a formally instructed context, by committing vocabulary items to memory and studying grammatical rules, usually followed by deliberate practice through various kinds of activities. The knowledge initially developed by L2 learners in these settings is predominantly explicit: conscious knowledge of facts and rules, which learners can access and use by making a deliberate effort. Even in more naturalistic settings, adults bring a level of metalinguistic awareness to language learning which is lacking in infants, such as the concepts of word and word meaning, which shapes their goals and expectations. They make a conscious effort to learn the language, which is more likely to result in explicit knowledge.

Of course, L1 speakers tend to have explicit knowledge of their language, too (e.g. due to education). Over-reliance on explicit knowledge, however, limits L2 attainment. While there is debate on the exact nature of the relationship between implicit and explicit knowledge, it is generally agreed that the development of implicit knowledge is necessary for the attainment of L2 proficiency (Krashen, 1981b, 1982; N. Ellis, 2005, 2011; DeKeyser, 2003; Hulstijn, 2002, 2005; R. Ellis, 2006c, 2012). Implicit knowledge, or at least, highly automatised explicit knowledge (DeKeyser, 2003; Li and DeKeyser, 2017) is the basis for fluent, automatic L2 processing in both comprehension and production. The question addressed in this thesis if whether implicit knowledge - specifically, knowledge that can be used in production - can be acquired directly, or whether learners must first go through a stage of explicit knowledge to be able to draw on it in production. The development of implicit knowledge, and the conditions which allow for it, are the focus of implicit learning (IL) research, a field of study originating in cognitive psychology. Initially developed with a focus on sequence learning using artificial stimuli (Reber, 1967), implicit learning paradigms are increasingly being used in the study of L2 acquisition, to investigate the extent to which we can directly acquire implicit knowledge of novel linguistic rules. The application of these paradigms to second language acquisition (SLA) research has revealed that we can implicitly acquire knowledge of structural aspects such as word segmentation (Saffran et al., 1996a), orthography (Pacton et al., 2001), phonotactics (Chambers et al., 2003), and syntax (Francis et al., 2009; Rebuschat and Williams, 2012), as well as novel form-meaning connections and their usage (Leung and Williams, 2011; Paciorek and Williams, 2015; Williams, 2005).

However, this strand of research tends to rely on comprehension tasks as measures of implicit knowledge: there is limited evidence that it is possible to acquire productive language skills through implicit learning, too. Being able to use linguistic rules in production is a fundamental skill in L2, which, like comprehension, is reliant on implicit knowledge to be fully developed. Therefore, the main goal of this dissertation is to investigate the following question: whether it may be possible to acquire implicit knowledge which can be directly used in spoken production, even while remaining unaware of it. This has clear pedagogical implications, but also touches on the nature of the relationship between production and comprehension in implicit knowledge. In implicit learning research, it is common to conclude that subjects have acquired implicit knowledge if they show evidence of discrimination in comprehension. The assumption underlying this claim is that, even though the knowledge is acquired in one domain (comprehension), it is not itself domain-specific. However, evidence from research comparing production and comprehension in L2 learning suggests that implicit knowledge may be limited to the domain in which it is acquired (DeKeyser and Sokalski, 1996; Li and DeKeyser, 2017). Therefore, the second goal of this dissertation is to investigate the relationship between production and comprehension in implicit learning, by testing whether implicit knowledge developed through a production task may also transfer to comprehension.

The contents of this dissertation are structured as follows:

- Chapter 2 provides an overview of the literature on implicit knowledge and implicit learning in the context of SLA research. It then focuses on the more concrete aspects of implicit learning, reviewing the literature on implicit and explicit learning in L2 instruction and the role of attention in the process. The final section focuses on the role of production in SLA and its interaction with implicit knowledge, in relation to both training and testing.

- Chapter 3 introduces the rule on which we based the materials for our study, derived from a pair of spatial prepositions found in Czech, and provides a rationale for the methodology we used, which was based on elicited oral recall. It then reports the findings of Experiment 1, in which we began to investigate elicited recall as a training and testing procedure for implicit productive knowledge.

- Chapter 4 reports the findings from Experiments 2 and 3, in which we introduced a new version of the testing methodology used in Experiment 1: this involved long-term recall and an element of generalisation, requiring participants to produce novel instances of the rule. We also developed the training methodology, testing different ways to induce implicit learning by either loading memory capacity (Experiment 2) or by directing attention to relevant aspects of the stimuli, by means of questions (Experiment 3).

- Chapter 6 explores the role of working memory and consolidation in the learning process. It reports the results of Experiment 4, in which we built on the procedure used in Experiment 2 by adding delayed testing after overnight consolidation, as well as measures of working memory to investigate its interaction with our learning paradigm.
- Chapter 7 carries out a comparative analysis of Experiments 2, 3 and 4, providing an account of their findings based on the training paradigm used and focusing on ways in which specific features of the training procedure could have influenced learning outcomes.
- Chapter 8 presents a general discussion of our findings, in relation to our research questions and to the existing literature.

Chapter 2

Implicit learning and production in L2 acquisition

2.1 Implicit and explicit knowledge

In cognitive psychology, the distinction between *explicit* and *implicit* knowledge in its basic form hinges on whether it is knowledge that we have conscious mental access to (Williams, 2009). Explicit knowledge is knowledge that we know that we know (Dienes and Perner, 1999) and can therefore use deliberately; implicit knowledge, by contrast, can guide our behaviour even if we are not aware that we have it (Cleeremans et al., 1998). In the field of SLA, the issue was first brought to the fore by Krashen's Monitor Theory (Krashen, 1981a, 1985), which famously drew a distinction between explicit *learning* and implicit *acquisition* in L2: learning is the process by which vocabulary and grammatical rules are consciously committed to memory and rehearsed, while acquisition is the subconscious internalisation of those rules, resulting in increased competence and fluency. In production, utterances are initiated by the implicit system, but formal explicit knowledge may be used to alter or correct these utterances by the conscious system, or *Monitor* (Krashen, 1981b, 1985). Learning is only available to the the conscious part (the *Monitor*), implicit knowledge can only be developed through acquisition (Krashen, 1982). In order for acquisition to happen, the learner needs to be exposed to and process input that is slightly more complex than their current competence level (the *Input hypothesis* (Krashen, 1994)). Krashen's distinction was initially criticised as loosely defined and unfalsifiable (McLaughlin, 1978); it does, however, come quite close to the psychological definition in including consciousness as a criterion. In SLA, the distinction between implicit and explicit knowledge was popularised again by N. Ellis (1994), who highlighted its importance in language processing and acquisition. The second hallmark of implicit knowledge, after a lack of conscious access, is lack of voluntary control: implicit knowledge guides behaviour without intention, while explicit knowledge can be used deliberately (Cleeremans and Jiménez, 2002; Dienes and Perner, 1999). Lack of voluntary control, however, does not necessarily equate with automaticity. According to graded theories of awareness (Cleeremans and Jiménez, 2002), in order to have voluntary control of a representation, it is necessary to be aware of it; implicit representations are outside of voluntary control because they are not accessible to conscious awareness. However, it is also possible for representations to be outside of voluntary control despite awareness, because their strength and stability is such that they guide behaviour automatically (Cleeremans and Jiménez, 2002): automatic processes can be "ballistic", in that they cannot stopped once initiated (Segalowitz, 2003). Implicit representations, on the other hand, may also be quite weak - too weak to enter conscious awareness, but still strong enough to influence behaviour.

There is evidence that the distinction between implicit and explicit knowledge is at least partly rooted in neuroanatomy, with implicit and explicit knowledge being subserved by different memory systems, the *declarative* and *procedural* systems (Paradis, 1994, 2004). The *declarative* system underlies knowledge of facts and events (knowledge "that"); declarative knowledge is usually explicit and can be consciously recollected (Paradis, 2004; Ullman, 2001). Declarative memory relies on the hippocampus, medial temporal lobes, and large portions of neocortex in the temporal and parietal lobes (Paradis, 1994; Suzuki and Amaral, 1994). The *procedural* system underlies the learning of new cognitive and motor skills (knowledge "how"), especially involving sequences, as well as their control once they have been established (Ullman, 2005). Procedural memory is subserved by a number of brain structures, which include the cerebellum, striatum and other basal ganglia, and parts of the left perisylvian cortical region (Paradis, 2004).

According to Paradis (1994, 2004), explicit, metalinguistic knowledge of language is subserved by the declarative system; implicit knowledge of language structure ("linguistic competence" in Paradis' terms), such as morphosyntax, is subserved by the procedural system (Paradis, 2004). Knowledge of lexical meaning, unlike structural knowledge, is explicit and rooted in the declarative system (Paradis, 1994). The Declarative/Procedural (DP) model of language (Ullman, 2001, 2005) emphasizes this distinction, hypothesizing that the processing of different aspects of language (lexicon and grammar) in native speakers is subserved by different systems. According to the DP model, vocabulary is mostly stored in declarative memory, while grammar, which requires the application of rules, is supported by procedural memory. While Paradis (1994, 2004) mostly identifies explicit knowledge with declarative memory and implicit knowledge with procedural memory, the DP model does not assume that all implicit knowledge is procedural knowledge: the procedural memory system is one among several system which underlie implicit representations (Williams, 2009).

The DP model also makes specific predictions with regards to second language acquisition. Evidence suggests the dimension in which late learners of an L2 are most lacking, compared to L1 acquisition, is grammar, while lexical processing is relatively spared (DeKeyser, 2000; Hahne and Friederici, 2001; Wartenburger et al., 2003; Weber-Fox and Neville, 1996). According to Ullman, this is because adults, due to "age-related attenuation of the procedural system" (Ullman, 2005, p. 152), tend to rely on the declarative system more than the procedural for both vocabulary and rules, particularly in a pedagogical setting. The use of the declarative system alone can lead to fairly high proficiency, but not to native-like processing. However, it is expected that practice will lead to procedural learning and increasingly native-like reliance on the procedural system (Ullman, 2001, 2005), a prediction which is supported by experimental studies on artificial language learning showing that, even with an implicit learning procedure, the learning progress is associated with declarative memory performance in the early stages of acquisition, and with procedural memory in the later stages (Hamrick, 2015; Morgan-Short et al., 2014; Ruiz et al., 2018).

2.1.1 Interface between implicit and explicit knowledge

The exact nature of the relation between implicit and explicit knowledge is the subject of ongoing debate. Theoretical positions on the matter range from "no interface" positions (Hulstijn, 2002, 2015; Krashen, 1981a, 1985; Paradis, 1994, 2004), which maintain that the two types of knowledge are completely separate from each other, through "weak interface" ones (Doughty and Williams, 1998; N. Ellis, 2005, 2015), which argue for separate but co-operative types of knowledge, to "strong interface" positions, according to which one kind of representation can be converted into the other (DeKeyser, 2003, 1997; Suzuki and DeKeyser, 2017).

According to the "no interface" position (Hulstijn, 2002, 2015; Krashen, 1981a, 1985; Paradis, 1994, 2004) there is no interaction between the two types of knowledge: explicit knowledge cannot be transformed into implicit knowledge through practice. In Krashen's view, learning (explicit) and acquisition (implicit) are entirely separate processes; it is also maintained that, from a neurocognitive perspective, it is not possible for representations stored in the declarative system to physically "transform" themselves in representations stored in different neural structures (Hulstijn, 2002, 2015).

In the "weak interface" view (Doughty and Williams, 1998; N. Ellis, 2005, 2015), explicit and implicit knowledge are dissociable and reliant on different neural networks, but also cooperative. There is a "dynamic interface" between the two: for instance, explicit knowledge of form-meaning connections plays a role in implicit learning, where explicit constructions from analogy and mnemonics can feed into proceduralisation and implicit learning (N. Ellis, 2005, 2015). Nick Ellis notes that "many aspects of a second language are unlearnable – or at best acquired very slowly – from implicit processes alone (N. Ellis, 2005, p. 307). While explicit knowledge, similarly to the "no interface" hypothesis, cannot be transformed into implicit representations, it can provide the conditions necessary for implicit knowledge to develop: "slot-and-frame patterns, drills, mnemonics, and declarative statements of pedagogical grammar [...] all contribute to the conscious creation of utterances that then partake in subsequent implicit learning and proceduralisation" (N. Ellis, 2005, p. 308).

Finally, according to the "strong interface" hypothesis, explicit knowledge can be converted into implicit knowledge, by means of gradual speeding up and automation (DeKeyser, 2003, 1997; Suzuki and DeKeyser, 2017). This process is often referred to as *proceduralisation*, following Anderson's ACT Theory of skill acquisition (Anderson, 1996, 1983; Anderson and Lebiere, 1998). In ACT theory, skill acquisition involves transitioning from a stage that is characterised by declarative knowledge, to one reliant on procedural knowledge: initially, a task is executed by retrieving the relevant declarative knowledge, which is used to apply a rule. With practice, as the same rule is applied repeatedly and consistently, its application becomes increasingly rapid and automatic: automatic processes are the result of "consistent mapping of the same input on the same pattern of activation over many trials" (McLaughlin, 1987, p. 34). This process of automatisation is thought to reflect a shift from declarative to procedural memory (DeKeyser, 2007). In order for proceduralisation to be possible, the right conditions need to be in place: the declarative knowledge required as well as appropriate tasks that require that knowledge to be employed (DeKeyser, 2007). Proponents of the strong interface acknowledge that implicit knowledge and automatised explicit knowledge are distinct constructs, because the former involves lack of awareness, while the latter does not (Suzuki and DeKeyser, 2015, 2017). However, they argue that automatised explicit knowledge is "functionally equivalent" to implicit knowledge, in that both involve rapid access to linguistic knowledge (DeKeyser, 2003; Suzuki and DeKeyser, 2017).

2.2 Implicit and explicit learning

The distinction between "implicit" and "explicit" knowledge can be extended to the learning process by which that knowledge is acquired, by distinguishing between explicit and implicit learning. *Explicit learning* is a conscious, usually intentional process, in which learners make and test hypotheses about structure (N. Ellis, 1994, 2015) and engage in a "process of concept formation and concept linking" (Hulstijn, 2002, p. 206). By contrast *implicit learning*, as the term is used in cognitive psychology, is learning that happens without intention to learn, and which results in implicit knowledge (Williams, 2009). The term "implicit learning" was first used by Arthur Reber in his seminal study on artificial grammar learning (AGL)(Reber, 1967). Participants were exposed to a series of letter sequences generated by a finite-state grammar (a set of possible letters together with rules for concatenating them), and were asked to memorise them. Following the exposure phase, they were informed that the strings followed a set of rules. They were then tested on a grammaticality judgment task with novel strings, and were asked to indicate whether the strings conformed to the grammar or not. Participants were significantly above chance level (79%) in discriminating between strings, even though they could not report what the underlying rule was. Therefore, Reber concluded that they had acquired implicit knowledge of the rule, and suggested that the same mechanism may be involved in other domains, including language acquisition (Reber, 1967). AGL learning is found even in subjects suffering from amnesia (Knowlton and Squire, 1996), which suggests that learning can occur in an entirely implicit way, without relying on the declarative system.

By separating the nature of the learning process from that of the resulting knowledge, we can further distinguish between *implicit* and *incidental* learning. In psychology, incidental learning refers to a situation in which subjects learn something without intending to, regardless of the nature of the knowledge acquired (Williams, 2009). In SLA, the term incidental learning, in its more general connotation, may be similarly defined as "learning without the intention to learn"; or, in a more specific interpretation, as the "learning of one stimulus aspect while paying attention to another stimulus aspect." (Hulstijn, 2003). For instance, incidental language learning may involve focusing on the acoustic properties of sounds while at the same time developing procedural knowledge of the articulatory movements required to produce them, or internalising the morphosyntactic features of an utterance while focusing on its semantic or pragmatic content (Paradis, 2004). In SLA research, the term *implicit* learning is sometimes used to refer to incidental learning (Schmidt, 1994; N. Ellis and Cadierno, 2009), while in the cognitive psychology literature, implicit learning refers specifically to a process of incidental learning which results in implicit knowledge (Reber, 1967; Williams, 2009; N. Ellis and Cadierno, 2009). Under this definition, an implicit learning process may result in explicit knowledge if learners reflect on the knowledge they have acquired implicitly; or, it is possible that a process of explicit learning directed at one feature may result in the incidental implicit learning of another feature (R. Ellis, 2009a, p. 6). Here, we will use the term *incidental* learning to refer to a process of "learning of one stimulus aspect while paying attention to another stimulus aspect." (Hulstijn, 2003), and the term *implicit* learning as incidental learning which results in implicit knowledge, following Williams (2009).

The processes involved in implicit learning have also been the object of study of a separate, but related, research strand in psychology, that on *statistical learning*. Rather than focusing on the learners' cognitive state, research on statistical language learning focuses on the complex statistical associations that exist between different elements of language, on the assumption that language acquisition involves acquiring information about those statistical regularities (Saffran, 2001; Saffran et al., 1996a). The mechanism was illustrated by Saffran et al. (1996a), who exposed 8-month old infants to a continuous stream of 3-syllable nonsense words (e.g. *babupudutabapidabu...*) repeated in random order by a speech synthesizer. Due to the lack of prosody, the only cue to word boundaries were the transitional probabilities between syllables (i.e. how likely a given syllable was to be followed by another one: very likely for syllables belonging to the same word, less likely for syllables on the boundary between two words). After only two minutes of exposure, they were exposed to 3-syllable combinations in a listening paradigm, where they could discriminate between "words" from the stream and other syllable combinations, even those made of part-words, showing that they were sensitive to the statistical properties of the input. Statistical learning is believed to be involved in the acquisition of sequences in general, even in domains other than language (Conway and Christiansen, 2005, 2009). While the defining feature of statistical learning is the use of statistical regularities in the input, the resulting knowledge is often unconscious, as is the case in implicit learning (Batterink et al., 2015b; Conway and Christiansen, 2005), although it may also involve a mixture of conscious and unconscious knowledge (Bertels et al., 2012; Franco et al., 2016). Neurophysiological evidence also suggests that both types of learning are supported by common neural mechanisms, involving the interplay of procedural and declarative knowledge (Batterink et al., 2019). Traditionally, statistical learning research has explained findings in terms of the computation of transitional probabilities, while implicit learning research has focused more on the acquisition of subsequences, or "chunks" (Perruchet, 2019). The other key distinction between the two traditions lies in the fact that implicit learning research has specifically focused on the *implicit* vs. *explicit* distinction, including ways of testing for awareness in participants. However, more recent research is bridging the gap between these two approaches, for instance by combining statistical learning paradigms with measures of awareness, showing that explicit knowledge can affect the way in which statistical learning operates (Monaghan et al., 2019). Due to the large amount of overlap between implicit and statistical learning, the term "implicit statistical learning" is increasingly being used (Christiansen, 2019; Reber, 2015; Walk and Conway, 2015) to stress the fact that they investigate the same phenomena, only from different angles.

In the field of SLA research, the operation of implicit statistical learning mechanisms on language acquisition can be captured by the notion of *contingency learning* (N. Ellis, 2006a,b; Shanks, 1995), also known as *associative learning*. Like statistical learning, the concept of contingency learning entails that language acquisition is a process of acquiring statistical regularities within the input; however, it emphasises the fact that the statistical regularities available to learners are not simply co-occurences between forms in the input, but also *cues* to an outcome (Williams, 2009). Often, this may be a form to meaning contingency (N. Ellis, 2006a), such as the final *-s* in *walks* in the sentence "He walks up the stairs", which is a cue to the fact that the verb is inflected in the 3rd person singular. The *validity* of a cue is a function of its availability (how frequent it is) and its reliability (how likely it is to predict the outcome) (N. Ellis, 2006a). In L1 acquisition, high-validity cues are the first ones children focus on (MacWhinney et al., 1985); these are purely statistical properties, akin to transitional probabilities. However, contingency learning can also account for non-statistical factors which may affect how likely learners are to attend to a given cue, such as perceptual saliency and competition between cues. For instance, the sentence "He walks up the stairs" contains redundant information. The ending *-s* is not the only cue to the verb inflection: the pronoun *he*, too, indicates that the verb is inflected in the 3rd person singular, and it too is a widely available and reliable cue. However, *he* is more perceptually salient than *-s*; therefore, it is more likely to be attended to, at the expense of *-s*. This is an attested phenomenon among

L2 English learners (Hawkins and Casillas, 2008; White, 2003) and can be seen as an instance of *overshadowing*, which happens when two cues jointly predict the same outcome, but one is more salient than the other (N. Ellis, 2006a).

2.2.1 Testing the knowledge derived from implicit learning

Experimental research on implicit learning, particularly in the tradition derived from Reber's 1967 artificial grammar learning experiments, relies on the ability to distinguish between explicit and implicit knowledge. In AGL paradigms, in order for learning to be implicit, participants must not be aware of the regularity they have learned, while demonstrating some knowledge of the rule: under the definition of implicit knowledge as "causally efficacious in the absence of awareness" (Cleeremans et al., 1998, p. 406), if a subject exhibits knowledge but is not aware of its content, we can conclude that the knowledge is implicit. Given the emphasis on ensuring that the knowledge being used is implicit, the way in which awareness is measured is critical to the outcome. Therefore, implicit learning studies aim to detect implicit knowledge by combining tests of knowledge (such as grammaticality judgment tasks) with measures of awareness. When knowledge is evaluated by a judgment task, as is often the case in AGL studies, awareness can be assessed by subjective measures: participants are asked to report the source of their answers (e.g. "guess", "memory" or "rule", or "intuition"), and to state how confident they are (Dienes and Scott, 2005; Rebuschat, 2008; Rebuschat and Williams, 2012). The premise of this paradigm is that it allows to distinguish between two kinds of knowledge: *structural knowledge*, which subjects have about the stimuli they have been exposed to, and *judgement knowledge*, which is the ability to recognise whether new items follow the same structure as previously encountered ones (Dienes and Scott, 2005). While performance on a test such as a GJT reveals structural knowledge, subjective measures allow researchers to investigate the nature of a subject's judgement knowledge. According to Dienes and Scott (2005), conscious structural knowledge should lead to conscious judgement knowledge (which subjects may attribute to "rule" or "memory"). However, it is possible for structural knowledge to be implicit while judgement knowledge is explicit - that would be a case of "intuition", which reflects the common situation of L1 speakers who may be very confident of their grammaticality judgements, even though they do not know why (Rebuschat, 2013). Finally, if a subject's accuracy is above chance even though they attribute their responses to "guess" (the *guessing* criterion), we can conclude that both their structural knowledge and their judgement knowledge are unconscious (Dienes et al., 1995). One potential problem with this methodology is response bias (for instance, some subjects may attribute their responses to "guess" unless they are absolutely sure, others may be more liberal). A second problem is that, by asking participants to consciously reflect on their answers, it may encourage rule-searching and cause participants to become aware of the rule.

Perhaps the most commonly used measure of awareness, which dates back to the original Reber (1967) study, is retrospective verbal report (Dienes et al., 1991; Reber, 1967; Williams, 2005). Participants are simply asked after the test phase whether they noticed any patterns or rules in the stimuli (Williams, 2005). If their test performance shows sensitivity to the regularity but they do not report any awareness of it, then it is assumed that the learning was implicit. One advantage of retrospective verbal report is that it is not task-specific: subjective measures are well suited for judgment tasks but not as appropriate for tasks involving spoken production or elicited recall, such as the ones used in our study. The use of retrospective verbal reporting is not without problems, however. It has been suggested that it may not be the most appropriate way to detect implicit knowledge, on the basis of what Shanks and St. John (1994) call the *Sensitivity* and *Information* criteria. By sensitivity, they mean that a measure of awareness needs to be sensitive enough to detect as much conscious knowledge as possible. With retrospective verbal reporting, there is a risk that participants may be under-reporting conscious knowledge, if they do not feel very confident (Rebuschat, 2013). The Information criterion, on the other hand, concerns the content of the knowledge being tested: for the measure of awareness to have any validity, it must be assessing conscious access to the knowledge that is actually driving participants' performance in the test. For instance, in an AGL paradigm using a finite-state grammar, it makes little sense to determine whether participants have conscious knowledge of the grammar if it's possible for participants to perform above chance by relying on means other than grammar knowledge, such as memory for chunks. While this is a genuine concern, it may be argued that it is less relevant for studies that do not involve abstract structural rules but focus instead on form-meaning connections and usage patterns, such as the study by Williams (2005) on animacy-based determiners. In these studies, there is no distinction to be drawn between patterns of co-occurrence and an abstract rule, since the co-occurrence patterns *are* the rule. In our study, we decided to use a debriefing questionnaire as a measure of awareness. Details of the questionnaire and motivations for using it can be found in Section 3.1.2.3.

Detecting implicit knowledge by the absence of awareness is only viable if subjects do not also have some explicit knowledge of the same regularity. While this may be obtained under controlled laboratory conditions, it is rarely the case with natural language, whether L1 or L2. For this reason, the ability to directly test for implicit and explicit knowledge is particularly important in SLA research. Implicit knowledge may be accessed through online measurements such as eye-tracking (Godfroid et al., 2015; Godfroid and Winke, 2015) or EEG (Batterink and Paller, 2017; Batterink et al., 2015b). For instance, morphosyntactic violations in the L1 normally elicit the P600 component (a positive shift in electric potential registered around the parietal lobe, 600ms after stimulus onset); as proficiency increases, a shift towards this pattern of processing can be detected in L2 learners (Osterhout et al., 2008). Investigations of incidental learning of an artificial language using EEG (Friederici et al., 2002; Morgan-Short et al., 2012b) have revealed native-like processing of incongruities after very little instruction. Additionally, EEG measurements can reveal sensitivity to newly acquired distinctions even though behavioural measures show no difference (McLaughlin et al., 2004; Tokowicz and MacWhinney, 2005).

Behaviourally, it is possible to detect implicit knowledge by focusing on its second hallmark, lack of voluntary control. Under the assumption that implicit knowledge affects behaviour automatically, while explicit knowledge can be used deliberately, it is possible to discriminate between the two with tasks that manipulate the degree of conscious control involved in performing the task. For instance, in a *process dissociation procedure* (Jacoby, 1991), subjects are instructed to either use the knowledge they acquired (inclusion task) or avoid using it (exclusion task). If participants are able to refrain from using the knowledge in the exclusion task, it assumed to be conscious, while failure to do so indicates unconscious knowledge. According to R. Ellis (2006c), implicit knowledge is best assessed through tasks that tap into what learners intuitively feel is correct, are time-pressured, place primary focus on meaning, and do not call for metalinguistic knowledge. Oral production tasks that encourage communicative and spontaneous L2 use with few constraints are also considered a good way to tap into implicit knowledge (Doughty, 2003; Norris and Ortega, 2003), although it can be difficult to elicit target structures under these conditions. Conversely, explicit knowledge should be assessed with tasks that encourage learners to use "rules", are not time-pressured, place primary focus on form, and invite the use of metalinguistic knowledge. For this reason, ungrammatical sentences in a GJT may be a better indicator of explicit knowledge than grammatical ones (Gutiérrez, 2013), because detecting them requires greater metalinguistic knowledge. However, evidence from eye-tracking suggests that the effect of ungrammatical items may be reduced under time pressure, which suppresses right-to-left gaze regressions and prevents reanalysis of ungrammatical items (Godfroid et al., 2015). In fact, the same task conducted with and without time pressure may tap into different types of knowledge. Untimed Grammaticality Judgment tasks (GJT) and metalinguistic tests tend to cluster together, while both elicited oral imitation and narration tasks correlate with performance on a time-pressured GJT, suggesting that they tap into implicit knowledge (R. Ellis, 2009b). Likewise, in a series of studies on the processing of L2 Spanish by English speakers, Tokowicz and colleagues find a dissociation between untimed GJT and online measures: learners showed sensitivity to morphosyntactic violations in both self-paced reading (Tokowicz and Warren, 2010) and EEG (Tokowicz and MacWhinney, 2005), but not in an untimed GJT. More generally, a lack of time constraints allows explicit knowledge to be used, especially in the case of written tasks, while performing a task under time pressure is more likely to engage implicit knowledge (N. Ellis, 2005; Norris and Ortega, 2003).

A commonly used measure of implicit knowledge which involves both time pressure and oral production is the Elicited Oral Imitation task (Erlam, 2006, 2009; Slobin and Welsh, 1973), which forms the basis for the paradigm used in this study. The task relies on the observation that, when processing language, the amount of material we can recall depends on the extent to which we have stored representations of it in memory (i.e. previous knowledge). For instance, English speakers are better at remembering lists of English words than Aymara words of the same length (Erlam, 2009). This facilitatory effect of existing representations is thought to come from regeneration: a sentence in a known language can be decoded and then re-encoded upon recall, whereas one in an unknown one can only be stored in short-term phonological memory as a string of phonemes (Lombardi and Potter, 1992; Potter and Lombardi, 1990, 1998; Slobin and Welsh, 1973). Therefore, the accuracy with which an L2 sentence is recalled by a learner provides a measure of the learner's L2 knowledge. Critics of this interpretation have suggested that need not be the case, and that it may be possible to do the task just by rote repetition (Bley-Vroman and Chaudron, 1994; Erlam, 2006; Spada et al., 2015; Vinther, 2002). However, there is empirical evidence in support of the "regeneration hypothesis" (Lombardi and Potter, 1992; Potter and Lombardi, 1990), such as the finding that intruding semantically related items during recall cause interference effects (Potter and Lombardi, 1990). Such effects are even found in bilinguals when the intervening task is in a different language from the sentence, and in monolinguals when the intervening task is picture-based (Lee and Williams, 1997). The hypothesis is also supported by the tendency to normalisation repeatedly found in elicited imitation studies: when asked to repeat ungrammatical sentences, participants will often correct any mistakes in the sentences without realising it (Erlam, 2009; Slobin and Welsh, 1973). According to Potter and Lombardi, the fact that recall is often verbatim is due to the fact that exposure to a sentence (whether for comprehension or recall) causes long-term activation of the lexical-semantic representations of the words in the sentence, which in turn makes it more likely that they will be accessed in production (Potter and Lombardi, 1990); syntactic priming can also contribute to making recall verbatim (Potter and Lombardi, 1998). Conversely, Rummer and colleagues (Rummer et al., 2013; Schweppe et al., 2015, 2009) argue that exposure to a sentence can cause long-term activation of representations on several levels, and that a verbatim recall task will cause both grammatical information and surface phonological representations to be maintained (Rummer et al., 2013).

Finally, it has also been suggested that the EOIT may not be a suitable way to tap into implicit knowledge, as it could be affected by metalinguistic knowledge, for instance through self-monitoring during production (Suzuki and DeKeyser, 2015). In meta-reviews, the EOIT has consistently been found to correlate with measures of implicit knowledge, such as time-pressured GJT and oral narration tasks (N. Ellis, 2005; Bowles, 2011; Erlam, 2009; Spada et al., 2015; Zhang, 2015). However, Suzuki and DeKeyser (2015) argue that performance the task should be seen as a reflection of automatised explicit knowledge, rather than true implicit knowledge, because it correlates with time-pressured judgement tasks but not with others which are deemed to be more robust to intrusion from explicit knowledge, such as word monitoring tasks (Suzuki and DeKeyser, 2015). Arguably, the role of automatised explicit knowledge is less of a concern for implicit learning studies using novel stimuli, where participants will not have had the amount of practice necessary for automatisation. It is possible, however, that the EOIT could generally reflect a contribution from explicit knowledge developed during exposure to the stimuli of implicit learning study, perhaps through monitoring. For this reason, the experiments in this study will also include retrospective verbal report as a measure of awareness, in order to establish the presence of implicit knowledge with greater confidence.

2.2.2 What can be learned implicitly in L2?

The application of implicit statistical learning paradigms to language acquisition research has shown that we can implicitly acquire knowledge of various aspects of language: from structural properties (such as word and phrase structure, orthographic and phonotactic rules, and syntactic structures) to form-meaning connections, such as novel determiners encoding animacy and thematic roles. Experimental evidence supporting these claims is reviewed below.

2.2.2.1 Structural properties

The finding that language learners can use transitional probabilities to segment individual words from the speech stream (Saffran et al., 1996a) has been replicated in numerous studies with both infants (Aslin et al., 1998; Hay et al., 2011;

Johnson and Jusczyk, 2001; Saffran, 2001) and adults (Karuza et al., 2013; Saffran et al., 1996b). It also appears that infants can use the words extracted through statistical learning as labels for new objects (Cunillera et al., 2010; Estes et al., 2007; Hay et al., 2011), which suggests that the mechanism could plausibly play a role in L1 vocabulary acquisition. There is evidence to suggest that different languages may require different segmenting strategies, however, as natural languages differ in the type of statistical cues they contain (Saksida et al., 2017). Phonotactic regularities, too, can be acquired implicitly: infants can use phonotactic cues as well as transitional probabilities to segment words (Chambers et al., 2003; Johnson and Jusczyk, 2001; Mattys and Jusczyk, 2001; Mattys et al., 1999), and adults can learn phonotactic constraints through implicit statistical learning (Dell et al., 2000; Warker and Dell, 2006), as well as orthographic regularities (Chetail, 2017; Pacton et al., 2001). Research using artificial stimuli has shown that implicit statistical learning can also support the acquisition of more complex structures, such as phrasal units (Morgan et al., 1987; Thompson and Newport, 2007) and non-adjacent dependencies (Amato and MacDonald, 2010; Ferry et al., 2016; Gómez, 2002; Lany and Gómez, 2008; Lany et al., 2007; Pacton and Perruchet, 2008). The evidence for non-adjacent dependencies (e.g. AxB) is in fact mixed (Newport and Aslin, 2004), but it appears that they can be acquired under facilitating conditions: for instance, if there is sufficient variety in intervening syllables (Gómez, 2002), if lexical meaning is added to the elements (Amato and MacDonald, 2010), if participants have their attention directed to the non-adjacent elements (Pacton and Perruchet, 2008) or if they are first exposed to simpler adjacent dependencies (Lany and Gómez, 2008; Lany et al., 2007). There is evidence that hierarchical structures, too, can be acquired through statistical learning (De Vries et al., 2012; Fitch and Hauser, 2004). In addition to finite state grammars (e.g. ABABAB), it appears that a nested grammar, too (e.g. AAABBB) can be learned implicitly (Fitch and Hauser, 2004); however, if embedded dependencies are added (e.g. $A_1A_2A_3B_3B_2B_1$), the grammar is still learnable, but only up to two-level embeddings (De Vries et al., 2012).

Perhaps of greater interest for SLA research is whether learners can implicitly acquire natural language syntax, too. Studies applying implicit learning paradigms to the acquisition of novel syntactic structures suggest that this is the case (Francis et al., 2009; Rebuschat, 2008; Rebuschat and Williams, 2012; Williams and Kuribara, 2008; Williams and Rebuschat, 2012). Rebuschat and Williams (2012, Exp. 2) exposed participants to sentences made up of English lexis, but arranged according to three possible German word order patterns. Participants were instructed to repeat each sentence and judge its semantic plausibility; after the exposure phase, they were tested on a GJT with subjective measures of awareness. A control group was also included, who did the GJT without any prior exposure to the stimuli. In the GJT, endorsement of novel grammatical items by unaware participants (who attributed their judgments to *intuition*) was higher than that of the control group and significantly above chance, indicating that some implicit learning of the novel structures had occurred, and that the knowledge could be generalised to new items. However, unaware participants were not able to reliably reject ungrammatical items, and they were more likely than the control group to endorse them. This suggests that they were at least partly basing their responses on memory for patterns encountered in training.

Rebuschat and Williams's findings (Rebuschat and Williams, 2012) illustrate the tension between two posited learning mechanisms underlying implicit learning, namely abstract rule learning and *chunking* (the memorisation of grammatical substrings from training items, which participants may then rely on when giving grammaticality judgement for novel items). Since the original Reber (1967) study, the issue of whether learning effects in AGL tasks are best explained in terms of abstract rule learning or chunking has been a long-standing debate in the field (Perruchet, 2019; Perruchet and Pacton, 2006).[1] Indeed, the findings that learners can automatically segment a sequence of syllables into words (Saffran et al., 1996a) and that they can automatically extract the abstract rules underlying letter strings in AGL (Reber, 1967) are seemingly at odds with each other. These are different ways of approaching a sequence of stimuli - one where transitional probabilities are used to segment the surface form into chunks, and one where the same transitional probabilities are used to extract abstract com-

[1] This was partly motivated by Reber's original claim that participants had acquired a "rule" (Reber, 1967). Reber has since distanced himself from the expression, and clarified that he sees the core operation of implicit learning as the "detection of patterns of covariation between elements in complex stimulus domains" (Reber, 2015, p. viii), rather than the acquisition of abstract grammatical rules.

binatorial rules beneath that surface form. It is not clear what should determine which of the two mechanisms is used given a specific set of stimuli.

On the one hand, there is evidence that participants in AGL experiments are acquiring something more abstract than letter chunks, as evidenced by transfer to different letter sets, and even to different modalities. A way to test for the transferability of learned regularities in AGL studies is to train participants on one grammar, and then test them on sequences generated by a second grammar, which is derived from the first by systematically swapping the letters with new ones. Studies adopting this method showed that participants were still performing significantly above chance in a GJT, even if accuracy was lower than for the training letter set (Knowlton and Squire, 1996; Mathews et al., 1989; Reber, 1967). Transfer is possible across modalities, too. Participants who were trained on an artificial grammar in the auditory modality still performed above chance when tested in the visual modality (Altmann et al., 1995; Bigand et al., 1998). Transfer from the visual to auditory modality, however, is not equally successful (Forkstam et al., 2009), perhaps because different modalities seem to affect implicit learning in different ways: comparisons of performance on artificial grammar learning tasks across visual, auditory and tactile modalities have found that learning outcomes are best when presentation was auditory, especially at faster presentation rates (Conway and Christiansen, 2005, 2009). On the other hand, the extent to which the input can be broken down into chunks has been shown to have an effect on participants' grammaticality judgements, too (Knowlton and Squire, 1996). Knowlton and Squire (1996) constructed an AGL experiment where chunk strength (the frequency of a given chunk in the input) was manipulated. In a GJT administered after exposure, participants were above chance when endorsing novel grammatical strings, with similar endorsement rates regardless of whether the items were made up of low- or high-frequency chunks. However, for ungrammatical items, chunk strength was shown to have an effect: novel ungrammatical strings containing frequent chunks were significantly more likely to be judged as grammatical. Knowlton and Squire (1996) conclude that participants were relying on both item-specific and abstract information when doing the task. Knowlton and Squire's findings have been replicated by Robinson (2005) using a natural language (Samoan). More recently, Frost and Monaghan

(2016) have shown that segmentation and the extraction of non-adjacent dependencies from the speech stream can occur simultaneously from the same cues, suggesting that both speech segmentation and the extraction of regularities depend on the same mechanism. Therefore, it would appear that both mechanisms are available and operating at the same time.

2.2.2.2 Form-meaning connections

The distinction between chunking and the extraction of abstract combinatorial rules is particularly relevant to the study of how learners process continuous streams of stimuli, such as those used in AGL and statistical learning paradigms. These stimuli are usually devoid of meaning, as the focus is on the acquisition of form. A more recent area of research within implicit learning, which is also the object of the present study, seeks to investigate the acquisition of novel form-meaning connections, using either natural or artificial languages. These studies have shown that it is possible to implicitly learn the meaning of novel determiners encoding animacy (Williams, 2005; Williams et al., 2004) or thematic roles (Leung and Williams, 2006, 2011, 2012), novel morphology (Marsden et al., 2013; Rogers et al., 2016) and verb usage patterns (Paciorek, 2012; Paciorek and Williams, 2015).

In Williams (2005, Exp. 1), L1 English speakers were exposed to a semi-artificial language in which novel determiners *gi*, *ro*, *ul* and *ne* were paired with English nouns. Participants were first introduced to the determiners and told that two of them were used with near objects (*gi* and *ro*), and the other two with far objects (*ul* and *ne*). Unknown to participants, the determiners were also alternating based on the objects' animacy: *gi* and *ul* only occurred with living things, *ro* and *ne* only with non-living ones. Participants were exposed to the determiner-noun pairs (e.g. *gi lion, ro table*) embedded in English carrier sentences (e.g. "I was terrified when I turned around and saw *gi lion* right behind me"). For each sentence, they had to indicate whether the object following the determiner was near or far, and whether it was living or non-living, by pressing the appropriate keys on a keyboard. After the exposure phase, participants did a forced choice task in which they were presented with individual nouns (including nouns not encountered during training) and had to select the correct determiner for each. Awareness of the animacy rule was assessed by verbal report using a debriefing questionnaire at the end of the experiment. Even participants who remained unaware of the rule were significantly above chance in the forced choice task even with novel nouns, indicating that they had acquired implicit knowledge of the distinction and could generalise it to new instances. The effect has been replicated with L1 Chinese speakers (Chen et al., 2011).

However, other replication attempts have yielded conflicting results (Faretta-Stutenberg et al., 2011; Hama and Leow, 2010; Rebuschat et al., 2015, 2013). Unlike Chen et al. (2011), which tested a different population with the same paradigm, these replication attempts aimed to further investigate the nature of the knowledge acquired, by using different methodologies to test for awareness both during and after the learning task. Hama and Leow (2010) reproduced Williams (2005) with the addition of a think-aloud protocol to assess awareness during the learning task, as well as a production task in the test phase. They found no evidence of learning in participants who were classed as unaware; similar results were obtained by Faretta-Stutenberg et al. (2011). Rebuschat et al. (2013) pointed out that both think-aloud and verbal retrospective verbal reports were limited in that they rely on verbalisation to assess awareness. They replicated Williams (2005) using confidence ratings and source attribution (*guess, intuition, memory, rule*) instead. They found that even when participants could verbalise the rule (70% could) some of the knowledge remained implicit, as evidenced by above chance performance on responses attributed to guess and intuition. In a subsequent replication, Rebuschat et al. (2015) directly compared source attribution and think-aloud reports as measures of awareness in a between-group design. The results from the source attribution group corroborated findings from Rebuschat et al. (2013). On the other hand, the think-aloud group showed reduced learning - specifically, no ability to generalise the rule to new items - leading the authors to suggest that the think-aloud protocol itself was interfering with the learning process, which may explain the results of Hama and Leow (2010).

The claim that novel form-meaning connections can be learned implicitly may seem to be in contrast with dual-system accounts, which assume that vocabulary knowledge is explicit, represented in declarative memory (Ullman, 2001, 2005) and dependent on hippocampal structures for its acquisition (N. Ellis, 1994, 2005), as evidenced by the fact that vocabulary learning is impaired in amnesics (Gabrieli et al., 1988). As Paciorek and Williams (2015) note, however, what is learned implicitly is not the novel arbitrary mapping between a form and its core (referential) meaning, but rather the usage patterns for the new form, which are also part of lexical meaning (collocational meaning) (Paciorek and Williams, 2015, p. 999).

In a development of Williams (2005), Leung and Williams (2006, 2012) used the same set of determiners (*gi*, *ro*, *ul* and *ne*) embedded in English sentences, but this time they were encoding thematic role (*agent* vs. *patient*) rather than animacy. Participants were informed that determiners *gi* and *ro* were used before personal names referring to adults, while *ul* and *ne* were used before those referring to children. They were not told that *gi* and *ul* were only used with agents, and *ro* and *ne* with patients. Leung and Williams also introduced a novel training and testing methodology based on reaction times. Participants heard each sentence in a fixed word order that did not give any thematic role cues (e.g. "Kiss *ul Mary* a boy on the face") while being a shown the corresponding picture (in this case, of a girl kissing a boy on the cheek). Then then had to indicate by button press whether the named character (Mary in this case) was on the left or right of the picture. If participants became sensitive to the fact that the determiner contained a cue to the thematic role of the named character, they would be able to direct their attention to the relevant character in the picture as soon as they heard the determiner, and therefore become faster in their responses. After giving their response, they were asked to reformulate the sentence using English word order but retaining the determiner; every two sentences, they had to recall the reformulated versions of the previous two items. All this was done to ensure that participants would pay attention to the sentences in their entirety, and process thematic roles as expected. Awareness of the thematic role rule was assessed by debriefing questionnaire. In the last block of the experiment, the mapping between determiners and thematic roles was inverted (violation items). Participants who remained unaware of the rule showed the expected decrease in reaction times as the experiment progressed, and significant increase in the violation block at the end, indicating that they had developed sensitivity to the thematic role markers, while remaining unaware of the regularity.

As noted by Leung and Williams (2011), animacy is a semantic feature of nouns and can be derived from the lexicon, while thematic role is contextually determined, which means that acquiring thematic role markers requires learners to "abstract over representations of instances of use of the word that contains contextually derived information." (Leung and Williams, 2011, p. 38). The finding that this can happen implicitly is compatible with usage-based approaches to SLA, which assume that individual instances are what is initially is acquired, and that their frequency distribution forms the basis for generalisation. Commenting on the role of frequency in language processing and learning, N. Ellis (2002) notes that:

> "...language learning is exemplar based. The evidence reviewed here suggests that the knowledge underlying fluent use of language is not grammar in the sense of abstract rules or structure but a huge collection of memories of previously experienced utterances. [...] Linguistic regularities emerge as central tendencies in the conspiracy of the database of memories of utterances."(N. Ellis, 2002, p. 166)

It is a well-known fact that morphological and morphosyntactic markers are a common stumbling block for L2 learners, in both production (Bardovi-Harlig, 2000; Clahsen and Felser, 2006; Klein, 1986), and comprehension (Jiang, 2004, 2007). Usage patterns, too, are notoriously difficult to acquire in L2. For instance, speakers of languages with no distinction between definite and indefinite articles (e.g. Slavonic languages) have little difficulty acquiring the meaning of the determiners *the* and *a*, but struggle with using them in the appropriate contexts. Likewise, the meaning of spatial and temporal prepositions in English and other languages is largely a function of their usage, which creates a complex set of distributional and idiomatic properties (e.g. cf. the expressions "at the time", "on time", and "in time"). These are areas of SLA which could plausibly benefit from the application of implicit learning paradigms. The following section outlines the work done on comparing the role of explicit and implicit instruction on SLA, highlighting some of the potential benefits of applying implicit learning paradigms to L2 instruction.

2.3 Applying implicit learning paradigms to L2 instruction

The research reviewed so far has focused on investigating implicit learning mechanisms in controlled experimental settings, with the aim of discovering what, if any, properties of language can be acquired implicitly. We will now turn to the application of implicit learning to SLA research and L2 pedagogy. In the light of Krashen's claims on the distinction between learning and acquisition, and the superiority of the latter (Krashen, 1981b), the idea that implicit L2 knowledge may be developed through implicit learning has held considerable interest for L2 pedagogy. Research on the application of implicit learning methodologies to SLA has tended to focus on learning outcomes in comparison with explicit learning: rather than asking whether a given property of language can be learned implicitly, the point of interest is whether it is *best* learned implicitly. It has been suggested that complex rules (e.g. relative clause structure) may be harder to teach explicitly than simple ones (e.g. morphological markers), and that better results should be obtained by teaching them through implicit learning, with meaning-based practice (Krashen, 1994, 1982). This is supported by evidence from AGL research: explicit rule search can be detrimental to performance on more complex sequences in AGL (Reber, 1976, 1989) and serial reaction tasks (Fletcher et al., 2005). However, the opposite point has also been made: Hulstijn et al. (1994) claim that "hard" (complex) rules are too difficult to notice in natural settings, and should therefore benefit from explicit teaching, while simpler rules are more easily noticed and can therefore be learned implicitly. In fact, the very definition of "simple" and "complex" rules can vary considerably between authors, depending on the criteria used (Spada and Tomita, 2010). Perhaps due to these methodological issues, research in this field has so far found limited evidence that implicit learning is preferable for complex rules (DeKeyser and Sokalski, 1996); however, there seems to be a large amount of evidence showing that explicit instruction is generally more effective than implicit learning (DeKeyser, 1995; Goo et al., 2015; Norris and Ortega, 2000; Robinson, 1996; Spada and Tomita, 2010).

DeKeyser (1995) investigated the acquisition of Implexan, a miniature artificial language with morphological inflections for gender, number and thematic role, comparing a deductive (explicit) and an inductive (implicit) methodology. The goal of the study was to test the hypothesis that deductive learning would be more effective for categorical rules, while "fuzzy" (probabilistic) rules would be best taught using inductive methods. During training, Implexan sentences appeared on screen one at a time, together with pictures depicting their meaning. Occasionally, participants were asked to indicate whether the content of the sentence matched the picture (incorrect sentences contained vocabulary errors, but were never ungrammatical). The training procedure was the same for both groups, with the exception that the explicit group were also given explicit rule instruction at various times during the training sessions. Participants were then tested on a written production task, where they had to type sentences matching a picture cue. Accuracy on vocabulary was similar for the two groups, but accuracy on categorical grammatical rules was significantly higher for the explicit group. Performance on fuzzy rules was somewhat better for the implicit group (they were closer to the expected prototipicality pattern in their responses), but not significantly so. Similarly, DeKeyser and Sokalski (1996) compared the acquisition of simple and complex English grammatical rules under implicit, incidental, rule-search and instructed conditions. Participants in the implicit conditions were instructed to focus on word order, while subjects in the incidental conditions were asked comprehension questions focusing on the meaning of the sentences. In a GJT administered at the end of the experiment, the instructed group outperformed all other groups on both simple and hard rules. The implicit and incidental group were not better than the instructed group on complex rules; however a comparison of the implicit and rule-search groups shows that the formed did better on complex rules and the latter on simple rules, which replicates Reber's findings with AGL (Reber, 1976).

Norris and Ortega (2000) carried out a meta-review of SLA studies to assess the effectiveness of different types of instruction. Explicit and implicit instruction were defined following the same criteria as DeKeyser and Sokalski (1996): explicit instruction involved explicit rule explanation (deductive, metalinguistic) and/or directing learners' attention to specific forms with the aim of figuring out the rules (inductive). Implicit instruction was defined as a treatment involving neither rule presentation nor directions to attend to particular forms (Norris and

Ortega, 2000, p. 437). Norris and Ortega (2000) found larger effect sizes for explicit instruction. However, the tests used in the studies they review were largely measures of explicit knowledge: 65% used constrained constructed response measures, 39% used selected response, 29% used metalinguistic judgments, and 16% used free constructed response (Norris and Ortega, 2000, p. 470). The closest thing to a test of implicit knowledge were "free constructed responses", defined by the authors as "measures that required participants to produce language with relatively few constraints and with meaningful communication as the goal for L2 production (e.g., oral interviews, written compositions)" (Norris and Ortega, 2000, p. 440). Free constructed response tests showed the smallest effect sizes across all studies (no information was provided about the specific effect of implicit instruction on this type of test). However, it should also be noted that due to the unconstrained and communicative nature of the tasks, scoring for accuracy on a particular feature depended on whether the feature was spontaneously produced by the learner (Norris and Ortega, 2000, p. 441), which may not necessarily provide an accurate picture of the learner's competence level.

Spada and Tomita (2010), using the same definitions of implicit and explicit instruction as Norris and Ortega (2000), carried out a meta-review comparing effect sizes obtained with explicit and implicit learning methodologies. Unlike Norris and Ortega (2000), they focused on studies on L2 English learning, including a variety of simple and complex grammatical features. Spada and Tomita (2010), too, found larger effect sizes for explicit than for implicit instruction, on both simple and complex structures. This is the case both for both "controlled" and "free" test measures (defined according to the same criteria used by Norris and Ortega (2000)), which should in theory tap into explicit and implicit knowledge respectively. However, Spada and Tomita (2010) suggest that some of the tasks classed as free response, such as picture-cued oral performance tasks, may rely on speeded explicit knowledge, rather than implicit (Spada and Tomita, 2010, p. 287). The "free response" category also included free written production tasks, which may be argued to rely on explicit knowledge. Goo et al. (2015) carried out a meta-review of SLA studies published between 1993 and 2011, following the same criteria as Norris and Ortega (2000) and Spada and Tomita (2010) for defining implicit and explicit instruction and for classifying the types of knowledge test used. As in previous meta-reviews, Goo et al. (2015), too, found explicit instruction to be more effective than implicit instruction on all measures, including free response.

2.3.1 Issues affecting the application of implicit learning to L2 instruction

Overall, the evidence reviewed above suggests a limited role for implicit learning paradigms in L2 pedagogy. However, there are a number of issues with the way in which implicit learning and knowledge are treated in these studies, which are crucial to learning outcomes: 1) the way in which knowledge was tested, 2) what sort of activities were designated as implicit learning, and 3) whether training and testing were based on comprehension or production activities.

The first issue is how the resulting knowledge was tested. The distinctive feature of implicit learning is that it can help develop implicit knowledge, yet the studies surveyed did not always use tasks designed to tap into implicit knowledge, possibly because they focused on comparing the relative effectiveness of implicit and explicit instruction for different types of rules instead. DeKeyser (1995) used an untimed written production task to test for learning; Robinson (1996) used a timed grammaticality judgement task. Similarly, the meta-reviews by Norris and Ortega (2000), Spada and Tomita (2010) and Goo et al. (2015) did not classify outcome measures depending on their suitability for detecting implicit or explicit knowledge (e.g. time-pressured vs. untimed task). Some, but not all, of these categories could be directly mapped onto a specific type of knowledge: for instance, tests involving "metalinguistic judgement" are good measures of explicit knowledge. The category closest to a set of implicit knowledge measures was "free constructed response"; however, even this category included tasks which are likely to benefit from explicit knowledge to a large extent, such as untimed writing tasks.

The second issue which affects these studies is the fact that implicit instruction in these studies was operationalised as simple exposure to the stimuli, without either explicit instruction or any manipulation of attention (DeKeyser and Sokalski, 1996; Norris and Ortega, 2000; Spada and Tomita, 2010). However, as VanPatten et al. (2013) note, the question of whether learners are provided with explicit rule instruction is separate from the explicit/implicit learning distinction: it is possible for learning to be explicit even if no rule instruction is provided (VanPatten et al., 2013). Likewise, directing learners' attention to particular aspects of the stimuli is compatible with both explicit and implicit learning, and may in some cases be necessary for L2 learning to occur at all. Section 2.4 below will explore the role of attention in SLA and in implicit learning specifically.

Finally, the studies included did not control for modality: for instance, whether training was done with comprehension tasks and testing with production tasks, or vice versa. Participants in DeKeyser (1995) were trained with comprehension-based activities, but tested on a written production task. Tasks classed as "free response" in the meta-reviews (Goo et al., 2015; Norris and Ortega, 2000; Spada and Tomita, 2010) were all production tasks, but there is no information on whether training was production-based. The issue of modality and the relation between production and comprehension in L2 learning will be explored in greater detail in Section 2.5.

2.4 The role of attention in SLA and implicit learning

In the studies reviewed in the previous section, what is meant by "implicit learning" is often mere exposure to the stimuli, as in DeKeyser (1995). In the meta-reviews, too, implicit instruction was defined in negative terms, as learning without explicit rule explanation, or where participants' attention was not directed to particular forms. However, we know that attention is a fundamental aspect of SLA, and that mere exposure often does not result in successful L2 learning in adults. Attention plays a central role in theories of L2 acquisition (Leow, 2015; Robinson, 2003; Schmidt, 1992a, 1990). According to Schmidt's Noticing Hypothesis (Schmidt, 1992b, 1990), only forms or form-meaning connections that are attended to - and therefore, noticed - can be learned: this is what Schmidt calls "awareness at the level of noticing", and is different from having explicit knowledge of the underlying regularity, which he calls "awareness at the level of understanding". While the latter is not necessarily a prerequisite for language learning, noticing of forms is (Schmidt, 2001). In naturalistic L2 learning, not everything is picked up by learners. For instance, observation of naturalistic language learning shows how L2 learners who acquired English in everyday settings settle on a "basic variety" of English with minimal morphology and inflection (Perdue, 1993; N. Ellis, 2006b; N. Ellis and Sagarra, 2010). This may be due to maturational constraints which specifically affect different components of language (Hyltenstam and Abrahamsson, 2003), or to a decrease in the ability to develop new procedural representations, in favour of declarative memory (Ullman, 2005). However, it can also be due to attentional bias: both explicit knowledge of the L2 and prior L1 experience can shape the way in which learners attend to the input.

It has been shown that awareness of regularities in the input can affect the way in which the input is attended to, not only in L2 learning, but even in basic L1 processing. For instance, in L1 priming, it is sufficient to manipulate whether primes can be consciously perceived (by making them either subliminal or visible) to alter the nature of the priming effects elicited (Alonso et al., 2006). Likewise, attentional bias exerts a powerful effect on L2 learning. In contingency learning, attention to L2 cues can be shaped by overshadowing and blocking mechanisms. Over time, overshadowing (competition between cues of different salience) gives rise to blocking: a cue that is initially overshadowed by a stronger one is then disregarded even if it becomes a good predictor, compared to another one with the same predictive power, but no previous overshadowing (Chapman and Robbins, 1990; N. Ellis, 2006a). N. Ellis and Sagarra examined the effect of blocking on the acquisition of morphological and lexical cues to temporal reference (specifically, verb inflection and temporal adverbs). L1 English speakers were asked to learn a number of Latin expressions and their English translations, which contained different kinds of temporal reference cues for past, present and future (adverbs such as *cras*, "tomorrow", and inflected verbs such as *cogitabo*, "I will think"). Before being trained with full sentences, they were pre-trained on either the adverbs or the verbal inflections, for past and present tense only. In a comprehension test, they were found to rely preferentially on the type of cue they had been trained on (as evidenced by bias in the case of conflicting cues). Strikingly, this was transferred even to expressions from a tense they had not being trained on (future tense) showing that they had learned a pattern of attention which they could apply to new input (N. Ellis and Sagarra, 2010, Exp. 1). In this sense, blocking is "the result of an automatically learned inattention" (N. Ellis, 2006a, p. 178).

Sensitivity to cues in L2 learning and processing is also affected by L1 (N. Ellis and Sagarra, 2010; Leung and Williams, 2014; MacWhinney, 2001; MacWhinney et al., 1984). In their study on the acquisition of Latin temporal adverbs and verbal inflection, N. Ellis and Sagarra (2010, Exp. 2) also compared English speakers to L1 speakers of Chinese, a language which does not have verbal inflection and relies on adverbs for temporal reference. Even if the two groups went through the same training, Chinese speakers were significantly more likely to rely on adverbs in a comprehension task, and more accurate in producing them than the English group (while they were less accurate in producing verbal morphology). Similarly, in MacWhinney et al. (1984), highly proficient German speakers of L2 English relied preferentially on animacy and verb agreement to identify the subject of a sentence, rather than pre-verbal placement (which is a primary cue in English, but not German), exhibiting what MacWhinney calls a "syntactic accent" (MacWhinney, 2001). Finally, in a study on the implicit learning of a classifier system derived from Mandarin Chinese, L1 Mandarin speakers showed evidence of learning, while L1 English speakers did not (Leung and Williams, 2014).

The shortcomings of naturalistic, exposure-only L2 learning have led authors such as Doughty (2003) to hypothesise that instruction may be "necessary to compensate for developmental changes that put adults at a cognitive disadvantage" (Doughty, 2003, p. 257). However, instruction can still be implicit in nature: the crucial thing is that is should direct attention to relevant form-meaning connections. Evidence from implicit learning and priming studies shows that it is possible to influence which contingencies associations participants become sensitive to, even while they remain unaware of them (Custers and Aarts, 2011; Jiménez and Méndez, 1999). Jiménez and Méndez (1999) conducted a SRT (serial reaction time) experiment in which participants saw symbols appear one at a time in one of four possible locations on screen, and had to press the button corresponding to that location as quickly as possible. The symbols used ("*",

"?", "!" and "x") appeared random, but there was a hidden connection between the shape of each symbol and the location of the following symbol in the sequence (e.g. whenever "!" was shown, the next symbol would appear in the rightmost location on screen). If participants were only told to pay attention to the position of the symbols and tap the corresponding key, they did not learn the hidden regularity. But if they were given a task which also required them to pay attention to symbol shape (e.g. counting how many times "*" appeared in the sequence), they became sensitive to the hidden regularity, as evidenced by a slowdown in reaction times if the regularity was violated. The study shows that being exposed to the stimuli was not enough for them to learn the regularity: in Schmidt's terms, they had to *notice* both character form and the position of the following stimulus in order to learn the association between the two, even if they did not have "awareness at the level of understanding" of the connection.

The findings by Jiménez and Méndez (1999) also show how attention can be selectively manipulated in implicit learning paradigms in order to promote learning of specific associations. The importance of directing attention to the relevant features of the input is also what motivates approaches to L2 teaching such as Processing Instruction (VanPatten, 1996; VanPatten and Cadierno, 1993), which aim to direct learners' attention to specific form-meaning contingencies in the input that participants may otherwise disregard. In PI, learners first receive explicit instruction on the target rule and then practice with a series of "input processing" activities, comprehension-based tasks (e.g. picture matching task) which require learners to attend to form-meaning connections specific to the target rule. Learners are also alerted to cognitive biases which may cause them to disregard less salient features of the input, as a way to direct their attention to such features. Processing instruction has been shown to be a highly effective methodology, leading to better performance on comprehension measures compared to traditional drill-based textbook methods (VanPatten, 2002; VanPatten and Cadierno, 1993). It is equivalent to traditional instruction on certain measures of production, such as sentence completion (VanPatten and Cadierno, 1993) and written narration tasks (VanPatten and Sanz, 1995). These results were replicated even when learners did not receive explicit rule instruction at the start (VanPatten and Oikkenon, 1996); this has led VanPatten to argue that the learning is a result of the way in which attention is directed during input processing activities, rather than conscious application of the rule during tests (VanPatten et al., 2013; VanPatten and Oikkenon, 1996). However, while it stresses the role of input processing activities over rule instruction, PI is still designed to be a type of explicit instruction (VanPatten et al., 2013); it does not seem to improve performance on measures of production that rely more on implicit knowledge, such as oral narration tasks (Marsden and Chen, 2011; VanPatten and Sanz, 1995).

Finally, it has been shown that prior explicit knowledge can have deep effects on L2 acquisition even if the rest of the training procedure is incidental; these effects may be detected by EEG, but not by behavioural measures (Batterink et al., 2015b; Morgan-Short et al., 2012a,b). Morgan-Short and colleagues (Morgan-Short et al., 2012a,b) trained participants on an artificial language, Brocanto2, similar in structure to Romance languages. Participants were trained over three sessions using either an “explicit” or “implicit” methodology. The explicit group were auditorily exposed to sentences in Brocanto2 together with explanation of its grammar, while the implicit group heard the sentences only (however, rule awareness was not assessed in this group; following the terminology used so far, the procedure they underwent would rather be classed as incidental learning). Following this initial phase, both groups engaged in communicative comprehension and production practice, where they had to play a chess-like board game by giving and receiving instructions in Brocanto2; corrective feedback was also provided. In the post-test, both groups performed similarly well in behavioural measures. However, only the “implicit” group showed L1-like ERP responses when exposed to ungrammatical Brocanto2 sentences, suggesting that the provision of explicit information at the start had fundamentally altered the way they had acquired the language, even when undergoing the same training procedure. Similarly, Batterink et al. (2015b) looked at the effect of explicit instruction administered prior to a statistical word segmentation task. They found that giving participants explicit instruction on the words prior to the task led to lower reaction times for predictable targets in a post-test; however, it also elicited a greater P300 response to the same targets, which suggests more controlled, effortful processing (Batterink et al., 2015b). From a memory system perspective, there is evidence that learning in declarative memory (which is promoted by explicit training) can inhibit procedural learning (Fletcher et al., 2005; Packard, 1999; Poldrack and Packard, 2003). Morgan-Short and colleagues suggest that this is implicated in their findings and that it may be related to blocking phenomena, where "...the retrieval of lexicalized knowledge (thought to rely on declarative memory) blocks the application of grammatical rules (thought to rely on procedural memory)" (Morgan-Short et al., 2012b, p. 12).

However, there is also evidence that explicit knowledge of one aspect of a language can speed up incidental learning of a different aspect of it. Monaghan et al. (2019) used a cross situational-learning paradigm to investigate the acquisition of an artificial language, comparing an entirely uninstructed condition to one where participants were given prior information on the syntactic structures of the language. The group who received explicit grammar instruction acquired the artificial language vocabulary (composed of novel nouns and verbs) earlier than the uninstructed group, showing higher accuracy during training. However, the effect was only temporary, and all participants eventually converged on similar levels of accuracy by the end of the training phase. Within the uninstructed group, participants who incidentally developed awareness of grammar rules also showed faster vocabulary learning compared to those who had remained unaware, but this difference eventually disappeared, too. Monaghan et al. (2019) then ran a second experiment designed to track the emergence of awareness during the learning phase under uninstructed conditions, which confirmed that the development of awareness was driving the increase in accuracy, rather than being the product of it.

In conclusion, attention is a crucial factor in both SLA and implicit learning. Implicit learning should not be conceived of as simple exposure to the stimuli, which, as we know from investigation of naturalistic L2 learning, is vulnerable to attentional effects such as blocking, overshadowing and L1 transfer. On the contrary, adequately designed implicit learning paradigms could help overcome these biases, directing attention to the relevant form-meaning connections while ensuring that learners engage with the material in a naturalistic, communicative way. In this respect, it may offer an advantage over a similarly attention-focused methodology such as PI, which includes highly targeted activities: the evidence reviewed above shows that focusing attention exclusively on one feature can be detrimental to the acquisition of other features (N. Ellis and Sagarra, 2010). Crucially, blocking effects may be difficult to spot in an instructed L2 setting, because they affect features other than the ones being taught and tested. For instance, if participants in N. Ellis and Sagarra (2010) had just been assessed on their ability to use and comprehend adverbs, as may often be the case in instructed L2 learning, they would have scored well and no detrimental effect would have been noticed. The other advantage of implicit learning over PI is the lack of explicit instruction: effects of explicit instruction on subsequent learning such as those reported by Morgan-Short et al. (2012b) suggest that developing explicit knowledge before exposure to the input may be detrimental to subsequent acquisition. For this reason, it may desirable to employ implicit learning paradigms specifically in the initial stages of exposure to an L2, focusing on directing attention to the relevant form-meaning connections rather than on overt rule explanation, which may be useful at a later stage. This is supported by evidence from an AGL study by Mathews et al. (1989), which shows that the best results in an AGL task with a biconditional grammar were obtained with an initial phase of implicit learning followed by a phase of explicit (rule-search) learning. In fact, the learning gains obtained in this way were greater than the sum of the scores obtained using either implicit or explicit learning alone (Mathews et al., 1989); this suggests that a synergy exists between implicit and explicit learning, which may also be beneficial to L2 acquisition. Finally, PI is based entirely on comprehension; perhaps unsurprisingly, one of its weaknesses seems to be a limited capacity to improve performance on implicit production tasks. The relation between production and comprehension, and how they interact in training and testing, is often not explicitly addressed in the literature on implicit learning and L2 acquisition, where the emphasis is on distinguishing between implicit and explicit abstract knowledge. However, it is of great consequence for L2 learning, and should be taken into consideration when designing implicit learning paradigms, too. The next section provides an overview of the role of production in SLA and in implicit learning, and then focuses on the question of whether the latter could be used to generate implicit productive knowledge in L2.

2.5 Production in SLA and implicit learning

2.5.1 Modelling the role of production in L2

The most commonly used framework of production in L2 research is Levelt's model of speech production (Levelt, 1993, 1995, 1999, 1989). It is made up of two principal components, the *rhetorical/semantic/syntactic system* and the *phonological/phonetic system*, each including sub-components responsible for different phases of speech production; syntactic and phonological rules are automatised and part of the encoding systems (Levelt, 1989). The model was expanded by De Bot (1992) and later further modified by Kormos (2009) to adapt it to bilingualism. We will refer here to Kormos' version (Fig. 2.1). In the first component, conceptual preparation (which has access to general and contextual world knowledge) takes place; the resulting "preverbal message" is then grammatically encoded, by retrieving lemmas and relevant syntactic information from the mental lexicon. The output of the first component is then fed into the second component (the phonological/phonetic system), where morpho-phonological encoding takes place, also drawing on information retrieved from the lexicon. This results in a "phonological score" (internal speech) which is then phonetically encoded by retrieving articulatory gestures from a dedicated store (the "syllabary"), and the resulting "articulatory score" is then used for articulation. The sub-component responsible for conceptual preparation can monitor the subsequent steps through a number of feedback loops: specifically, it can check the preverbal message, the phonological score (produced by morpho-phonological encoding) and the final speech output. In addition to world knowledge, the lexicon and the syllabary (which are shared between L1 and L2), Kormos' version of the model (Kormos, 2009) also includes a store of declarative L2 rule knowledge, which can be accessed during both grammatical and morpho-phonological encoding (Kormos, 2009, p. 168). Since the lexicon and syllabary are shared between the L1 and L2, L1 transfer errors can occur if the syntactic and lexical information for grammatical encoding is retrieved from the L1 lemma instead of the L2. For instance, Hungarian speakers of English may often say "enter into a room" because they transfer the VP + PP structure that the Hungarian equivalent of "enter" points to (Kormos, 2009,

p. 172). In the next phase of grammatical encoding, which involves phrase and clause building, L2 learners may deviate from native speakers in various ways. L1 and advanced L2 speakers will rely on procedural knowledge of syntactic and morphological rules, and apply those rules automatically. By contrast, lower proficiency L2 speakers may have declarative knowledge of the rules, and use them consciously. If they have no knowledge of the rules in any form, they may use communicative strategies such as deliberate transfer of L1 structures or simple juxtaposition of lexically encoded concepts (Kormos, 2009, p. 172).

The presence of loops monitoring internal and external speech allows for hypothesis testing and can aid noticing of errors in the learner's own output, such as in case of a mismatch between preverbal message and phonological score (De Bot, 1996). According to the Output Hypothesis (Swain, 1985, 2005; Swain and Lapkin, 1995), this property of speech production makes it a valuable part of L2 learning. Production contributes to the development of accuracy, by drawing attention to form and forcing learners to process language at a deeper level than input does: output may stimulate learners to move from the semantics-based processing prevalent in comprehension to the complete grammatical processing needed for accurate production, meaning that output plays a significant role in the development of syntax and morphology (Swain, 1993). It also allows learners to "notice" gaps in their L2 (Schmidt and Frota, 1986), which in turn enables them to acquire the correct target form. This applies both to output generally and to "pushed" output, that is, attempting to produce output slightly above one's own capabilities (Swain, 1985). Pushed output is usually obtained by means of a text reconstruction or "dictogloss" task, in which students are asked to take notes while listening to a text, and then reconstruct it (Kowal and Swain, 1994, 1997; LaPierre, 1995; Wajnryb and Maley, 1990). Studies conducted by Izumi et al. using this paradigm show that it leads to improvements in both comprehension and untimed written production, which was tested with a sentence combination task (combining two sentences into a main and a relative clause) and a cued sentence completion task (Izumi, 2002; Izumi et al., 1999). These production tasks are quite explicit in nature; as already argued by LaPierre (1995), this suggests that the benefits of pushed output through text reconstruction are largely due to increased awareness (and therefore, explicit knowledge), which supports the

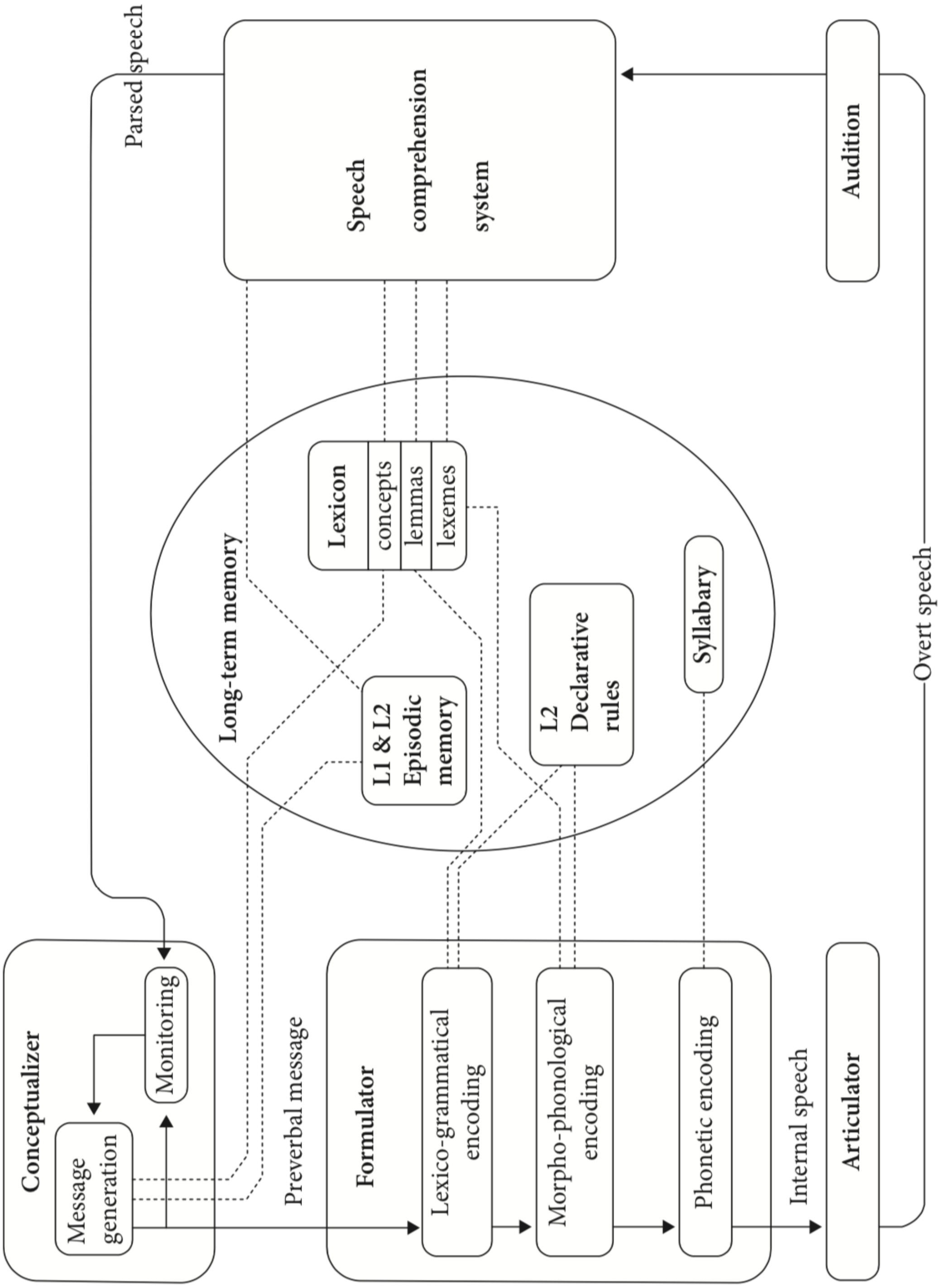


Figure 2.1: Kormos' model of bilingual speech production (Kormos, 2009, p. 168). Copyright ©2006 by Lawrence Erlbaum Associates, Inc. Used by permission.

Output hypothesis (Swain, 1985, 1993).

Besides raising awareness, it is widely acknowledged that productive practice can improve speed and fluency in production (Gass and Mackey, 2007; Lee and VanPatten, 2003; Schmidt, 1992b; Swain, 1985; R. Ellis, 1988). However, whether this amounts to creating new implicit knowledge is a point of contention, which hinges on the nature of implicit knowledge and of the relation between production and comprehension.

2.5.2 Relation between production and comprehension

Production practice can help generate awareness and improve accuracy on explicit tasks, and it can help improve fluency. But can it generate new implicit knowledge, which could then be used in production? De Bot (1996) argues that learning during production can also take place through proceduralisation, as connections between lemmas and procedures are strengthened by being repeatedly made in the grammatical encoding phase. Repetition priming effects in production have been observed (Lupker, 1988; Mitchell and Brown, 1988), with reductions in picture naming latencies observed even several months after priming (Cave, 1997). The duration of the priming effects observed has led some to argue that these should be viewed as learning effects rather than just priming (Griffin and Ferreira, 2006). In fact, in usage-based views of language acquisition (N. Ellis, 1998; Barlow and Kemmer, 2000) priming and learning are fundamentally part of the same mechanism, on the assumption that "short-term effects lead to long-term effects [...] as the individual increments of learning integrate over time to form the processes, representations, and attentional biases that constitute human minds." (N. Ellis and Sagarra, 2010, p. 570).

However, in the field of implicit L2 learning research, few studies to date have targeted production measures (for an exception, see De Jong (2005)). In IL studies those that did include measures of production, the aim to was to evaluate abstract rule knowledge, rather than productive skills *per se* (Hama and Leow, 2010; Ruiz et al., 2018; Wonnacott, 2011; Wonnacott et al., 2017) or to compare explicit (instructed) learning to learning under incidental conditions (DeKeyser, 1995; Doughty, 1991; Goo et al., 2015; Norris and Ortega, 2000; Spada and Tomita, 2010). Ruiz et al. (2018) adapted the implicit learning paradigm from Rebuschat (2008) and Rebuschat and Williams (2012), using an artificial language made of English vocabulary arranged following German word order rules (unlike previous studies, each sentence also contained a novel pseudoword). Training was based on reading for comprehension: one group (*incidental* group) was only tasked with judging semantic plausibility and learning the novel pseudowords; the second group (*intentional* group) was additionally instructed to discover the word order rules. Testing included a 4-alternative forced choice task (4AFC), a grammaticality judgment task and a written sentence production task. While both the 4AFC and grammaticality judgment task showed evidence of learning, accuracy on the production task was generally low, both for the incidental group (26%) and for the intentional one (44%). Similarly, studies comparing implicit with instructed learning have found little to no learning gains when participants were tested on production tasks following implicit instruction. However, two key factors were generally not controlled in these studies: the type of production task (tapping into implicit or explicit knowledge) and the type of training activity (comprehension or production-based). For instance, DeKeyser (1995) compared implicit learning of an artificial language to instructed learning; for the implicit group, DeKeyser found no evidence of learning in a written production task, even after lengthy exposure. However, participants were trained on comprehension, but tested on production: they were evaluated on a task best suited for assessing explicit knowledge, and in a modality other than the one they were trained in. Therefore, it is difficult to establish whether the outcome is due to the fact that the training procedure did not generate any implicit knowledge, or whether implicit knowledge acquired through comprehension cannot be deployed in a production task.

To know whether we can directly create implicit knowledge that can be used in production, we need to know what the nature of that knowledge is. Does the same abstract knowledge underlie both comprehension and production? In L2 acquisition, it is common for comprehension skills to develop before production. This is partly because comprehension can rely on semantic and pragmatic cues to compensate for gaps in syntactic (Clahsen and Felser, 2006, 2018) and morphological (Clahsen et al., 2010) knowledge. However, the same is not possible for production. There are two possible explanations for this: either comprehension and production rely on the same knowledge, but production posits additional processing difficulty, or they are separate skills.

Some researchers in the field of SLA have assumed that both L2 comprehension and production rely on the same abstract knowledge (Bates and MacWhinney, 1989; Krashen, 1985; Schwartz, 1993; VanPatten and Cadierno, 1993), even though they acknowledge that the requirements of actual processing may differ considerably between modalities (Bates and MacWhinney, 1989). The assumption is that production requires the same basic knowledge or skill as comprehension, with the addition of extraneous factors such as word retrieval and motor skills, which may mask the true extent of linguistic knowledge. Comprehension, under this view, is a "purer" measure of linguistic knowledge, offering a window into abstract language knowledge; while production can be affected by lexical access or word retrieval problems which can make it hard to gauge the actual extent of rule knowledge (Chondrogianni et al., 2015). The hypothesis of shared representations is supported by syntactic priming: comprehension of specific syntactic structures can prime their production (Bock et al., 2007; Branigan et al., 2000, 1995; Potter and Lombardi, 1998), and vice versa (Branigan et al., 1995). There is evidence that syntactic priming modulates activity in the same brain areas (inferior frontal gyrus, left middle temporal gyrus and bilateral supplementary motor area) and to the same extent, regardless of whether it occurs within modalities (e.g. production to production) or across them (Segaert et al., 2012). Morphosyntactic and lexical properties such as gender and number can also be primed across comprehension and production (Pickering et al., 2000).

Others, most notably DeKeyser (DeKeyser, 2017, 1997; DeKeyser and Sokalski, 1996; Li and DeKeyser, 2017) have taken the view that comprehension and production are separate skills, and develop in accordance with the power law of skill acquisition (Anderson, 1993): once proceduralised, knowledge is highly skill specific. This predicts that production training will lead to gains in production only, and comprehension tasks will likewise only lead to improvement on measures of comprehension. These different positions on the relation between production and comprehension generate different predictions with regards to our question: in Krashen's view, comprehension-based training should be enough to generate implicit knowledge that can be used in production. In DeKeyser's view, production practice should be needed to develop implicit productive knowledge through proceduralisation.

2.5.3 Can we generate implicit productive knowledge?

According Krashen's Input Hypothesis, input is what drives acquisition: specifically, comprehensible input that is slightly more complex than the learner's current level, or "i+1", where i is the learner's current state (Krashen, 1989, 1994, 1981b, 1982). In this view, the only role output can serve in acquisition is that of generating self-input for the learner. Similarly, VanPatten argues that Input Processing activities are sufficient to develop both comprehension and productive skills. While he acknowledges that output practice is useful for developing fluency at a later stage (Lee and VanPatten, 2003), VanPatten maintains that production practice in the early stages of acquisition may in fact be detrimental: it would be like "putting the cart before the horse" by requiring learners "to produce when the developing system has not yet had the relevant intake data" (VanPatten and Cadierno, 1993, p. 436). In a similar vein, it has been pointed out that practice may enhance fluency, but it does not necessarily improve accuracy (R. Ellis, 1988; Schmidt, 1992b). In Krashen's and VanPatten's view, accuracy in production can be obtained simply through adequate exposure to input.

A number of studies comparing Processing Instruction (PI) with output-based tasks have shown that students trained with PI performed better than output groups on interpretation tasks, and on the same level as output groups in production tasks (Cheng, 1995; VanPatten and Cadierno, 1993; VanPatten and Sanz, 1995; VanPatten and Wong, 2004), which would seem to support the hypothesis that input practice is sufficient for developing competence in production. VanPatten and Cadierno (1993) investigated the acquisition of Spanish word order by L1 English learners. They trained one group using input processing – a methodology that involves explicit explanation and a range of comprehension tasks forcing subjects to attend to specific form-meaning connections, but no production tasks. They compared this to a second group who received traditional instruction based on a current university textbook, and to a control group who received no instruction. After two days of instruction, participants were tested on a comprehension task (picture matching) and a production task (gap filling). They found that only the Input Processing group showed significant improvement in the comprehension task, relative to pre-test. On the production task, both groups showed the same improvement, which is striking because the PI training protocol did not include any production task, unlike the traditional instruction group. VanPatten and Cadierno conclude from this that input processing has a significant effect on learners' L2 development, which is detectable in output as well as input, despite the fact that learners did not produce any output during training (similar results were obtained by Cheng (1995)). However, they used a very constrained written production task (gap filling), with no time constraints. To examine what the outcome would be in a freer, more communicative context, VanPatten and Sanz (1995) carried out a partial replication of VanPatten and Cadierno (1993) using two groups (one input processing, one control) but with three tasks for testing production: the sentence completion task used in VanPatten and Cadierno (1993) and a video narration task, administered both orally and in writing. They found that, in the written mode, the PI group significantly improved on both the sentence completion and video narration task. However, in the oral mode they only showed significant improvement in the sentence completion task and not in the video narration one (arguably, the only one requiring fluency). Likewise, in VanPatten and Uludag (2011), participants who were trained with PI and were then tested on a text reconstruction task showed no difference in performance from participants who were trained with a dictogloss task, which promotes explicit knowledge. Similar results were obtained by Erlam et al. (2009) in a study on the role of output in the acquisition of the definite/indefinite/zero article alternation in English, which incorporated measures of both explicit knowledge (untimed grammaticality judgment task (UGJT)) and implicit knowledge (elicited oral imitation task (EOIT)) in testing . Subjects were given explicit instruction on the rule and then performed a battery of tasks, either production- or comprehension-based: production activities consisted of free meaning-oriented practice with target structure, while comprehension activities were based on PI protocols (VanPatten and Cadierno, 1993). Both groups made significant improvements on both UGJT and EOIT, which suggests both treatments had increased implicit as well as explicit knowledge. However, only the production group performed significantly better than a control group on the EOIT, and only they showed generalisation to new items in a delayed EOIT two weeks later, which suggests that output activities were instrumental in developing productive implicit L2 knowledge. Overall, the evidence covered so far suggests that input training can lead to improved performance on explicit production tasks, but is not as successful with productive tasks that require automatic knowledge, as also noted by Marsden and Chen (2011). Instead, production activities seem better suited to develop productive automatic L2 knowledge.

The importance of production activities to develop productive skills is also stressed by DeKeyser (DeKeyser, 2007, 1997; DeKeyser and Sokalski, 1996), who argues that production and comprehension do not operate based on the same abstract information, but are separate skills. Production training is not simply useful for developing fluency, but it is necessary to develop implicit productive knowledge (which, in DeKeyser's framework, is obtained through the automatisation of declarative knowledge). To test this hypothesis, DeKeyser (1997) investigated the acquisition of an artificial language (Autopractan) over 8 weeks of training: participants were trained with either comprehension tasks only, production tasks only or with a mixture of both. Comprehension was trained and tested with a picture-matching task; production, with a written production task in which participants had to type sentences cued a picture. Each task was administered under both single- and dual-task conditions, the latter to tap into automatic processing. Participants were significantly better in the practiced skill than in the reverse, with regards to both speed and accuracy. They still showed some improvement in the unpracticed domain, however, even if not as great as the practiced one. According to DeKeyser (1997), this is due to the fact that performance in the trained domain was driven by proceduralized knowledge, while transfer to the other domain relied on declarative knowledge. However, there was one exception to this pattern: the production-only and mixed groups actually had higher accuracy than the comprehension group on one measure of comprehension, a picture-matching task done under dual-task conditions (designed to tap into automatic knowledge). DeKeyser attributes this result to the task having low reliability, rather than to an effect of production training, however, it suggests that not only is production training required to develop implicit productive knowledge, but that it may be beneficial for comprehension, too.

Taken together, results from VanPatten's and DeKeyser's studies suggest that while transfer between comprehension and production is possible, production training may be required to develop implicit productive knowledge. On the one hand, results from VanPatten and Oikkenon (1996); VanPatten and Sanz (1995); VanPatten and Uludag (2011) show that comprehension-based practice can lead to productive performance on explicit, but not implicit, measures; this is compatible with findings by DeKeyser (1997) and DeKeyser and Sokalski (1996), who found that performance on production measures, under both single- and dual-task conditions, was better for subjects who received productive training on the relevant structure. DeKeyser's findings also show evidence of transfer between comprehension and production, which is compatible with shared representations. However, production practice actually led to higher accuracy than comprehension-based practice on an automatic comprehension task; the reverse was not true. This is compatible with shared representations, but it also suggests a primacy for production, providing support for models that see production and comprehension as instances of action and action perception respectively (Dell and Chang, 2014; Pickering and Garrod, 2013). In Pickering and Garrod's (2013) integrated model of comprehension and production, the two processes are interwoven in language use. Specifically, production is recruited during comprehension as comprehenders make "covert models" of production during comprehension, which allows them to make predictions about upcoming input. Based on this model, we would expect the development of production to have an effect on comprehension - not just due to shared representations, but because production is harnessed by comprehension. Indeed, EEG studies on implicit learning which found learners had developed L1-like processing of a miniature artificial language, when tested with a comprehension task (GJT) (Friederici et al., 2002; Morgan-Short et al., 2012a,b), used a communicative production task to train participants, which lends some support to this hypothesis.

2.6 Research questions

Research on implicit learning in SLA has shown that it is possible for learners to acquire both novel forms and form-meaning connections from the input in an implicit way. However, evidence for this claim comes primarily from studies using comprehension tasks to test for implicit knowledge. Little attention has been devoted so far to the relation between comprehension and production in implicit learning, focusing instead on the acquisition of abstract implicit knowledge, assumed to underlie both processes.

Our first research question was whether it would be possible for learners to develop implicit knowledge of a novel linguistic rule and use it correctly in production. Based on the the evidence reviewed here, we hypothesised that it would be be possible for the knowledge acquired through implicit learning to be used productively, provided that en element of production training were included in the learning phase. To test our hypothesis, we devised a novel learning and testing methodology based on the elicited oral imitation task, in order to have a production task that would focus attention on form-meaning connections without encouraging the application of explicit rules.

Our second research question was whether developing implicit knowledge through a production task would also lead to learning gains in comprehension. Here, different theoretical positions predicted different outcomes, allowing us to test different accounts of the relationship between comprehension and production. If both production and comprehension relied on the same abstract knowledge, with additional processing demands for production, we hypothesised that that the knowledge acquired through a production task should immediately be available in comprehension, too. On other hand, if production and comprehension were different skills relying on different kinds of automatised knowledge, we would not expect the knowledge acquired in a production task to readily transfer to a comprehension one, without comprehension-based training. To address this question, we included comprehension tasks in the study, specifically a reaction time listening task (Experiment 1) and a recognition memory task (Experiments 2 & 3).

Chapter 3

Experiment 1: Elicited recall of novel prepositions

3.1 Introduction

As discussed in the previous chapter, the existing literature suggests that it may be possible for learners to acquire implicit knowledge that they can then use in production. To test this hypothesis, we turned to a feature of language that commonly presents problems for adult L2 learners, namely prepositions. As function words with limited semantic content, prepositions are often neglected by L2 learners (VanPatten and Oikkenon, 1996). Furthermore, their correct usage often depends on semantic properties of the content words they are paired with, rather than just the lexical meaning of the prepositions themselves. For this reason, they may be particularly suited to benefit from a training paradigm which directs learners' attention to relevant form-meaning connections and does so in a holistic manner, requiring participants to pay attention to the association of prepositions with different nouns, rather than just to the prepositions themselves. Section 3.1.1 provides an introduction to the materials we used as stimuli, which were based on a pair of spatial prepositions found in Czech, *v* and *na*.

The literature on implicit learning and production which we surveyed suggests that, in order for learners to develop productive implicit knowledge, production-based training may be necessary. At the same time, it is crucial that the task used to test learners should be well-suited to detect implicit knowledge. Therefore, in order to evaluate production skills, we needed a productive task that could directly tap into implicit knowledge, minimising interference from conscious control. In order to meet both requirements - the need for production-based training, and for a suitable test of implicit knowledge - we decided to use an elicited oral imitation task, which would serve as both training and testing methodology. It is widely accepted that the EOIT is a suitable test of implicit knowledge (see Section 2.2.1); this chapter (Section 3.1.2.1) will lay out the rationale for using the task as a learning tool, too. Finally, the existing literature also suggests that knowledge developed through implicit learning may transfer to a comprehension task, although the nature of that transfer process is not clear. If participants are able to acquire abstract, implicit rule knowledge through the production task, we may expect to them to display sensitivity to the rule in a comprehension test, too. On the other hand, if transfer between modalities relies on explicit declarative memory, as DeKeyser (1997) argues, we may expect to see a dissociation between productive and receptive measures of implicit knowledge, with participants scoring above chance in a production task, but not in a comprehension one. We selected a reaction time (RT) task as measure of receptive implicit knowledge, which will be described in Section 3.1.2.2. Section 3.2 will then present the first experiment we ran using this methodology, which aimed to test the effect of a hidden regularity on participants' recall accuracy.

3.1.1 Czech spatial prepositions: *v* and *na*

As the basis for our study, we selected a specific set of spatial prepositions found in Czech: *v* and *na*. In Czech, as in other Slavonic languages, the same spatial preposition may exist in different forms, which usage depends on the physical properties of the space being referred to. Both *v* and *na* indicate a stationary position (which may be variously translated as "at", "in", or "on"), but they are used before different entities. Nouns denoting open spaces tend to be preceded by the preposition *na*, whereas enclosed spaces are normally preceded by *v*. In this sense, there is partial overlap between the Czech rule and the English *in/on* distinction: generally, nouns that take *on* in English are likely to take *na* in

Czech (e.g. *na balkoně*, "on the balcony"). However, the reverse is not true, as the range of nouns that take *na* in Czech is much wider than that (e.g. *na hřišti*, "in the playground" or *na koncertě*, "at a concert"). In fact, it is not merely a distinction between open and enclosed spaces. It is a fuzzy category, which can fully be captured only by a variety of rules, some of which may at times conflict with each other. Hrdlička (2000) identifies tendencies driving usage, based on properties of the physical location being referred to (e.g. how bounded or hidden it is, whether it is lower or higher then the surroundings), "social importance" (e.g. institutions), type of locality (administrative unit, etc.), physical properties of a substance / locality (e.g. water, earth), and intensity of contact between agent and patient. The complex nature of this category makes it a common stumbling point for L2 learners of Czech, to whom the distinction often appears arbitrary. Conversely, native speaker intuitions on this subject are usually strong and uniform, even for novel items such as newly introduced loanwords. We can also identify certain morphological regularities among nouns, which may serve as cue for the *v/na* distinction. For instance, nouns ending in *–iště* (e.g. *hřiště*, "playground"; *parkoviště*, "car park"), which usually denote open spaces, overwhelmingly tend to use *na* as preposition. Conversely, nouns ending in *–árna*, which normally denote enclosed spaces (e.g. *kavárna*, "café"; *čekárna*, "waiting room"), will usually occur with *v* (Lukšija, 2010). It is possible that native speakers rely on implicit knowledge of these distributional properties when acquiring the rule, as well as semantic properties of the nouns. As the examples above show, the *v/na* alternation in its original form is a very complex rule, where semantic generalisations interact with both pragmatic and morphological factors. Therefore, in order to avoid confounds and ensure that participants were sensitive to the basic spatial distinction underlying the rule, we thought it necessary to begin by testing a simplified version of it. Therefore, we decided to focus on the core part of the rule, namely the physical distinction between open/outdoor and enclosed/indoor spaces. Place nouns used for each category were selected based on frequency and natural usage in Czech; more details on how study materials were prepared can be found in Section 3.2.2.

3.1.2 Methodology

3.1.2.1 Using elicited oral imitation as a learning tool

As mentioned previously, the EOIT is commonly used as a test of implicit knowledge in L2 research. However, elicited imitation has also been used in implicit learning research, as a way to get subjects to engage with the training material on a deeper level than could be achieved by simple exposure. In Leung and Williams (2006, 2011), participants were auditorily exposed to sentences in a semi-artificial language (English lexis with novel case markers, and either VSO or VOS word order) while performing a RT task. After giving their response, they had to reformulate the sentence they had just heard using normal English word order, but retaining the case markers; every two trials, they had to recall these reformulated versions. According to the authors, this was done to "stimulate concurrent active processing of the action portrayed and the article used" (Leung and Williams, 2006, p. 467). Similarly, in a study on the implicit learning of novel syntactic structures by Rebuschat and Williams (2012, Experiment 2), the addition of an elicited imitation task during training improved learning outcomes. Rebuschat and Williams attribute this to the fact that the addition of a recall task required participants to process word order more directly, thus facilitating the acquisition of the different syntactic structures.

Elicited oral imitation combines a number of different processes which have been shown to be beneficial for learning: memorisation, retrieval and production. The use of memorisation as a way to get subjects to engage with stimuli has been a common feature of studies on artificial grammar learning: in a typical AGL experiment, participants are instructed to memorise strings, under the pretence that they are doing a memory task (Mathews et al., 1989; Reber, 1976). Memorisation has been shown to yield better learning effects than rule-search (Reber, 1976, 1989), which suggests that it offers benefits beyond those stemming from simple exposure to the material. It appears to be a kind of "desirable difficulty" (Bjork, 1994) which can aid processing, by forcing participants to process material in greater depth than they would if they were just reading for comprehension. Greater depth of processing can lead to better item recall even when memorisation is not the goal of the task. For instance, when reading an ambiguous sentence, it has been shown that making an effort to comprehend the sentences for several seconds before receiving a disambiguating cue enhances sentence recall later on, compared to receiving the cue immediately, even if it does not affect comprehension scores (Zaromb et al., 2010).

Retrieval practice, too, has been shown to be beneficial for learning, a phenomenon known as the "testing effect". Testing learners on newly acquired information improves retention of that information, more so than simply giving subjects the opportunity for restudy (Carpenter and DeLosh, 2005; Cull, 2000; Karpicke and Roediger, 2008; **?**). The beneficial effect of testing is not just due the repeated exposure subjects receive during testing: for instance, Zaromb and Roediger (2010) show that, if we substitute learning episodes with testing (free recall), performance at delayed testing significantly increases, suggesting that the effect is not simply due to repeated exposure to the material. Crucially, the benefits of testing are found even if learners do not receive any feedback. In a typical study, participants may be tested on word lists (Karpicke and Roediger, 2007; Zaromb and Roediger, 2010), and simply be asked to recall as many words as they can remember; or, if they are asked to learn word translation pairs (Karpicke and Roediger, 2008), the testing phase may just require them to type the English translation. Either way, they are not told whether the response is correct; the effort of retrieval is what seems to drive the learning effect. According to the "elaborative retrieval" hypothesis, this effect of immediate memory testing on long-term retention occurs because, as subjects search they memory for the required piece of information, "the information activated during retrieval may spread to other related concepts and eventually activate an elaborative semantic network with multiple pathways leading to the correct target." (Carpenter, 2009, p. 1564). This would not happen during restudy, as the information would be readily available. On the other hand, we would expect this effect to be triggered during elicited oral recall, which will require participants to activate the relevant information even in the absence of further exposure.

While the testing effect has been shown to enhance retention for individual items, the aim of this study is not for participants to just learn individual items: the focus is on whether they can extract patterns of regularities from the input and use them productively. However, following a usage-based account of L2 acquisition, we assume that learners will first need to store individual representations in memory in order to abstract from them; therefore, a paradigm which aids the retention of individual items should ultimately also aid rule extraction. Indeed, evidence suggests that memorisation and recall can support the acquisition of abstract regularities, too. In a study on the acquisition of novel affixes, cued recall was used by Tamminen et al. (2015) to promote the learning of stem + affix combinations; participants were then able to generalise the affixes to new stems in a comprehension task, showing that they had abstracted away from the individual combinations they were taught. Based on the regeneration hypothesis of sentence recall (Potter and Lombardi, 1990), we also hypothesise that the memory load coming from having to recall large amounts of material will provide an incentive to decode and re-encode sentences wherever possible, rather than relying solely on phonological working memory, thus making subjects more likely to extract the regularity from individual items. Research on rule induction by Radulescu and colleagues (Radulescu, 2014; Radulescu et al., 2019) provides some support for our hypothesis: it suggests that the extraction of regularities from a string of letters is a function of channel size (memory capacity) and stimulus complexity. In two AGL studies using strings of syllables as stimuli, increasing internal string complexity (e.g. the number of possible X syllables in a string of type XXY) made participants more likely to extract categorical generalisations from the input (Radulescu et al., 2019).

In this study, we will attempt to put pressure on memory capacity not by increasing stimulus complexity, but rather by increasing the amount of material to be stored in short-term memory. Memory load is a crucial component of the EOIT: in order for the task to be effective, it should not allow subject to rely entirely on phonological memory, so as to tap any internalised grammatical knowledge. To achieve this, researchers can manipulate various factors including stimulus length, time pressure, and interval between stimuli (Erlam, 2009). However, once parameters such as sentence length and the time interval are set, they will be the same for all participants. This means that individual differences such as working memory may still play a role, making it difficult to determine the relative contribution of phonological short-term memory across subjects. Our experiment will circumvent this problem by using a within-subject design, where each participant will be given the exact same set of sentences: one half of the sentences will follow the Czech preposition rule, while the other will not. If participants have acquired the rule, it should be possible for them to re-analyse the rule-based sentences, which in turn should make it easier to recall them, compared to those with no underlying rule. The task we developed will be described in greater detail in Section 3.2.3.2.

3.1.2.2 Measuring receptive knowledge

In order to explore the relation between production and comprehension in the development of implicit rule knowledge, our study will also include a measure of receptive knowledge. Similarly to the production task, the comprehension task should be able to tap into any existing implicit knowledge, without prompting rule searching. Following Leung and Williams (2006, 2011), we selected a reaction time (RT) task based on aural presentation of the sentences, which will require participants to give speeded responses based on information that can be inferred from the critical items. The assumption behind this task is that, in a time-pressured situation, participants will use any cues from the input that may allow them to carry out the task as quickly as possible. In rule-based sentences, the specific preposition used will depend on place type, thereby offering a cue as to which type of noun will follow. In the other sentences, by contrast, there will be no such relation, giving the listener no means to predict which noun will follow. Therefore, a subject who has acquired the underlying rule could be expected to react faster to the rule-based sentences than to ones with no rule. Details of the task can be found in Section 3.2.3.2.

3.1.2.3 Measuring awareness

As means of assessing awareness of the Czech preposition rule, we decided to use retrospective verbal reports, which will be collected by means of a debriefing questionnaire at the end of each experiment. While the use verbal report as a measure of awareness is not without drawbacks, we judged it to be the most appropriate measure for our task. We decided not to use a now commonly employed measure, judgement source attribution (Dienes and Scott, 2005), because it would not be suitable for our paradigm. As pointed out by Batterink et al. (2015a), the guessing and zero-correlation criteria which underpin the use of source attribution apply to *judgement knowledge*, defined as the ability to recognise whether a particular test item has the same structure as training items (Dienes and Scott, 2005). That is not not applicable to our task, which involves cued recall and production, rather than judgement. Pragmatically, asking for the source of each response in a recall task would not be appropriate, since the premise of the task is that participants are relying on their memory to produce the sentences. Furthermore, it may make participants more self-conscious and less spontaneous in their behaviour, which may be particularly detrimental in a production task; it might also make them more inclined to search for rules. In order to avoid the problem of under-reporting, our questionnaire includes multiple indirect questions; for instance, participants will not only be asked whether they think there were any rules underlying the use of pseudowords, but also to attempt a translation for them. A copy of the questionnaire used can be found in Appendix A.

3.2 Experiment 1

The aim of this first experiment was to explore the effects of incidental exposure to a novel rule on subsequent recall, as an initial measure of production. In order to make it as similar as possible to a naturalistic L2 learning situation, we selected a rule borrowed from a natural language, Czech, keeping only its core semantic distinction between open and enclosed spaces. When choosing tasks to test production and comprehension, the main criterion was that they should be able to tap into implicit knowledge, while discouraging rule searching and minimising the chances of participants relying on explicit knowledge during testing. We settled on a time-pressured elicited oral imitation task (Erlam, 2006; 2009), and on a reaction time task (RT) following Leung and Williams (2006).

If participants acquire implicit knowledge of the rule, we would expect them to make fewer recall errors when recalling rule-based items compared to the other items, even if they have no explicit knowledge of the rule. Likewise, if participants acquire implicit knowledge that can be used in comprehension, too, we would expect them to be faster when responding to items which follow the rule, and are therefore potentially predictable, compared to those which do not.

3.2.1 Participants

28 native English speakers aged 18-45 were recruited from the University of Cambridge and surrounding community (17 females, mean age 24, SD = 5.92), and received £8 as compensation for taking part in the study. Most participants had knowledge of at least one foreign language, but only two subjects reported having any knowledge of a Slavonic language (Russian), which could have provided an advantage due to its similarities with Czech. Other participants reported knowledge of languages belonging to the Romance (n = 32), Chinese (n = 11), Germanic (n = 8) and Indo-Aryan (n = 6) families, as well as Greek (n = 4), Malay (n = 2), Korean (n = 1), Navajo (n = 1) and Hawaiian (n = 1).

3.2.2 Materials

We first extracted lists of the nouns used most frequently in Czech with either *v* or *na*, using the SYN2015 corpus of written Czech as source (Křen et al., 2015). From these, we eliminated nouns that did not conform to the simplified rule we were going to use, such as: names of countries being used with *v*, or names of institutions being used with *na*. We kept only those nouns referring to physical locations (large enough for a person to occupy them), which could clearly be ascribed to either category by virtue of their physical properties (enclosed vs. open space). We translated the nouns into English and ranked them based on their frequency in the British National Corpus (2007). The 32 most frequent nouns were kept for each category, for a total of 64 unique place nouns (mean frequency 42.5 per million words). The full list of place nouns used and corresponding pictures is provided in Appendix C.

Place nouns were embedded in simple sentences, all with the structure *subject* – "is" – *preposition* – *place noun*, e.g. "Harry is *gi* desert". The subject was always one of two characters, "Harry" or "Lucy" (which corresponded to a stick figure drawing of either a male or a female), randomly assigned at each trial. Four pseudowords ("gi", "ro", "wa" and "ne") were used as prepositions, and for each participant they were randomly assigned to one of two conditions –

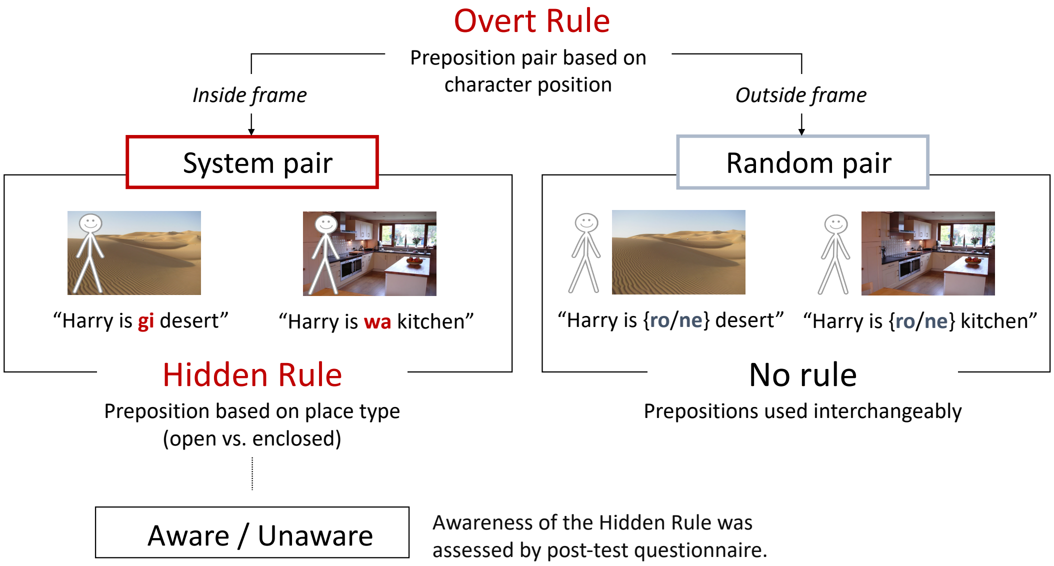


Figure 3.1: Description of rules used in experiment (Overt and Hidden)

System or Random. The two prepositions in the System condition followed the simplified Czech rule: one preposition could only be used with open space noun, the other with enclosed space ones (Overt rule). The two prepositions assigned to the Random condition, on the other hand, were used interchangeably (Fig. 3.1). Each Random preposition appeared with both types of noun over the course of the experiment, an equal number of times with each place type. The assignment of specific prepositions to the Random and System condition was counterbalanced across participants.

Audio recordings of the sentences were produced using Apple's OS X 10.11 built-in VoiceOver voice synthesizer. All place nouns as well as other sentence parts were recorded separately. Subject + predicate pairs ("Harry is" / "Lucy is") were recorded as one unit and edited to have the same length (0.812s). The four prepositions were recorded individually and were also of equal length (0.447s). Sentences were then formed online by concatenating the audio clips during presentation. In this way, we ensured that all critical word onsets (prepositions and place nouns) occurred at the exact same time point for all sentences. During training, each sentence would be accompanied by a visual representation, composed of a drawing representing either Harry or Lucy, and a picture depicting the place noun. Place pictures were sourced from the ImageNet (Deng et al., 2009) and SUN (Xiao et al., 2010) databases. The place picture always occupied the centre of the screen, while the position of the character drawing varied depending on sentence type. In System sentences (which followed the simplified Czech rule) the character was superimposed on the left side of the picture, whereas in the Random condition the character appeared beside the picture – this was the Overt rule. Participants were invited to pay attention to the position of the character on screen, encouraging them to discover its association with a particular pair of prepositions.

3.2.3 Procedure

Participants were seated at approximately 60cm from a 19" LCD screen, which was connected to a Windows desktop PC and a standard computer keyboard. Visual stimuli and instructions were displayed on the screen while audio stimuli were played through a pair of headphones, which participants wore throughout the experiment. During recall trials, their responses were recorded by a microphone positioned on the desk and connected to the PC. Both stimulus presentation and recording were carried out using PsychoPy software (Peirce et al., 2019). The procedure consisted of a training phase, a recall testing phase and a comprehension testing phase (reaction time task), delivered in this order. Each task was preceded by a set of written instructions, which participants read at their own pace. At the start, participants were informed that they would hear a series of sentences, each accompanied by a drawing of the scene it described. Every sentence would be almost entirely in English apart from one word, which would be a foreign word. Participants were encouraged to discover the Overt rule: they were told that different words were associated with different character positions on screen, and that they would be tested on the association later (full instructions are included in Appendix B). They were not encouraged to discover the meaning of the novel words. After instructions, participants would do a practice block for each task, and then begin the task. At the end, participants were asked to fill in a debriefing questionnaire, before being informed of the nature of the experiment and the Hidden rule.

3.2.3.1 Training phase

The training phase consisted of 96 sentences, divided into 8 blocks of 16. Each block was further divided into four groups of 4 sentences each: in each group, the four sentences were first presented in sequence, followed by four recall trials (in random order), one for each sentence. There was one exposure trial and one recall trial per sentence, for a total of 32 trials per block; there was a compulsory 10-second break at the end of every block of 32 trials. In an exposure trial, the audio recording of the sentence was played while the corresponding image (character + place picture combination) was displayed on screen. As soon as the sentence was over, the exposure trial for the next sentence began. Presentation was rapid, with minimal interval between stimuli. This was done on the basis of the results of a pilot study carried out beforehand. During piloting, we found that a greater interval between exposure trials had two disadvantages: it allowed participants to rehearse the sentences they had just heard, and this in turn encouraged many of them to employ mnemonic strategies (e.g. associations of ideas) to help them remember the sentences. After four exposure trials, the screen background colour changed from white to grey, and participants did four recall trials on the same sentences they had just heard (Fig. 3.2). In a recall trial, the image was displayed on screen for 6s, but no audio recording was played. Instead, participants were instructed to repeat the sentence corresponding to that image. Recall trials were shuffled: participants had to repeat the sentences they had heard in the last four exposure trials, but not in the same order as they had heard them. This, too, was based on the pilot study: we found that if recall trials followed the same order as the exposure ones, participants could simply memorise the sequence of prepositions (the only aspect of a sentence which could not be immediately retrieved from a picture) over the course of the four trials, without forming an association between each pictures and preposition. Shuffling the recall trials, on the other hand, ensured that participants had to form an association between the picture and the preposition in order to remember the sentences. The 96 sentences of the first task were made up of a mix of 64 matched sentences and 32 unmatched

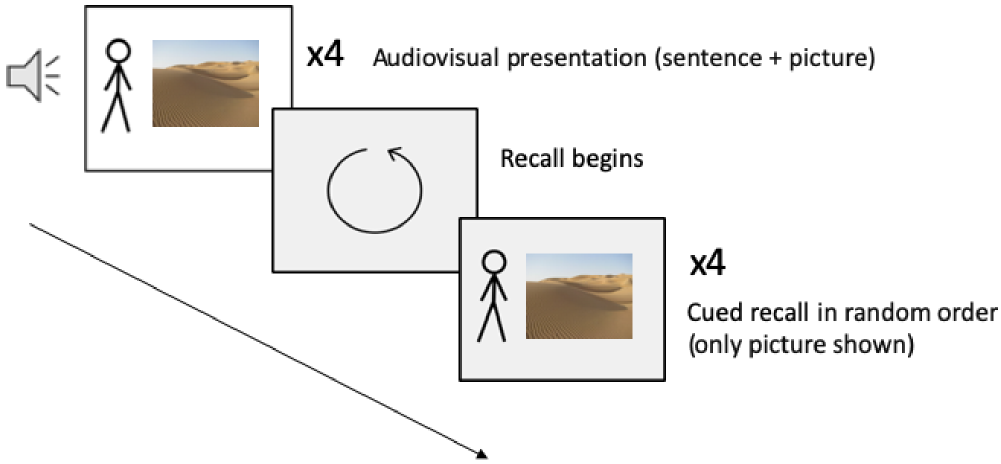


Figure 3.2: Training phase procedure for Experiment 1

sentences. The 64 matched sentences were based on a subset of 32 place nouns (half of the entire noun set). These were equally distributed between place types, and each appeared in two sentences – once in the System and once in the Random condition (32 nouns x 2 conditions = 64 matched sentences). The 32 unmatched sentences were based on the other half of the place noun set, which only appeared in one condition, counterbalanced between participants (32 nouns x 1 condition = 32 unmatched sentences).

3.2.3.2 Testing phase

Recall task The recall task used the same items as the training tasks and was identical to it in all respects, apart from the internal composition of each block. Blocks were still made up of 16 trials, but recall was every 8 trials, instead of every 4 (Fig. 3.3); the order in which items were presented for recall was still randomized. As in the training phase, there was a compulsory 10s break after every block of 16 trials. The choice of the number of exposure/recall trials for each task (4 for training, 8 for recall) was also based on the results of the pilot study. While in the training phase we wanted to minimize the number of recall

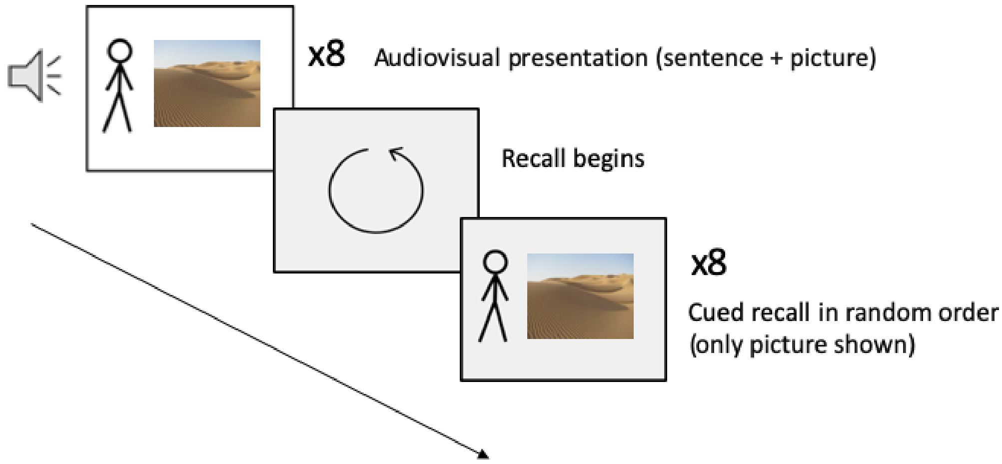


Figure 3.3: Recall task procedure

errors (in order to prevent incorrect associations from being formed), in the recall phase we wanted to maximise it, in order for the task to be as sensitive as possible and to avoid ceiling effects. On the other hand, the memory load could not be so great as to discourage participants from carrying out the task altogether. We found that a recall span of 8 sentences seemed to offer the best balance, taking account of individual differences.

Listening task Each trial began with a fixation cross which remained on screen for 0.5s. After that, the audio of the sentence started playing, while the screen remained blank. At 0.8s from the beginning of the sentence, simultaneously with the onset of the preposition, two pictures representing different locations (without any characters) appeared on screen, next to each other (Fig. 3.4). One of the pictures depicted the place mentioned in the sentence (*Target picture*), while the other (*Distractor picture*) was always drawn from the opposite category (either open or enclosed spaces). All pictures had already been used in both previous tasks and were therefore familiar to the participants. The target and distractor pictures were randomly assigned to either the left or the right side of the screen; participants were instructed to press a key on the keyboard ("D" for left, "K" for

Figure 3.4: Sample trial from RT task. The target and distractor picture are always drawn from different place categories.

right) as quickly as possible to indicate which of the two pictures depicted the place mentioned in the sentence. Reaction times were measured from the onset of the preposition. At the end of each block, participants received feedback on how fast their responses had been on average, encouraging them to try and lower their response times further. Similarly to previous tasks there was a compulsory break between blocks, in this case of 15s. The task consisted of a total of 128 trials, divided into 8 blocks of 16. The first 6 blocks were made up of the same 96 sentences already used in the training and recall phases, followed by one Generalisation and one Violation block. The stimuli in the generalisation and violation blocks were derived from the 32 unmatched sentences included in the training phase: if a place noun appeared in the training phase in the System condition, it would then appear in the generalisation block as Random, and vice versa (this was counterbalanced between participants). Both Generalisation and Violation items, then, were composed of previously seen pictures in novel combinations; additionally System items in the Violation block had reversed preposition assignment, so as to make them ungrammatical.

3.2.3.3 Debriefing questionnaire

Awareness of both the Overt and the Hidden rule was assessed by a questionnaire administered at the end of the experiment (Appendix A). Throughout the experiment and in the questionnaire, the prepositions were never described as such, but either mentioned specifically word by word or referred to simply as the "foreign" words. This was done so as not to condition participants' responses and gain an unbiased picture of their intuitions concerning the novel words. To assess awareness of the Overt rule (which they had been encouraged to discover), participants were asked whether they thought that the choice of preposition used depended on the character's position on the screen. To assess awareness of the Hidden rule (the Czech-like rule), on the other hand, participants were not explicitly asked if they thought that the nature of the place depicted in the picture determined preposition use. Instead, they were simply asked if they had spotted any other rules besides the Overt one. This was done in order to avoid confirmation bias on the subjects' part, and also based on findings from the pilot study. We found that providing a description of the rule, even in the form of a question, could in fact cause participants to become aware of it – which made it an unreliable tool for assessing prior awareness during the experiment.

3.2.4 Results

3.2.4.1 Recall task scoring

Each trial in the recall task was marked as either correct or incorrect depending on whether participants successfully reproduced the correct preposition during sentence recall; failure to produce the correct character name or place name was not marked as error. Occasionally, subjects would produce non-target-like versions of the phonological forms of the prepositions (e.g. "gwa" for *wa*, "bro" for *ro*, "glee" for *gi*). As a general rule, if a non-target-like form was unambiguously derived from a target form (usually by consonant/glide epenthesis on the syllable onset), and was used consistently in place of the target form, it was scored as the target form. On the other hand, non-target-like forms that consisted of a mix of different target forms (e.g. "nee" or "gwe"), and/or were used inconsistently alongside the target forms, were rejected. When scoring the task, we extracted two measures of accuracy: Overt rule and Hidden rule accuracy. Overt rule accuracy was the proportion of all recall trials in which participants correctly used the Overt rule, that is, when out of four possible prepositions, they used one of the two which were appropriate for that condition (System or Hidden, as cued by the character's position on screen relative to the place picture). To calculate Hidden rule accuracy, we only retained trials in which participants had used the Overt rule correctly, and calculated the proportion of those trials for which participants correctly recalled the exact preposition used in the item during training. Since the choice was between two possible preposition, chance level for Hidden rule accuracy was 50%.

3.2.4.2 Rule awareness

Participants' awareness of the Overt and Hidden rules was assessed using the debriefing questionnaire (Appendix A). Awareness of the Overt rule was assessed by Question 1 ("Did you think the use of words *ro*, *wa*, *ne* and *gi* was governed by any rules?") and Question 2 ("Did you think it depended on whether the character was inside or outside/near the place pictured?"). If participants answered "Yes" to Question 2, they were classed as having discovered the Overt rule and were included in the analysis.

Awareness of the Hidden rule was assessed by Question 1, Question 3 ("Did you think there were any other rules? [...] If you answered 'Yes', please describe the rules") and Question 6 ("Could you give a rough translation of the words *ro*, *gi*, *wa* and *ne*?"). Both questions served the same purpose, but Participants were classed as Aware if their answer to either Question 3 or 6 made reference to the rule. In Question 3, they would need to answer "Yes" and describe the rule by mentioning the relevant distinction, e.g. "open" vs. "enclosed" spaces, "indoor" vs. "outdoor" spaces, or "inside" vs. "outside", or in any other recognizable way, even indirectly. For instance, the following answer would class a participant as Aware: "I thought that wa was used for places which you could be inside, i.e. rooms, caves, whereas gi was used for places which you could be at but not inside, e.g. dock, savannah". Participants were counted as Aware regardless of whether they gave assignments for specific prepositions in Question 3 (e.g. "I thought *ro* was used for outdoor spaces" as opposed to "It depended on whether the space was indoor or outdoor"). In marking Question 6, we followed similar criteria: participants would need to provide translations which referenced the Hidden rule distinction, in any of the ways just described. Many participants incorrectly extended the Hidden rule to the Random pair of prepositions (e.g. "gi = inside building/room, ne = at or on terrace / external place, ro = far away from enclosed space, wa = far away from open space"); they were classed as Aware regardless of whether they did so, as long as they detected the presence of the Hidden rule.

3.2.4.3 Analysis

Participants who failed to discover the Overt rule (n = 6) were not included in the analysis; a further 2 participants were excluded due to low accuracy (<80%) in the RT task. In total, 21 subjects were included in the analysis (13 females, mean age 23 years, SD = 4.33); of these, 12 remained unaware of the Hidden rule, as assessed by debriefing questionnaire. Based on Hidden rule awareness, participants were divided into two groups: Unaware (n = 12) and Aware (n = 9).

Using mixed-effect modelling allowed us to look at trial-level data and account for variation between subjects and items within the same model, eliminating the need for by-subject and by-item analysis. Additionally, mixed-effect modelling is robust to missing data (Baayen et al., 2008). This feature was particularly useful when analysing Hidden Rule accuracy data, which was derived from a subset of the data based on Overt Rule accuracy (meaning that it would probably not be balanced, since the the selection criterion was blind to condition treatment).

We analysed the data using mixed-effect modelling, implemented in R using the *lme4* package (Bates et al., 2015b). We used a generalized linear mixed-effect model (GLMER) for binomial data to analyze accuracy scores, and a linear mixed-effect model (LMER) for RT data. In both cases, following Barr et al. (2013), we adopted a design-driven, rather data-driven, approach to random structure selection. As recommended by Bates et al. (2015a) and Matuschek et al. (2017), we followed a parsimonious approach rather than starting from a maximal random structure, including random effects as justified by the experimental design. We constructed an initial model including the maximal fixed structure and random intercepts for subjects and items. We then gradually added random slopes as justified by the design, only keeping them if they significantly improved model fit. We measured this by using maximum likelihood ratio tests (LRT), as recommended by Bolker et al. (2009), comparing models with and without the random effect of interest. Continuous fixed effects were scaled and centred to improve model fitting and interpretability in interactions.

To assess the statistical significance of effects in the logistic mixed-effect model (GLMER) on accuracy data we used Wald χ^2 tests following Bolker et al. (2009), who warns against using LRT for comparing fixed effects in GLMERs as they are unreliable for small to moderate sample sizes, and recommends Wald χ^2 tests instead (or Wald Z tests if the model is overdispersed). We implemented the Wald χ^2 tests using the *Anova()* function of package *car* (Fox and Weisberg, 2011). To assess statistical significance for the linear mixed-effect models (LMER), we computed F statistics using the *anova()* function of package *lmerTest* (Kuznetsova et al., 2017), which calculates degrees of freedom using the Satterthwaite approximation. We chose this method over gradual model simplification by LRT, as it has been shown to be better than LRT at containing Type I error for LMERs (Luke, 2017).

3.2.4.4 Recall task

Overt rule accuracy was relatively high for both groups, but more so for the Aware group (Aware: System 94%, Random 92%; Unaware: System 86%, Random 79%)(Fig. 3.5). We measured Hidden rule accuracy by only retaining trials in which subjects had got the Overt rule correctly – since the Hidden rule was a sub-rule of the Overt rule, it could only be applied if the latter had been correctly applied. Hidden rule accuracy (Fig. 3.6) was comparatively lower (Aware: System 91%, Random: 72%; Unaware: System 72%, Random 74%). To analyse the data, we built a GLMER with a random structure which included random intercepts for subjects only (since adding random intercepts for items caused singular fit due to insufficient variance in the Item effect) and correlated random slopes for

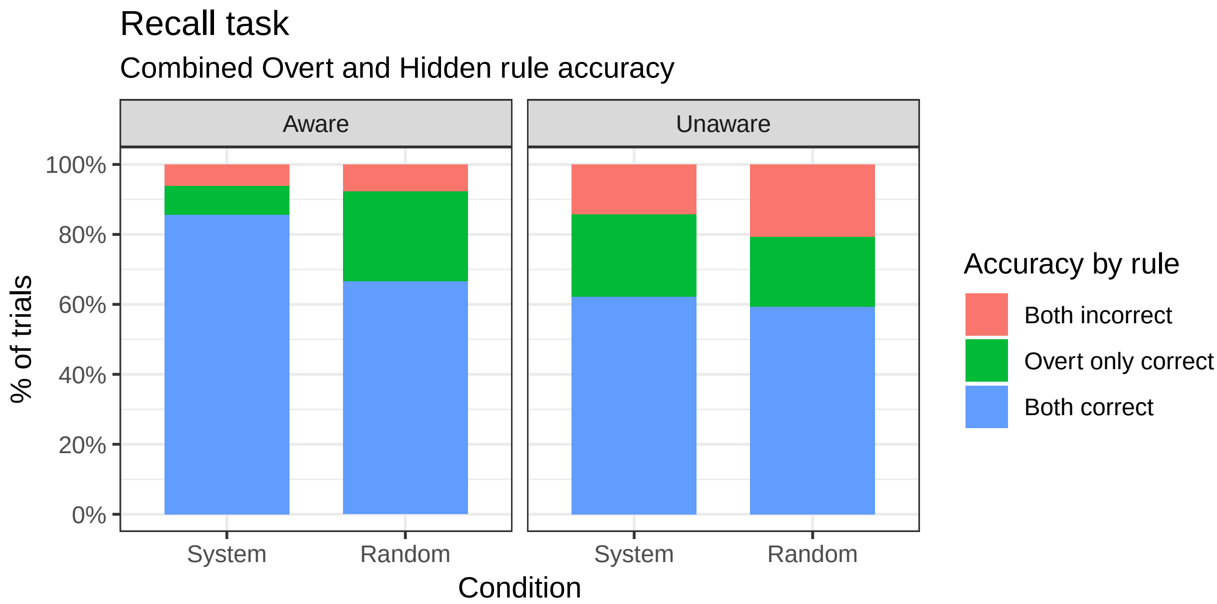


Figure 3.5: Combined Overt and Hidden rule accuracy by Group and Condition.

subjects by Condition. We then ran the model using this random structure and a fixed structure including Group, Condition and Block as main effects together will all their possible interactions (Table 3.1). We found significant main effects for Condition ($\chi^2(1) = 6.768$, $p = .009$) and Block ($\chi^2(1) = 4.256$, $p = .039$) as well as a strong interaction between Group and Condition ($\chi^2(1) = 11.408$, $p = .001$), together with a marginally significant 3-way interaction between Group, Condition and Block ($\chi^2(1) = 3.889$, $p = .049$).

We carried out post-hoc comparisons of means using the Bonferroni correction, focusing on the Group x Condition interaction first. Post-hoc tests showed that there was an effect of Condition for the Aware group only ($\chi^2(1) = 21.296$, $p < .001$), while there was no difference for the Unaware group ($\chi^2(1) = 0.123$, $p = 1$); the main effect of Condition seems to be driven by the size of the effect for the Aware group alone (Fig. 3.6). The effect was modulated by Block: a post-hoc comparison of the Group x Condition x Block interaction showed that the slope coefficient for Block was significantly greater for System compared to Random items in the Aware group (System: $\beta = 0.765$, Random: $\beta = 0.092$, $p = .014$), while no difference was observed for the Unaware group. This is compatible with visual inspection of the data (Fig. 3.7), which shows a gradual increase fof

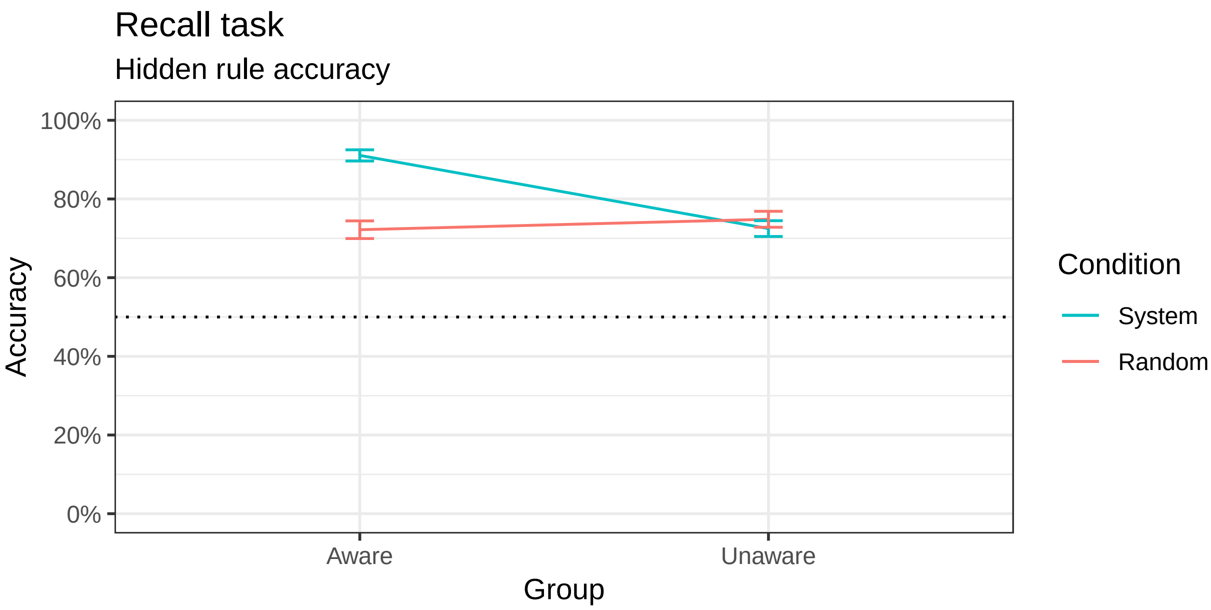


Figure 3.6: Average Hidden rule accuracy by Group and Condition. Error bars represent SE of the mean, dotted line marks 50% chance level.

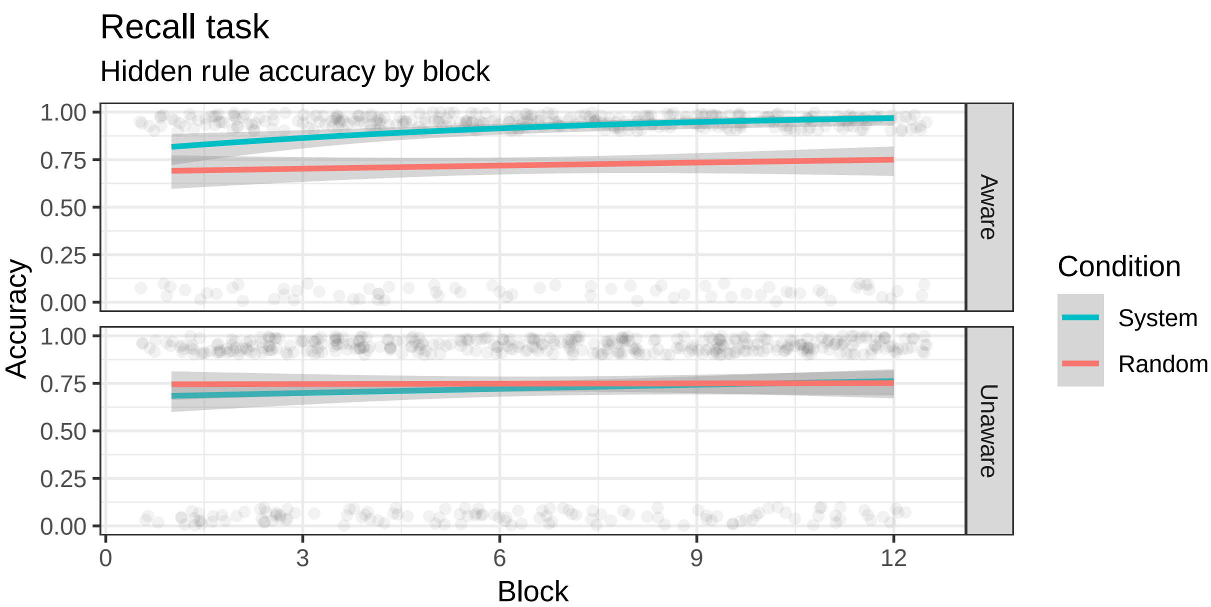


Figure 3.7: Hidden rule accuracy by block with fitted linear trends.

Hidden rule accuracy, Recall task

	Coefficient	SE	χ^2	df	p	
Group	-2.305	0.731	0.544	1	0.461	
Condition	-2.588	0.563	6.768	1	0.009	**
Block	0.765	0.221	4.256	1	0.039	*
Group x Condition	2.452	0.677	11.408	1	0.001	***
Group x Block	-0.652	0.248	3.257	1	0.071	
Condition x Block	-0.673	0.249	3.773	1	0.052	
Group x Condition x Block	0.579	0.294	3.889	1	0.049	*

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 3.1: Summary of GLMER model for Hidden rule accuracy.

the difference between Random and System items for the Aware group over the course of the task.

3.2.4.5 Listening task

Raw reaction times for first-attempt correct responses were recorded, and outliers at more than +/-2.5 standard deviations from the mean (amounting to 4% of total data) were removed. Average reaction times by Group and Condition across the task are shown in Fig. 3.8. To analyse reaction time data, we ran separate LMERs for the thee phases of the RT task (Training items, Generalisation items and Violation items).

Training items We selected a random structure with random intercepts for Subjects and Items and a correlated random slopes for Trial and Subjects and for Trial and Items; more complex models failed to converge. In the resulting model, we found a significant main effect of Trial (F(1,39.2) = 15.029, $p < .001$) and significant interactions for Group x Condition (F(1,1657.7) = 5.332, p = .021) and Condition x Trial (F(1, 554.2) = 16.491, $p < .001$), as well as a Group x Condition x Trial interaction (F(1,1679.2) = 7.101, p = .008) (Table 3.2). Plotting the residuals (the difference between actual data points and the values predicted by the model) shows that they are homoskedastic and normally distributed (Fig. 3.10) indicating that the model's linear fit is appropriate for the

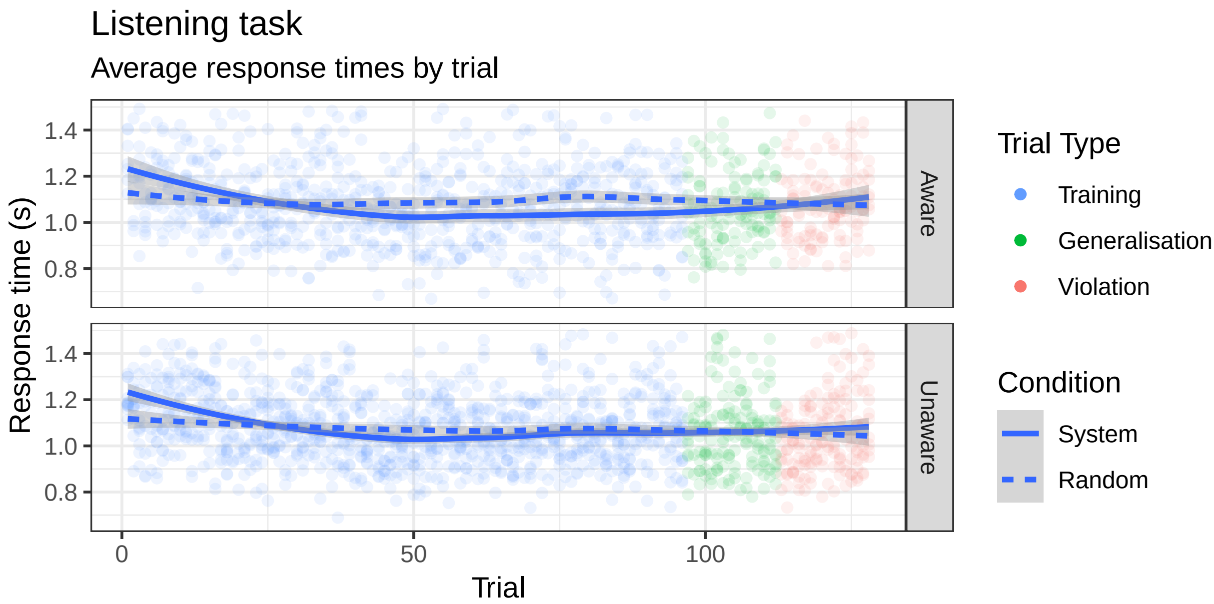


Figure 3.8: Response times with averages by Group and Condition.

data.

We carried out post-hoc comparisons using the Bonferroni correction to explore the interactions found in the model. The effect of Trial was significant for System items only, and it remained significant at the group level for both Aware ($\chi^2(1) = 24.530$, $p < .00001$) and Unaware subjects ($\chi^2(1) = 16.633$, $p < .0001$), reflecting a gradual speed-up of responses to System items in both groups. However, the difference between Conditions was only significant for the Aware group ($\chi^2(1) = 5.470$, $p = .039$), and it was significantly affected by Trial ($\chi^2(1) = 21.117$, $p < .0001$). This reflects the gradual increase in RT difference between Random and System items, which was greater for Aware subjects than for the Unaware (Fig. 3.9).

Generalisation items For Generalisation items, we used a simpler random structure which only included random intercepts for Items and Subjects; more complex models failed to converge, possibly due to the fact that we had fewer data points for generalisation trials compared to Training trials. In the resulting model (Table 3.3), the only statistically significant effect was a Condition x Trial interaction ($F(1{,}104.2) = 10.012$, $p = .002$). The plotted residuals are

Training items, Listening task

	Coefficient	SE	df	F	p	
Group	0.012	0.023	20.892	0.009	0.924	
Condition	0.025	0.010	317.356	1.864	0.173	
Trial	-0.054	0.011	39.137	15.029	0.000	***
Group x Condition	-0.029	0.013	1657.669	5.332	0.021	*
Group x Trial	0.013	0.011	20.328	0.168	0.686	
Condition x Trial	0.050	0.011	554.166	16.491	0.000	***
Group x Condition x Trial	-0.034	0.013	1679.176	7.101	0.008	**

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 3.2: Summary of LMER model for Training items.

Generalisation items, Listening task

	Coefficient	SE	df	F	p	
Group	-0.004	0.043	20.657	0.001	0.971	
Condition	0.005	0.023	247.655	0.271	0.603	
Trial	0.054	0.017	73.327	2.909	0.092	
Group x Condition	0.006	0.032	254.528	0.033	0.856	
Group x Trial	-0.029	0.022	262.897	0.712	0.400	
Condition x Trial	-0.067	0.024	104.159	10.012	0.002	**
Group x Condition x Trial	0.032	0.032	264.320	0.993	0.320	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 3.3: Summary of LMER model for Generalisation items.

homoskedastic and normally distributed (Fig. 3.11) indicating that the model's linear fit is appropriate for the data.

Post-hoc comparisons using the Bonferroni correction to explore the Condition x Trial interaction showed that the effect of Trial was only significant for System items, and then only for the Aware group ($\chi^2(1) = 10.509$, $p = .005$; Unaware: $\chi^2(1) = 2.878$, $p = .36$). Accross both groups, the slope coefficient estimate for Trial was positive for System items ($\beta = 0.04$) but negative for Random items ($\beta = -0.012$) which reflects a slowdown in RT for System items during the generalisation phase which affected both groups, even though it is more pronounced for the Aware group (Fig. 3.9).

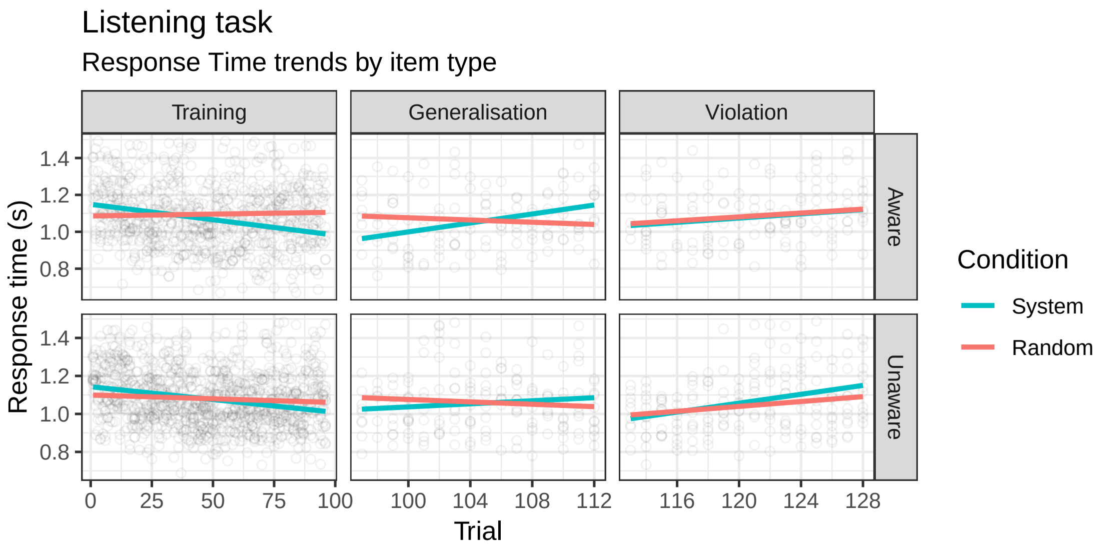


Figure 3.9: Response times by Item Type, Group and Condition with fitted linear trends.

Violation items We used only included random intercepts for subjects in the random structure, as more complex models failed to converge. In the resulting model (Table 3.4), we found a main effect of Trial ($F(1{,}270.1) = 16.191$, $p < .001$) with a positive coefficient estimate ($\beta = 0.019$) which reflects a general increase in RTs during this phase of the test (Fig. 3.9). The residuals are homoskedastic and normally distributed (Fig. 3.12) indicating that the model's linear fit is appropriate for the data.

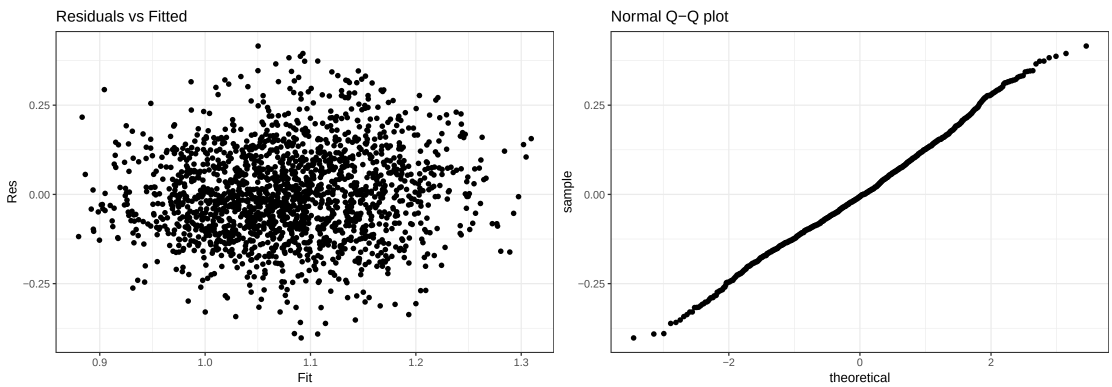


Figure 3.10: Plotted residuals for Training items LMER.

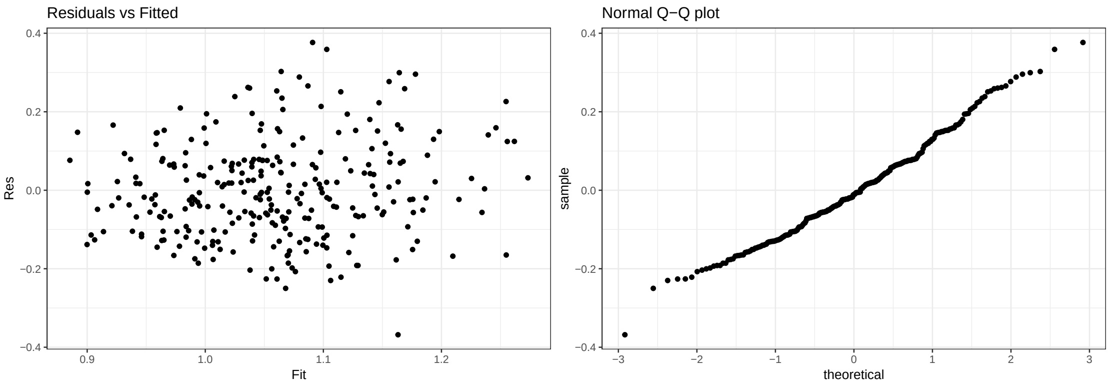


Figure 3.11: Plotted residuals for generalisation items LMER.

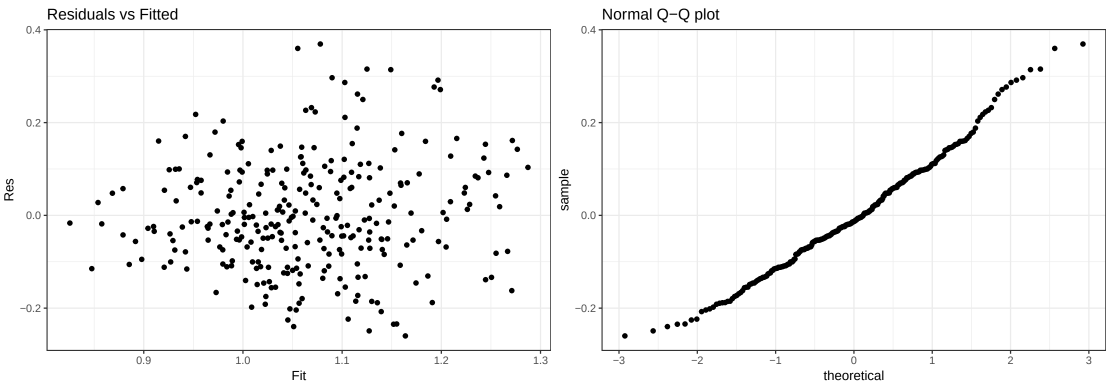


Figure 3.12: Plotted residuals for Violation items LMER.

Violation items, Listening task

	Coefficient	SE	df	F	p	
Group	-0.025	0.044	21.104	0.675	0.420	
Condition	0.000	0.023	269.702	0.343	0.559	
Trial	0.019	0.017	270.125	16.191	0.000	***
Group x Condition	-0.018	0.029	269.702	0.355	0.552	
Group x Trial	0.031	0.021	270.125	1.600	0.207	
Condition x Trial	0.003	0.023	270.223	0.335	0.563	
Group x Condition x Trial	-0.024	0.030	270.223	0.632	0.427	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 3.4: Summary of LMER model for Violation items.

3.2.5 Discussion

3.2.5.1 Development of productive skills

If participants had acquired implicit productive knowledge of the Hidden rule, we expected to see higher recall accuracy rates for System items relative to Random items in the Recall task, regardless of rule awareness. Our predictions were not confirmed: we did observe an effect of Condition on Hidden rule recall, but only for the Aware group; there was no difference in accuracy between conditions for the Unaware. Given that we did not find a significant effect in production among the Unaware, we assume that Aware participants' performance in the test was driven by explicit knowledge.

With regards to the Unaware group, we see two possible explanations for the lack of a visible learning effect among these subjects: either the Unaware group did not acquire any knowledge of the Hidden rule (whether explicit or implicit), or the task used was not sensitive enough to detect implicit knowledge. In the latter case, it may possible that Unaware participants had developed implicit knowledge of the rule, but that it was not detected by the task. The assumption of the Elicited Oral imitation task, on which our testing paradigm was based, is that where a large amount of linguistic material needs to be stored and recalled from memory, knowledge of the language rules will supplement short-term phonological memory capacity, allowing participants to remember longer stretches of material and even unconsciously correct mistakes in the original input (Erlam, 2006; 2009). Therefore, if the testing task was not demanding enough in this respect, it is possible that even if Unaware participants did have some representation of the rule, they had no need to rely on it, in a situation were item memory was still strong. Hidden rule accuracy for Random items (which could only be correctly reproduced by relying on item memory) was about 75% for both groups, against a chance level of 50%, suggesting that they could still rely on item memory to a large extent. To address this point, the next experiment will include a use recall task and introduce novel (generalisation) items, in order to create situations where subjects have minimal or non-existent item memory. This will allow us to determine whether the failure of the current task to detect any knowledge among the Unaware group should be attributed to a lack of rule knowledge or to low task sensitivity.

For the Aware group, we also observed an interaction of Block with Condition: the difference in accuracy between Random and System items increased as the task went on. It could be, then, that the Recall task was not simply testing knowledge developed during the Training phase, but was providing further opportunities for learning. However, while it is true that the effect increases over time for Aware participants, a separation between conditions is already visible at the start of the task, and increases very rapidly (Fig. 3.7). There is more than one possible interpretation for this pattern: it could be reflecting emerging awareness of the rule, or further development of implicit rule knowledge from training. To address this point, in the next experiments we will retain and analyse data from the Training phase, too, in order to gain a fuller picture of the learning process undergone by participants.

3.2.5.2 Development of comprehension skills

In the first phase of the Listening task (Training items), we expected participants to respond faster to System items than to Random ones, if they had acquired any knowledge of the Hidden rule, because it would have allowed them to anticipate the target picture before hearing the final noun. What we observed was a significant speed-up for System items in both groups over the course of the task, but the difference between System and Random was only significant for the Aware; the Unaware group showed greater decrease in response time for Random items, too (Fig. 3.8). The interaction of Condition with trial in the Aware group suggests any receptive sensitivity to the rule was developed as a consequence of the task itself, rather than being carried over from the previous tasks. Indeed, visual inspection of the data shows that, at the beginning of the listening task (Fig. 3.8), both groups were slower when responding to System items than to Random ones; only later does the pattern get reversed. If the listening task was tapping into pre-existing knowledge, we would have expected the difference in RT to appear from the beginning. Rather than being a test of implicit knowledge developed during the previous tasks, then, the listening task may have served as a learning opportunity in its own right, developing a different type of domain-specific knowledge. This is compatible with the findings of previous IL studies which used time listening tasks as both training and testing (Leung and Williams, 2006, 2011).

The fact that comprehension skills appear to emerge (in the Aware group) over the course of the listening task supports a skill-specific account (DeKeyser and Sokalski, 1996). It would appear that comprehension practice was required in order to develop receptive skills, and that participants could not initially make use of the knowledge they relied on during the production task. However, while the same trend was observed for the Unaware, the difference between conditions was only significant for Aware participants, suggesting a possible connection with the outcomes of the production task. It could be a causal connection - awareness or greater accuracy of the Hidden rule during the production task may have favoured performance in the Listening task. This, too, would be compatible with DeKeyser's account of transfer between modalities (DeKeyser and Sokalski, 1996): it is possible that the declarative knowledge acquired by Aware participants in the first phase of the experiment served as basis for proceduralization during the listening task. It would appear that Aware participants could rely on some explicit knowledge derived from the previous task, but that skill-specific practice was needed for it to develop into receptive sensitivity. However, there could also be extraneous cognitive factors affecting Aware and Unaware participants differently, underlying both performance in production, the development of conscious knowledge and the development of comprehension skills in the listening task.

With regards to Generalisation items, we expected to see the same pattern we had predicted for Training items: a sustained difference between conditions, with faster response times to System items relative to Random ones. Instead, we observed a slowdown for System items relative to Random ones; this was significant for Aware participants only (Fig. 3.8 & 3.9). Visual inspection of the data suggests this is due to a loss of the advantage gained by System items over the course of the previous phase of the task, as RTs to System items go back down to the level of Random items, for both groups (Fig. 3.8); the slowdown is not significant for the Unaware participants, however, because the difference between conditions was not significant for them in the first instance.

A possible explanation explanation for this finding may be picture familiarity. System items in this phase of the task were still grammatical; the only thing that distinguished them from Training items was that participants were hearing them for the first time, as they were not included in the Training phase. They were novel picture + preposition combinations, using pictures that had already appeared once in training (assigned to the opposite condition). By contrast, the pictures used in matched training items appeared twice during training, so would have been more familiar. The slowdown in response to generalisation items may therefore be due to a lack of familiarity: either to the specific picture + preposition combinations (which were novel), or to the pictures themselves, which had only appeared once in training, instead of twice. However, in that case we would expect to see a similar slowdown to for Random items, too, since they were equally unfamiliar. Instead, response times for the two conditions seemed to converge in this phase of the task, which would suggest a lack of ability to generalise the rule even for the Aware participants. This is puzzling if we assume that the speedup for System items in the previous phase (Training items) was a result of the task itself: if participants were gaining a new, modality-specific skill, then there ought to be no difference between Training and Generalisation items, since they were all equally 'new' in the context of the Listening task (being presented in this modality for the first time). On the other hand, the fact that the System items are processed differently from Random items suggest that there is something about the rule which participants are sensitive to. A familiarity-based account is compatible with the hypothesis that the speedup for Training items in the System condition was due to ease of recognition (cognitive fluency), rather than predictive processing. The fact that the effect was found for Training items but not Generalisation items also suggests that it may have been a low-level effect rather than abstract generalisation: perhaps the presence of the rule simply eased recognition of System items that had already been encountered during training, rather than allowing Aware subjects to predict upcoming content. This is compatible with research showing that the ability to make predictions based on morphosyntactic information is generally impaired in L2 speakers. While there is evidence that L1 speakers anticipate the content of upcoming stimuli based on syntactic (Tanenhaus et al., 1995) and morphophonological cues (DeLong et al., 2005), this does not seem to be the case with L2 learners (Martin et al., 2013). A visual world paradigm study of adult learners of Spanish as L2 found that they would not use forward-looking morphological cues (determiner gender) to rapidly direct their gaze to a fitting noun, unlike native Spanish speakers (Lew-Williams and Fernald, 2010). In a study on the implicit learning of a direct object marking construction sensitive to animacy, Andringa and Curcic (2015) found that even subjects who were given explicit instruction at the start of training (and who performed well in a grammaticality judgement task) were no faster at picking the correct picture when animacy could be predicted based on the DOM construction, and did not show any preferential gazing either.

Finally, our prediction for Violation items was to observe the opposite pattern to Training and Generalisation items: we expected a slowdown in responses to (now ungrammatical) System items relative to Random ones. Instead, we observed an overall slowdown, with no differences between groups or condition. Given that the violation trials were at the very end of the task, this could simply be an effect of fatigue. However, it could also be that experiencing unexpected outcomes for System items generated a surprisal effect which then affected Random items, too, by causing participants to become more hesitant in their responses. Based on the available data, however, it is not possible to conclusively attribute the slowdown to either of these suggested reasons.

3.3 Conclusion

In this first experiment, we did not find evidence of implicit rule learning which could translate into production skills. Only those who were aware of the Hidden rule exhibited higher recall accuracy for System items, suggesting that they were relying on their explicit knowledge of the rule when performing the task. The lack of evidence of implicit rule knowledge may be due the task not being sensitive enough, or to the fact that groups did not in fact develop implicit knowledge. Accordingly, with regards to comprehension skills, we did not find evidence of implicit knowledge being transferred from the production to the comprehension task for either group: some sensitivity to the rule appeared to emerge during the course of the task, but only for aware participants, which suggests it may have benefitted from existing declarative rule knowledge. In the next experiment, we will address these issues by modifying both production and comprehension task, in order to make them more sensitive to implicit knowledge.

Chapter 4

Experiment 2: Developing productive generalisation

4.1 Introduction

The findings of Experiment 1 suggest that only Aware participants had developed sensitivity to the Hidden rule, as evidenced by a recall advantage for System items, which was probably driven by their explicit knowledge of the rule. The lack of a significant effect for the Unaware group would seem to indicate that participants did not develop implicit knowledge of the rule by doing the production task. However, we have suggested that this could also be due to the fact that the task was not sensitive enough to detect implicit knowledge. To address this possibility, we ran two further experiments, in which we modified the testing phase in order to make it more sensitive to implicit knowledge. This was achieved by employing a harder recall task and introducing cued production of novel (generalisation) items, as explained in Section 4.2.1. We also made a number of changes to the training methodology: based on the results of Experiment 1, we modified the training paradigm to reduce the opportunity for explicit learning; the modified paradigm is described in Section 4.2.2. Finally, given the inconclusive results of the comprehension task from Experiment 1, we decided to use a different measure of receptive knowledge, employing a recognition memory task instead of the reaction time task used previously (Section 4.2.3). This chapter and the next report the findings of two further experiments we conducted, using two variations of the modified training paradigm from Experiment 1. Experiment 2, reported in this chapter (Section 4.3) introduces long-term recall of training items and generalisation items, as well as a new comprehension task. Experiment 3 (Chapter 5) employs a modified version of the training paradigm used in Experiment 2, designed to increase engagement with the stimuli and improve acquisition of both Overt and Hidden rules.

4.2 Methodology

4.2.1 Introducing productive generalisation

In Experiment 1, we found a significant effect of condition on recall accuracy for the Aware group, but not for the Unaware group. While this may indicate that participants had developed no implicit knowledge of the Hidden rule, it could also be the case that the task was not sensitive enough. It is possible that, if both item memory and implicit rule knowledge are available as sources of information, item memory may be preferentially used, masking any potential evidence of implicit knowledge. The testing phase in Experiment 1 required participants to recall 8 items at a time; accuracy for Random items was roughly 75% for both groups, far above the 50% chance level, which indicates that item memory was still strong. In order to explore this possibility, we introduced a long-term recall task, designed to minimise the effect of item memory. We also decided to include a generalisation component, which will require participants to generate new sentences according to the Hidden rule, while remaining under the impression that they are doing a recall task. By adding generalisation items, we can examine what happens when no item memory trace is present, which may compete with rule use. If unaware participants have acquired any knowledge of the rule, we would expect it to emerge in these circumstances.

4.2.2 Changes to the training procedure

In Experiment 1, as well as in further pilot studies, we observed that the proportion of participants becoming aware of the Hidden rule was relatively high (9 out of 21), compared to similar studies (e.g. Williams, 2005). Given that the focus of the study was on the development of implicit knowledge, which requires a sufficient sample of unaware participants to be studied, we decided to modify the training procedure used in Experiment 1 to discourage rule discovery, by reducing the number of items recalled in each trial from four to two. We hypothesised that having four items per trial facilitated hypothesis testing by allowing participants to directly compare all four prepositions and their usage; reducing the number of items to two per trial ensured that participants would never be exposed to the full system in any given trial. To maintain a high memory load and compensate for the reduction in items per trial, we used different means in each of the subsequent experiments: in Experiment 2, we replaced a subset of the place nouns with pseudowords; in Experiment 3, we inserted questions after each item, which probed Overt rule knowledge and drew subjects' attention to the physical properties of the place pictured. Detailed descriptions of the training procedures used in Experiments 2 and 3 can be found in Sections 4.3.3 and 5.2.3.

4.2.3 Measuring receptive knowledge

The results of the listening task used in Experiment 1 did not show conclusive evidence of knowledge transfer from production to comprehension. The Aware group, which was the only group to show a significant learning effect in the production task, showed a difference in response times between conditions for trained items; however, the difference emerged over time, suggesting that it may have developed as a result of the task itself. This would be compatible with a skill acquisition approach (DeKeyser, 1997; DeKeyser and Sokalski, 1996) in which production and comprehension skills need separate, modality-specific practice in order to develop. Furthermore, the effect did not persist in generalisation trials, which raises the possibility that the difference in response times observed during the first phase of the listening task may have been due to ease of recognition, rather than rule acquisition. Finally, as noted in Section 3.2.5, it is also possible that the task we used was not sensitive enough to pick up newly acquired implicit knowledge, since it relied on predictive mechanisms which are often lacking in early L2 learners. It is possible that, if participants had developed implicit abstract knowledge of the rule, the task would not have been able to detect it. Therefore, in the next experiments we decided to use a different measure of comprehension: we abandoned the listening task and adopted a recognition memory task instead, as previously done by similar studies on implicit and incidental learning of new linguistic forms (Paciorek and Williams, 2015; Tamminen et al., 2015).

4.3 Experiment 2

The aim of this experiment was to test whether participants would be able to generalise the Hidden rule to new items, in order to establish whether they had acquired productive knowledge of the rule. It also aimed to provide a more sensitive measure of implicit knowledge, by creating conditions for the knowledge to emerge without strong competition from item memory, by adding long-term recall of training items as well as generalisation items. We also used a different measure of comprehension, a recognition memory task, in order to address some of the issues that emerged in the previous experiment with regards to comprehension testing. As in Experiment 1, the aim of including the comprehension task was to test whether the knowledge acquired through the recall paradigm would be available to comprehension as well as production processing. In the production task, if participants had acquired implicit knowledge of the rule, we would expect there to be a recall advantage for System items in the long-term recall task. We would also expect participants to perform at above chance level when producing new items in the System condition during generalisation testing, which required productive application of the Hidden rule. In the recognition memory task, too, we would expect endorsement rates to vary as a function of condition, with a bias towards endorsing rule-based items over random ones, and grammatical ones over ungrammatical ones.

4.3.1 Participants

42 native English speakers (29 females, mean age 20.5 years) from the University of Cambridge and surrounding community took part in the experiment, receiving £6 as compensation. Of those who took part, two reported some knowledge of a Slavonic language (Russian). Other foreign languages spoken by participants were French (n = 13), German (n = 10), Spanish (n = 6), Italian (n = 4), Mandarin Chinese (n = 1), Irish (n = 1), Greek (n = 1), Hindi (n = 1), Urdu (n = 1) and Arabic (n = 1).

4.3.2 Materials

We used the same items as Experiment 1 , with the addition of a new set of items to be used in generalisation. Using the same procedure outlined in Experiment 1, we selected a further 16 place nouns from a list of the most frequent nouns used with *do* and *na* prepositions in Czech, which added to the 64 already used in Experiment 1, for a total of 80 unique nouns. Pictures depicting the nouns were obtained using the same procedure as Experiment 1.

4.3.3 Procedure

The general structure of the experiment was similar to that used in Experiment 1, with a training phase followed by production and comprehension testing. However, there were important differences in every task, most notably changes to the training procedure (Fig. 4.1), the introduction of long-term recall and generalisation items in the Production task (Fig. 4.2) and a new type of comprehension task (Recognition memory task).

4.3.3.1 Training phase

During the training phase, participants were exposed to 112 sentences: of these, 64 were the same matched items used in Experiment 1 (32 unique nouns appearing once in each condition), the remaining 48 were unmatched items, 48 unique nouns appearing only in one condition during training (counterbalanced across participants). The training phase followed the same basic procedure as Experiment 1, with two differences: the number of items per trial, and the introduction of non-words (Fig. 4.1).

The first difference was the number of items per trial: recall was every two sentences instead of every four. The previous experiment and piloting for the current experiment had yielded a high rate of insight; one possible explanation for that was that each trial could contain instances of all four prepositions, allowing for direct comparison which could facilitate hypothesis formation and testing. To account for this, the number of items per trial was reduced to two (ensuring that they did not contain the same preposition), so that participants would never be exposed to the full system in any given trial.

The second difference was the introduction of non-words replacing a subset of place nouns in the training phase. Reducing the number of items per trial meant that the memory load for each trial was now significantly smaller, and the task was not demanding enough. In order to both increase memory load and keep participants engaged in the task, in half of the matched items (32 sentences) the place noun was replaced by a four-phoneme pseudoword in the audio recording, which participants had to repeat together with the rest of the sentence (e.g. "Harry is *gi dreet*"). The full list of pseudowords used is provided in Appendix C, p. 201.

4.3.3.2 Testing phase

Unlike Experiment 1, the production task in this experiment did not include any further exposure to the stimulus sentences. It consisted entirely of recall screens: participants were simply shown the graphic representation for each sentence (the character + picture combination) and were asked to produce the corresponding sentence. In this task, items already seen in training were intermixed with generalisation items; this means that, even though all trials followed the same procedure, the task was in fact a combination of a long-term recall task and generalisation task (Fig. 4.2). At this stage, participants were not explicitly told to repeat the pseudoword matching the place picture, only to recall the sentence inclusive of preposition.

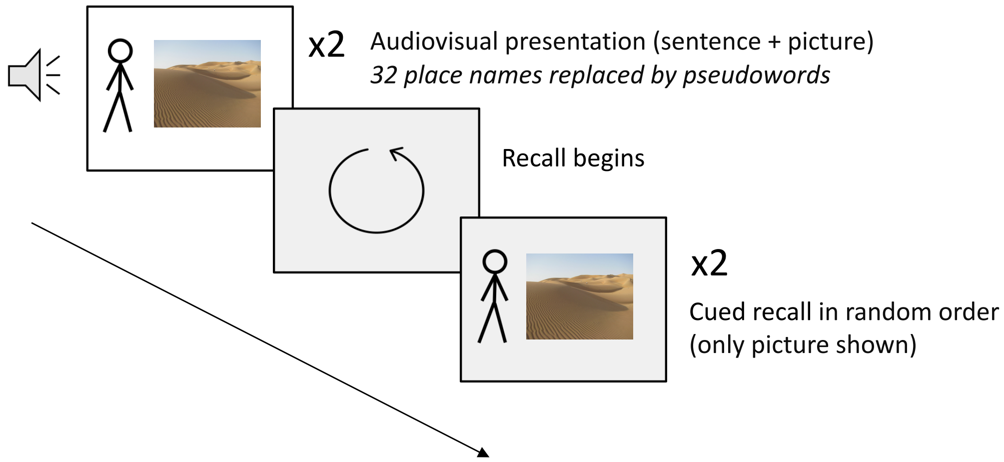


Figure 4.1: Training phase procedure for Experiment 2

The testing phase consisted of 88 items in total, which were composed of 64 recall items (the 64 matched items used in training) and 24 generalisation ones, which were half of the 48 nouns used in unmatched items during training, presented in the opposite condition (counterbalanced across participants). The items were presented in 11 blocks of 8 items each (7 blocks with recall items, 3 with generalisation ones), presented in random order, with generalisation and recall blocks intermixed.

Long-term recall In long-term recall trials, participants were shown the recall cues (character + picture combinations) for the 64 matched items already encountered during training (32 unique pictures, shown once in the System and once in the Random condition).

Generalisation In generalisation trials, participants were shown 24 novel character + picture combinations (generalisation items). Generalisation items were derived from a subset of training items: we selected half of the 48 unmatched items from the training phase (counterbalancing them across participants) and presented them in the opposite condition during the production task. The gen-

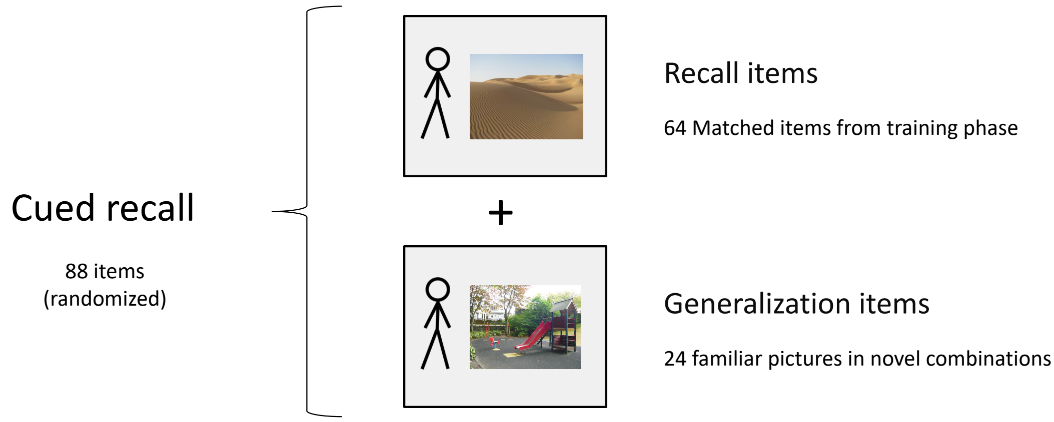


Figure 4.2: Production task procedure for Experiment 2

eralisation items were composed of pictures participants had already seen during training, but the character would be in a different position, requiring a different preposition. Participants were not told about the insertion of these novel combinations and believed that all picture combinations they saw had already been presented during training. None of the participants reported becoming aware of this manipulation at any stage during or after the experiment.

Recognition memory task Given the results of the RT task from Experiment 1, we decided to use a different task to assess for receptive rule knowledge of the end of the experiment. We opted for a recognition memory task, following Paciorek (2012). During each trial, participants heard stimulus sentences while the corresponding place picture was displayed on screen, but without any character superimposed. They were then prompted to respond by button press whether they had heard the sentence before over the course of the experiment, on a scale from 1 ("Definitely no") to 6 ("Definitely yes"). They were instructed to base their answer on the sentence, rather than on the picture alone. The task consisted of 48 sentences, all using pictures participants had already seen during training. 24 of these were the a subset of the 64 matched items used in training (Old items);

the other 24 were New items, i.e. half of the 48 nouns from unmatched training items, presented in the opposite condition (the other 24 had already been used as generalisation items for the generalisation task – this was counterbalanced across participants). For New items in the System condition, prepositions were assigned violating the Hidden rule, so as to make the items ungrammatical (Table 4.1);

	System	**Random**
Old	Grammatical	Neutral
New	Ungrammatical	Neutral

Table 4.1: Summary of item types used in Recognition memory task and their grammaticality.

4.3.3.3 Debriefing questionnaire

Participants were debriefed using the same questionnaire employed in Experiment 1 (Appendix A) to assess awareness of the Overt and Hidden rules.

4.3.4 Results

4.3.4.1 Task scoring

Scoring for the production task followed the same procedure outlined in Experiment 1 for recall task scoring (Section 3.2.4.1, p. 63).

4.3.4.2 Rule awareness

Awareness of the Overt and Hidden rule was assessed by debriefing questionnaire (Appendix A), as already done in Experiment 1. With regards to the Hidden rule, we initially divided participants into Aware and Unaware groups, following the criteria already used for Experiment 1 (Section 3.2.4.2, p. 64). However, we noticed that a number of participants did not explicitly mention the distinction between open and enclosed spaces (or any of the other formulations accepted in Experiment 1), but when giving translation for the pseudowords in Questions 3 and 6, they used English prepositions“in” and “on” (and near synonyms) as translations for the enclosed and open space System prepositions respectively (the equivalents of Czech *v* and *na*). We decided to code these answers separately and created a third *Transfer* category, in which we entered participants who mapped the English "in"/"on" distinction onto the Hidden rule, in addition to Aware and Unaware.

4.3.4.3 Analysis

We excluded participants who did not become aware of the Overt rule (N = 6). A further 3 participants were excluded for failing to perform all the tasks in the experiment. A total of 33 participants were included in the analysis (22 females, mean age 20 years, SD = 2.09). Based on their awareness of the Hidden rule, participants were initially divided into *Aware* (N = 10), *Transfer* (N = 6), and *Unaware* (N = 17). We decided to merge the Transfer group into the Aware group for analysis purposes, based on the following assessment.

To assess the extent to which transfer of English categories may aid Hidden rule recall in our task, we investigated the amount of overlap between the two systems in our items set. We first ran an online norming study to elicit judgments from native English speakers, to establish which of the two English prepositions (*in* or *on*) participants would be most likely to associate with the place nouns used in our study. 49 native English speakers, recruited from the University of Cambridge and surrounding community, took part in the norming study. A comparison of the two rules revealed a high amount of overlap in our item set: 80% of System items could be correctly produced by transferring the English rule *in* vs *on* distinction onto our Hidden rule (mapping *in* onto the enclosed space preposition in our study, and *on* onto the open space one). Therefore, given the magnitude of the advantage, we decided to enter participants from the Transfer group into the Aware group for the purposes of our analysis.

Data from the training phase and production task was analysed with separate logistic mixed-effect models (GLMERs) following the same procedure as Experiment 1 for model selection and for drawing statistical inferences. Data from the recognition memory task was analysed using mixed-effects proportional odds models (also known as cumulative link mixed models, CLMMs) for ordinal re-

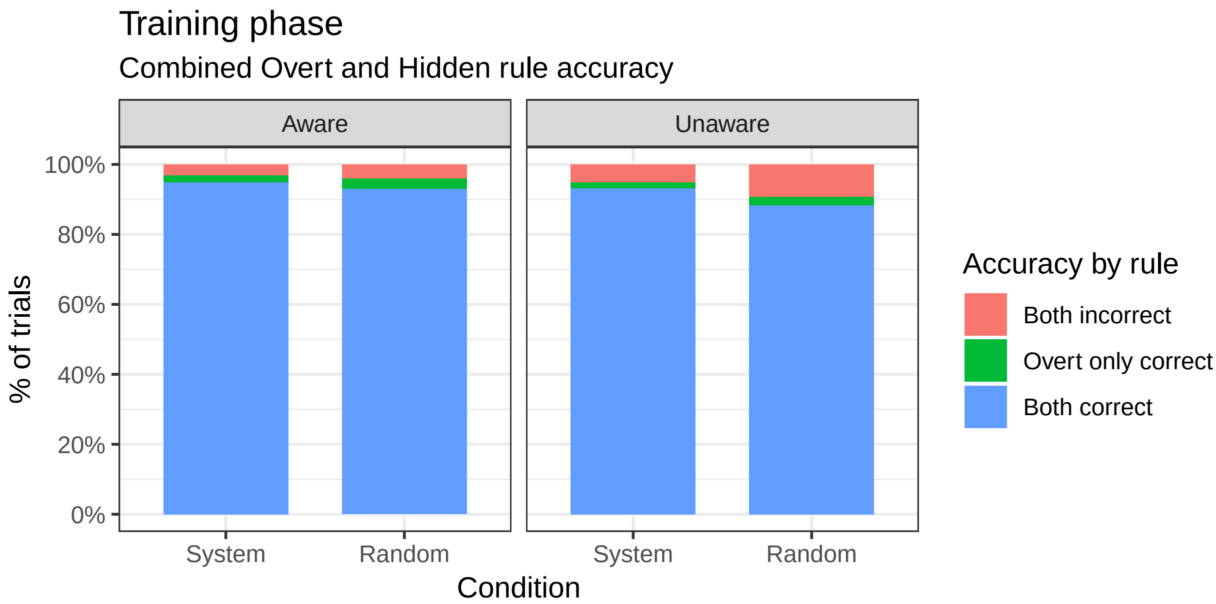


Figure 4.3: Combined Overt and Hidden rule accuracy by Group and Condition during training phase.

gression, as the data was in the form of Likert-scale ratings. We implemented the CLMMs in R using the package *ordinal* (Christensen, 2019). We selected the random structure as done with other models, gradually adding random effects to the maximal fixed structure and only retaining them if they significantly improved model fit. Since there is currently no way to simultaneously assess significance for the effects included in a CLMM, we used likelihood-ratio testing (LRT) to assess the statistical significance of individual effects by comparing models with and without the effect of interest, starting form the maximal structure and gradually removing effects.

Since the *phia* package is not capable of making comparisons based on CLMMs, posthoc comparisons of means for comprehension task data were carried out by first constructing an Anova table with Group and Condition as factor (and their interaction if it was found to be significant in the CLMM) and then computing multiple comparisons of groups means from the Anova using the Bonferroni correction, with package *emmeans*.

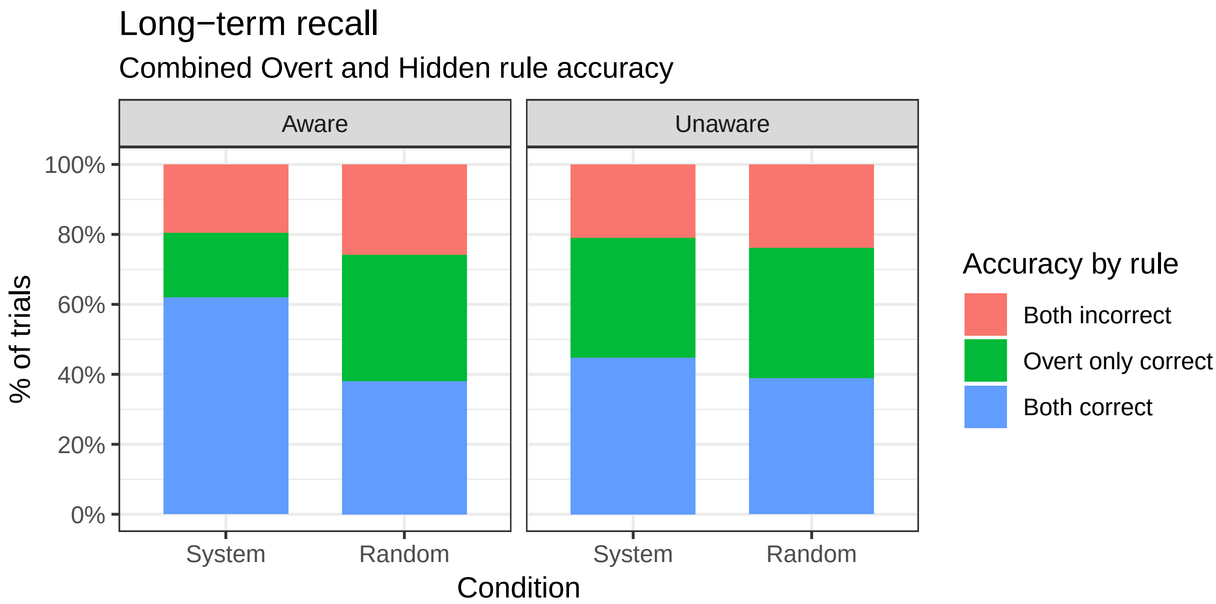


Figure 4.4: Combined Overt and Hidden rule accuracy by Group and Condition during long-term recall (Production task).

4.3.4.4 Training phase

Both Overt rule accuracy (Aware: System 97%, Random 96%; Unaware: System 95%, Random 91%) and Hidden rule accuracy (Aware: System 98%, Random 97%; Unaware: System 98%, Random 97%) were very high during the training phase (Fig. 4.3).

There were no significant differences in Hidden rule accuracy: we fitted a GLMER with the maximal fixed effects structure (Group, Condition and Block together with all their interactions), random intercepts for subjects and items, random slopes for subjects by block and for items by condition; no effects were found for any of the predictors.

4.3.4.5 Testing phase

Long-term recall Overt rule accuracy, Aware: System 80%, Random 74%; Unaware: System 79%, Random 76% (Fig. 4.4). Hidden rule accuracy, Aware: System 77%, Random 51%; Unaware: System 57%, Random 51% (Fig. 4.5). To analyse Hidden rule accuracy, we fitted a GLMER including Group, Condition,

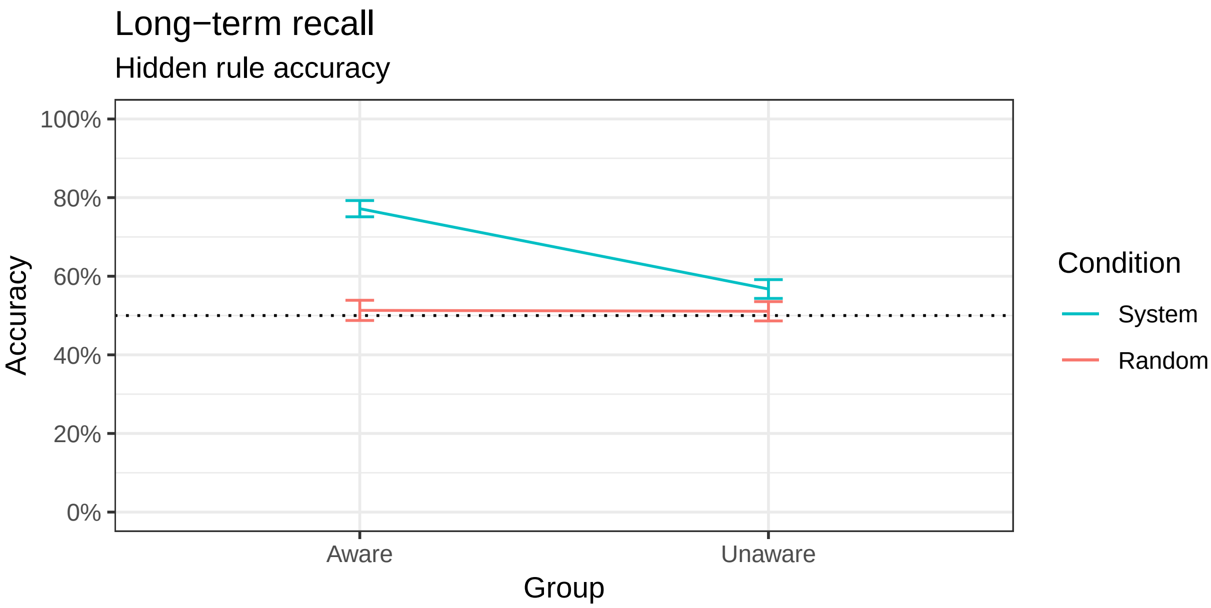


Figure 4.5: Average Hidden rule accuracy by Group and Condition during Long-term recall. Error bars represent SE of the mean, dotted line marks 50% chance level.

Block and their interactions as fixed effects, and random slopes for subjects and items (more complex models failed to converge). We found significant effects of Condition ($\chi^2(1) = 40.018$, $p < .0001$) and Group ($\chi^2(1) = 4.465$, $p = .035$), as well as an interaction between Group and Condition ($\chi^2(1) = 22.511$, $p <$ .0001)(Table 4.2); there was no effect of Block on accuracy. Post-hoc comparisons of the Group x Condition interaction using the Bonferroni correction showed that the effect of Condition was significant for the Aware group ($\chi^2(1) = 59$, $p < .0001$) but not for the Unaware ($\chi^2(1) = 3$, $p = .17$). Hidden rule recall for Random items was at chance level for both groups (Aware: 51%, $t(20) = 0.8$, $p = .4$; Unaware: 51%, $t(20) =$ -0.2, $p = .8$), which suggests that participants had no item memory. Conversely, accuracy on System items was above chance for Aware participants (77%, $t(20) = 3.7$, $p = .002$) but only marginally so for the Unaware (57%, $t(20) = 2.1$, $p = .05$).

Generalisation Generalisation items were novel System items derived made from pictures previously seen in training, but in the opposite condition, and they

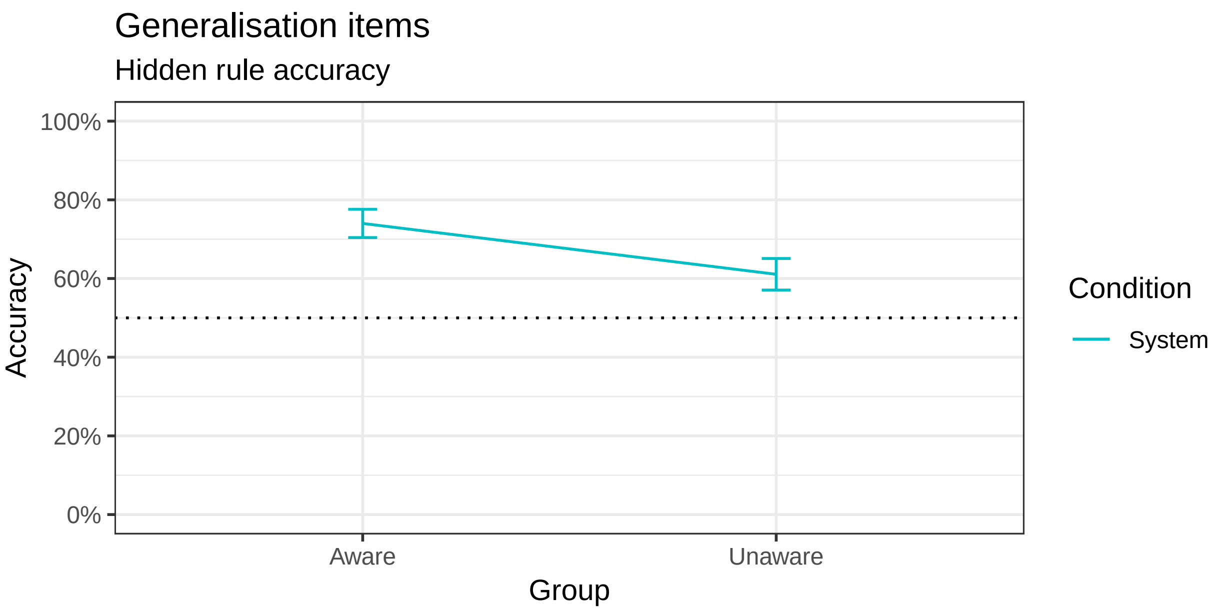


Figure 4.6: Average Hidden rule accuracy by Group for generalisation items. Error bars represent SE of the mean, dotted line marks 50% chance level.

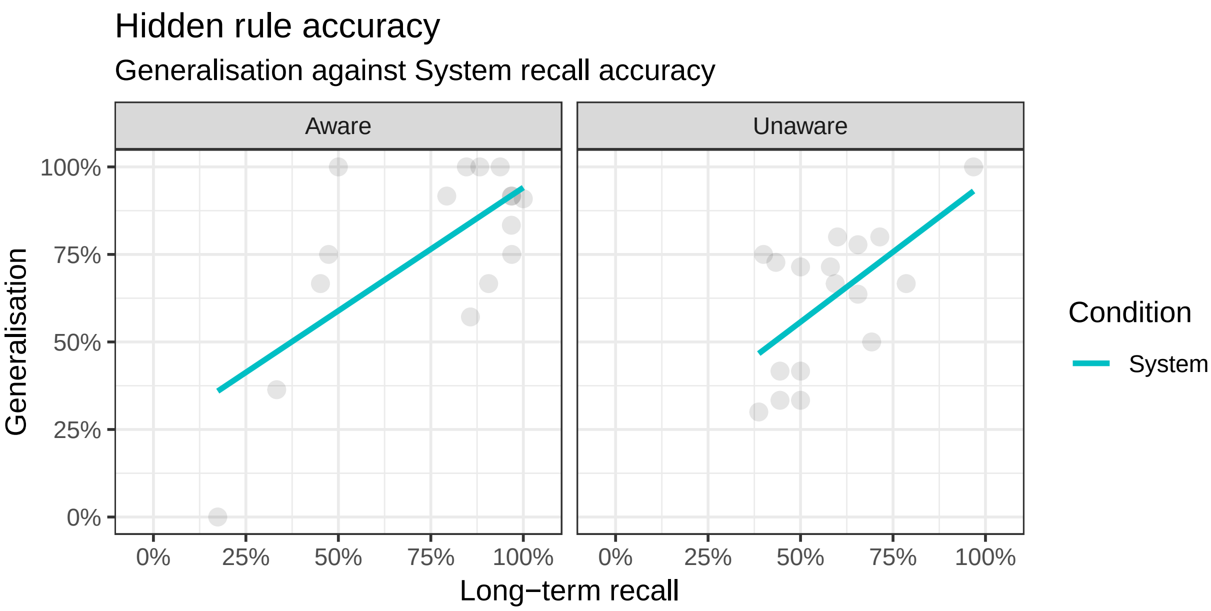


Figure 4.7: Average Hidden rule accuracy during Generalisation plotted against average accuracy for System items during Long-term recall task.

Hidden rule accuracy, Long-term recall

	Coefficient	SE	χ^2	df	p	
Group	-1.030	0.26	4.465	1	0.035	*
Condition	-1.288	0.17	40.018	1	0.000	***
Block	0.076	0.12	1.221	1	0.269	
Group x Condition	1.045	0.22	22.511	1	0.000	***
Group x Block	0.092	0.16	0.002	1	0.961	
Condition x Block	-0.031	0.17	1.562	1	0.211	
Group x Condition x Block	-0.183	0.22	0.701	1	0.402	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 4.2: Summary of GLMER model for Hidden rule accuracy, Long-term recall.

were presented intermixed with Recall items during the Production task. Hidden rule accuracy on Generalisation items was 74% for Aware participants, and 61% for Unaware (Fig. 4.6). We carried out one-sample t-tests against chance level (50%) to assess whether participants were above chance in their accuracy. Both groups were significantly above chance (Aware: $t(15) = 3.9$, $p = .001$; Unaware: $t(16) = 2.5$, $p = .025$). Accuracy on generalisation items positively correlated with average recall accuracy for System items in the Long-term recall task for both Aware (Pearson's $R = .69$, $p = .003$) and Unaware participants (Pearson's $R = .61$, $p = .009$) (Fig. 4.7).

Recognition memory task Average ratings for Old items (Fig. 4.8) were higher for System items than for Random ones (Aware: System 4.6, Random 3.8; Unaware: System 4.6, Random 4.2). We constructed a CLMM with random intercepts for subjects and items and correlated random slopes for subjects by condition, entering Group and Condition as fixed factors. Stepwise model simplification revealed a main effect of Condition ($\chi^2(1) = 9.71$, $p = .002$), which was significant for both Aware and Unaware in posthoc paired t-tests (Aware: $t(15) = -2.6$, $p = .019$; Unaware: $t(21) = -2.3$, $p = .033$).

Average ratings for New items (which included ungrammatical System items) (Fig. 4.8) followed different trends for Aware and Unaware, with a tendency

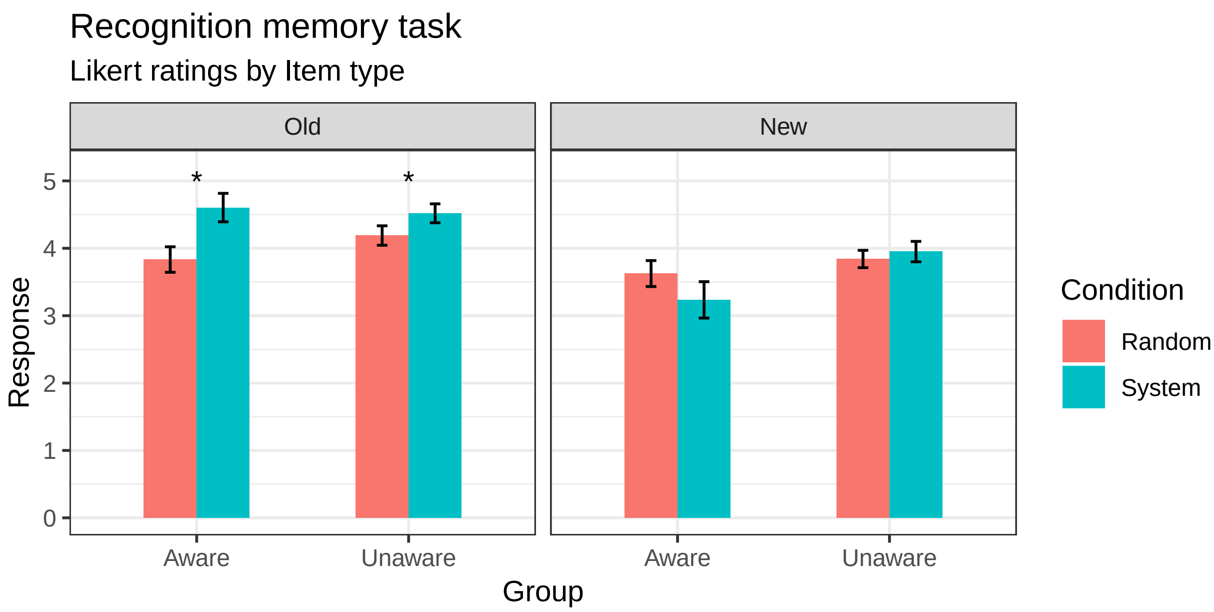


Figure 4.8: Mean ratings in the Recognition memory task by Item type, Group and Condition.

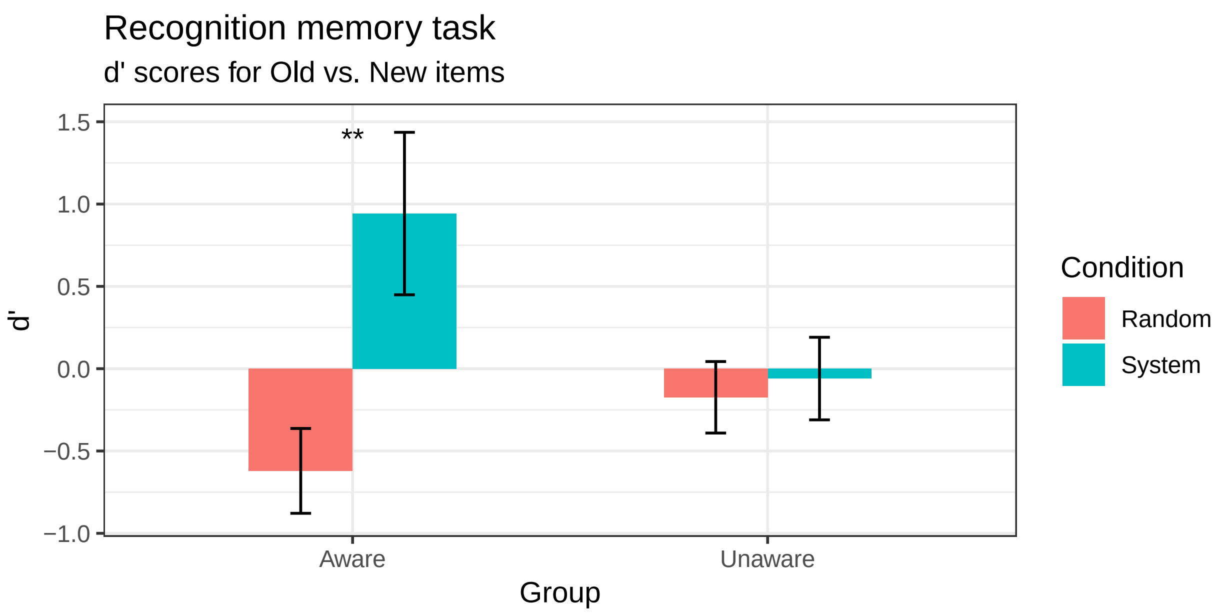


Figure 4.9: Mean d-prime scores for Old vs. New items, by Group and Condition.

among the Aware to give lower ratings for ungrammatical System items (Aware: System 3.2, Random 3.6) while the opposite was true for the Unaware group (Unaware: System 4.0, Random 3.8). We constructed a CLMM with random intercepts for subjects and items and correlated random slopes for subjects by condition, entering Group and Condition as fixed factors. There were no significant effects of any factors (Fig. 4.8).

To explore the extent to which participants were sensitive to the rule in comprehension, we carried out a comparison of Old vs. New items (Fig. 4.9). New items in the System condition were ungrammatical, while new Random item could not be, due to the lack of an underlying rule. If participants had developed sensitivity to the Hidden rule, we expected them to show sensitivity to novelty in the System condition, where it would have made items ungrammatical, than in the Random condition. In order to account for response bias (either towards or against endorsement), we compared response to Old and New items using signal detection measures (*d'*) (Fig. 5.10). Since we had used a 1-to-6 Likert scale as response variable in the recognition memory task, we first derived binary measures of endorsement by splitting the scale into two halves (responses 1 to 3: "No", 4 to 6: "Yes"). Following Tamminen et al. (2015), we then obtained *d'* scores by calculating the difference between the z-transformed proportion of correct "Yes" responses to Old items (hits) and incorrect "Yes" responses to New items (false alarms) for each subject. We entered the *d'* scores into a mixed ANOVA with Group as between-subject factors and Condition as within-subject factor, which revealed main effects of Condition ($F(1{,}36) = 5.99$, $p = .019$), and an interaction between Group and Condition ($F(1{,}36) = 5.85$, $p = .021$). Post-hoc comparisons using the Bonferroni correction showed that the effect of Condition was only significant for the Aware group ($\chi^2(1) = 19.6$, $p = .003$).

4.3.4.6 Additional analyses

L1 transfer Our findings suggest that participants had successfully developed productive knowledge of the Hidden rule, based on the the Czech *v/na* alternation. However, as previously pointed out, this distinction is in some ways similar to one found in English, the *in/on* alternation. Even though, as remarked in Sec-

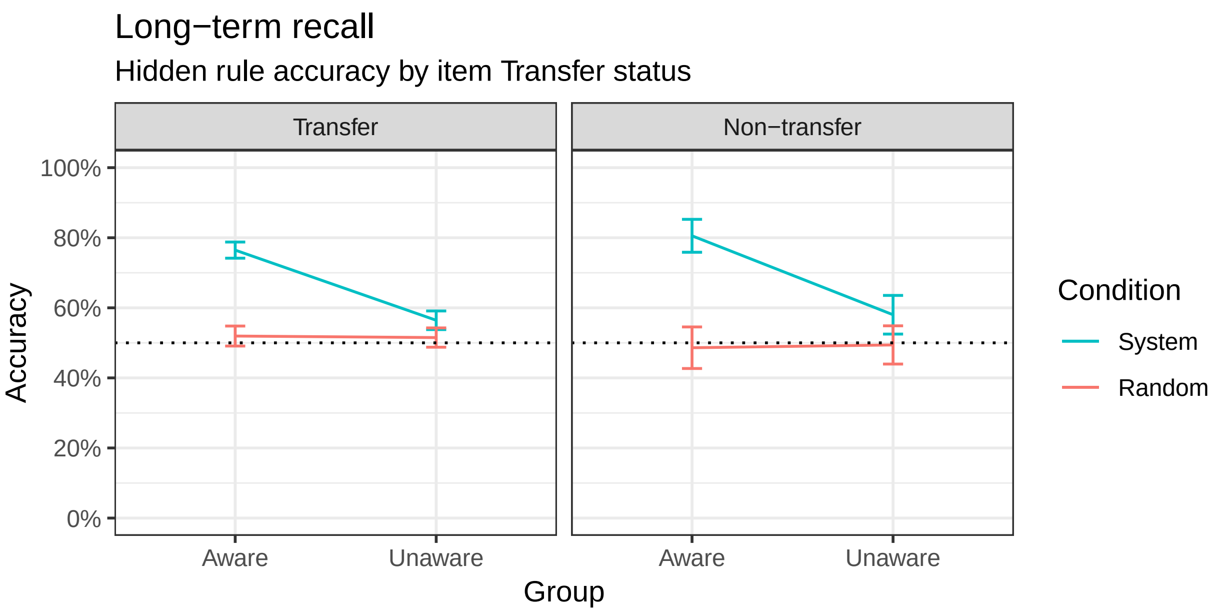


Figure 4.10: Average Hidden rule accuracy by Group, Condition and Transfer status during Long-term recall. Error bars represent SE of the mean, dotted line marks 50% chance level.

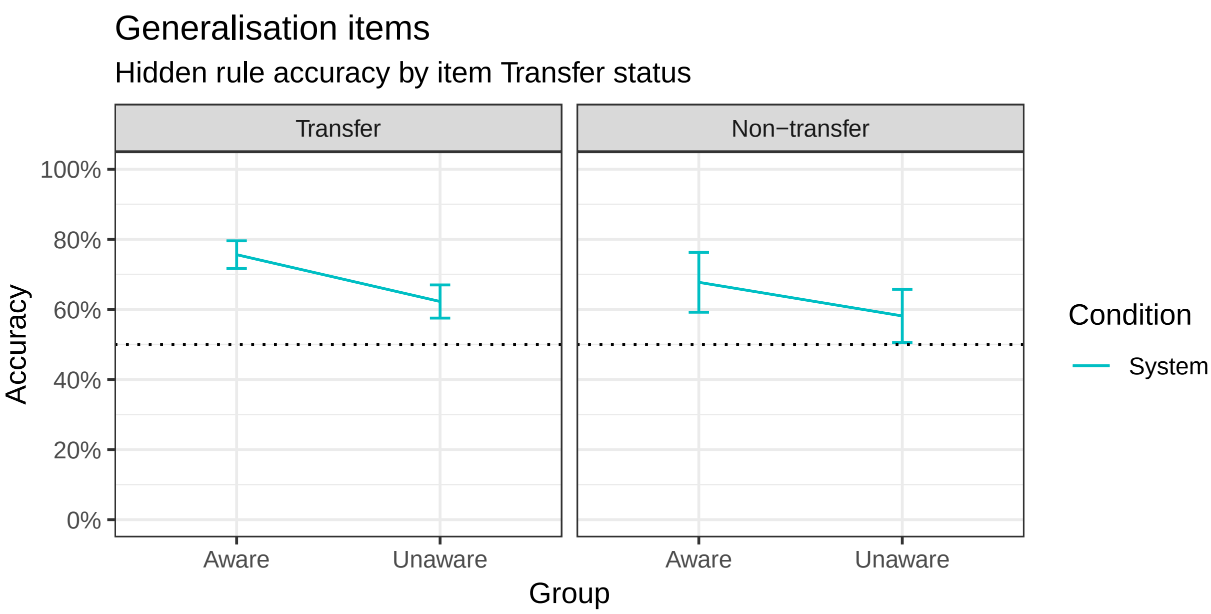


Figure 4.11: Average Hidden rule accuracy by Group and Transfer status for Generalisation items. Error bars represent SE of the mean, dotted line marks 50% chance level.

tion 3.1.1, the two rules differ in their distribution, we decided to run a norming study (as described in Section 4.3.4.3) to determine the exact extent to which L1 transfer from English could drive Hidden rule accuracy. We found that 80% of the items in our study could, in theory, be produced correctly by transferring the English *in/on* distinction. For this reason, participants who reported "discovering" the English *in/on* alternation as the rule governing preposition use in the experiment were included in the Aware group, as previously done in Experiment 1. Given previous studies showing that L1 transfer can be decisive in determining the implicit acquisition of a novel rule (Leung and Williams, 2014), it was possible that Unaware participants, too, may have unconsciously relied on the English rule in the production task. To investigate this possibility, we coded the items in our study as either Transfer or Non-transfer depending on whether they could be produced correctly by transferring the *in/on* distinction, based on the results of the norming study. We found no significant effect of Transfer on Hidden rule accuracy in either Long-term recall ($\chi^2(1) = 0.34$, $p = .56$) or Generalisation items ($\chi^2(1) = 0.60$, $p = .44$). Visual inspection of the averages (Fig. 4.10 & 4.11) shows that there is a numerical difference for both groups, with generally higher recall and generalisation accuracy for Transfer items relative to Non-transfer items. Therefore, it is possible that there may be have been some facilitation from transfer of the English *in/on* alternation alongside learning of the Hidden rule; however, this is not sufficient to explain our findings, suggesting that participants also developed sensitivity to the open vs. enclosed regularity encoded by the Hidden rule.

Breakdown of Aware group In our analysis, we included in the Aware group even participants who could not report the exact rule, but who believed that the prepositions were encoding the *in/on* distinction found in English (Section 4.3.4.3). This was taken as a conservative approach, given the high overlap between our target rule and the English one. The rationale for this choice was that these participants may benefit from applying explicit knowledge of the English rule, which, even though it was not the target rule, was close enough to the target rule to get a significant proportion of items correct. Therefore, including them in the Unaware group would potentially artificially inflate accuracy rates in the

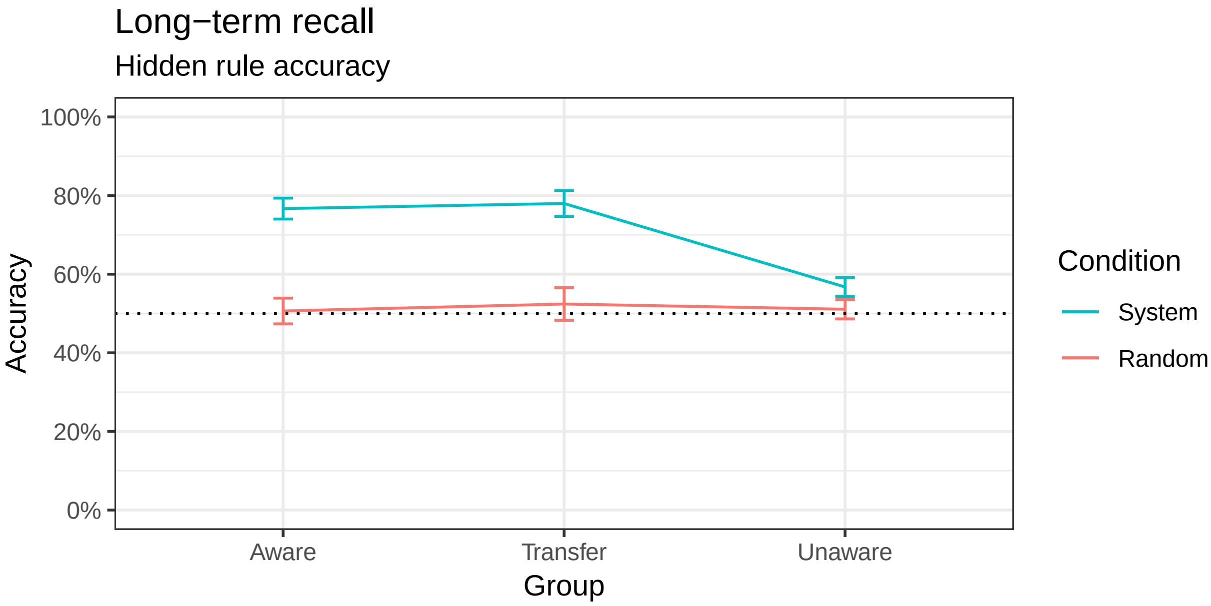


Figure 4.12: Average Hidden rule accuracy by Condition for Aware, Transfer and Unaware subjects during Long-term recall. Error bars represent SE of the mean, dotted line marks 50% chance level.

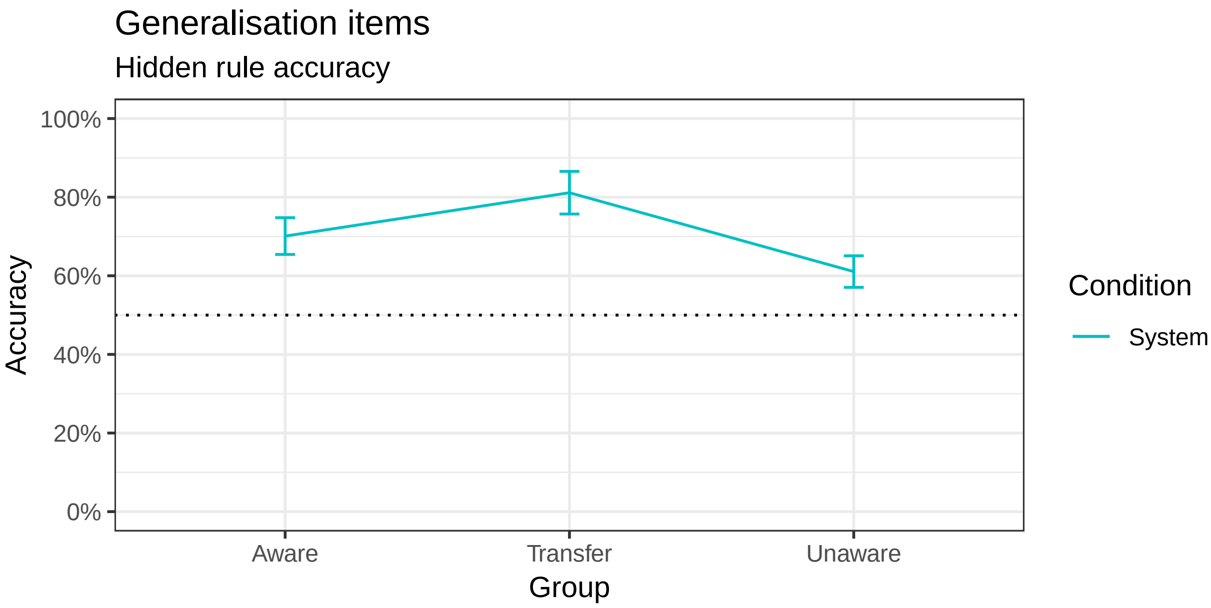


Figure 4.13: Average Hidden generalisation rule accuracy for Aware, Transfer and Unaware subjects. Error bars represent SE of the mean, dotted line marks 50% chance level.

group, masking the extent to which unaware participants had acquired implicit knowledge of the Hidden rule.

At the same time, however, further by-item analysis revealed that transfer of the English *in/on* distinction did not seem to play a significant role in participants' performance. Additionally, it may be argued that participants who mapped the novel prepositions onto the English rule were not 'aware' of our rule in any traditional sense of the word, and should therefore not be treated as such. In fact, their questionnaire responses could even indicate that they had strong implicit knowledge of the Hidden rule, but were not able to verbalise it; therefore, they resorted to the closest rule they could verbalise, that is, the English distinction. To explore these possibilities, we conducted a further analysis breaking down participants into three groups: Aware (who could verbalise the Hidden rule), Transfer (who reported the English rule), and Unaware (who could not report any rule).

In recall, Transfer participants performed very similarly to the Aware group (Fig. 4.12). A glmer with Group and Condition revealed an interaction between Group and Condition ($\chi^2(1) = 22.62$, p ¡ .0001). In the Random condition, all groups were at chance level. In the System condition, there was a significant difference between Aware and Unaware ($\chi^2(1) = 11.28$, $p = .005$) and between Transfer and Unaware ($\chi^2(1) = 9.51$, $p = .012$), but not between Aware and Transfer. In generalisation, we found no significant effect of Group in a glmer with Group as fixed effect, and random intercepts for subjects and items. ($\chi^2(1) = 3.72$, $p = .16$). However, visual inspection shows that the Transfer group was closer to the Aware group than to the Unaware groups in terms of accuracy, and had the highest accuracy overall numerically (Fig. 4.13)

We then looked again at the effect of item Transfer status, this time entering the three groups (Aware, Transfer and Unaware) as levels for Group, in order to assess whether the Transfer group would benefit more from Transfer items relative to the other groups. In long-term recall, there was no significant effect of Transfer status ($\chi^2(1) = .33$, $p = .57$): all groups performed similarly regardless of item Transfer status (Fig. 4.14). In generalisation, we found no statistically significant effect of Transfer status, either ($\chi^2(1) = 3.20$, $p = .20$). Visual inspection of the data (Fig. 4.15) shows opposite trends for the Aware group, who performed better

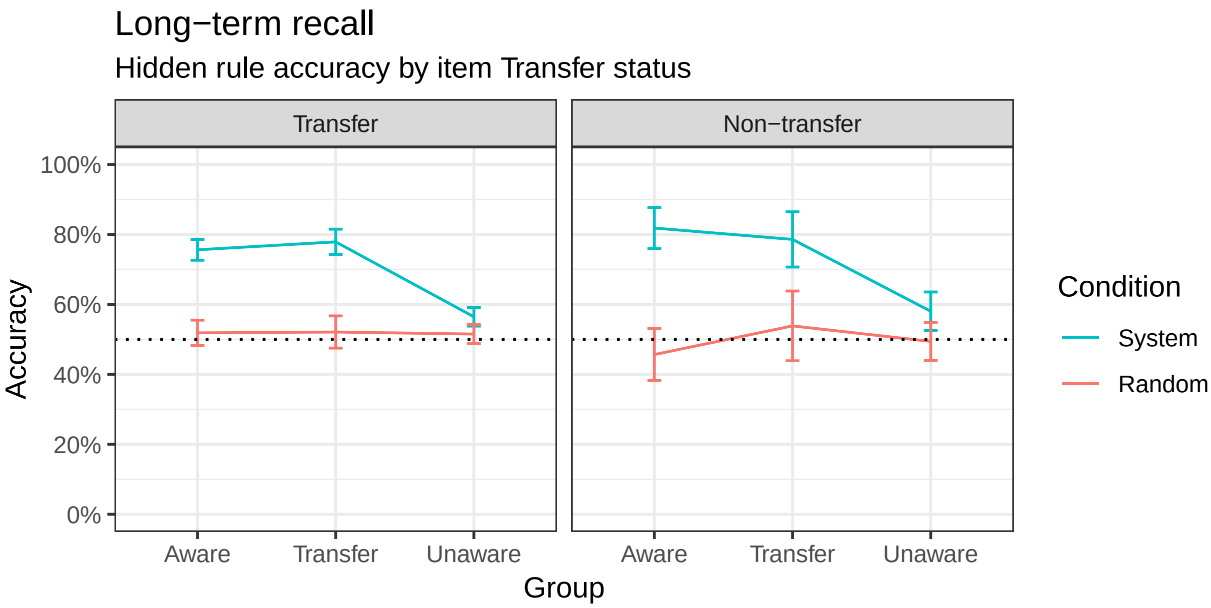


Figure 4.14: Average Hidden rule accuracy by Condition and Transfer status during Long-term recall for Aware, Transfer and Unaware subjects. Error bars represent SE of the mean, dotted line marks 50% chance level.

on Transfer items, and for the Transfer group, who, somewhat counterintuitively, tended to perform better on non-Transfer items. Numerically, the Transfer group performed best overall (however, it should also be emphasised that the data sample for this analysis was quite small, given the small amount of generalisation items relative to recall ones, combined with the fact that only 20% of items where classed as Transfer, so there could be spurious variation).

4.3.5 Discussion

4.3.5.1 Development of productive knowledge

In this experiment, as in the previous one, we predicted that if participants had acquired knowledge of the Hidden rule, they should be more accurate when recalling System items relative to Random ones. Indeed, we found a significant advantage for the System condition in Hidden rule recall; however, at the group level this was only significant for Aware participants, even though the Unaware showed a trend in the same direction. We also introduced generalisation items

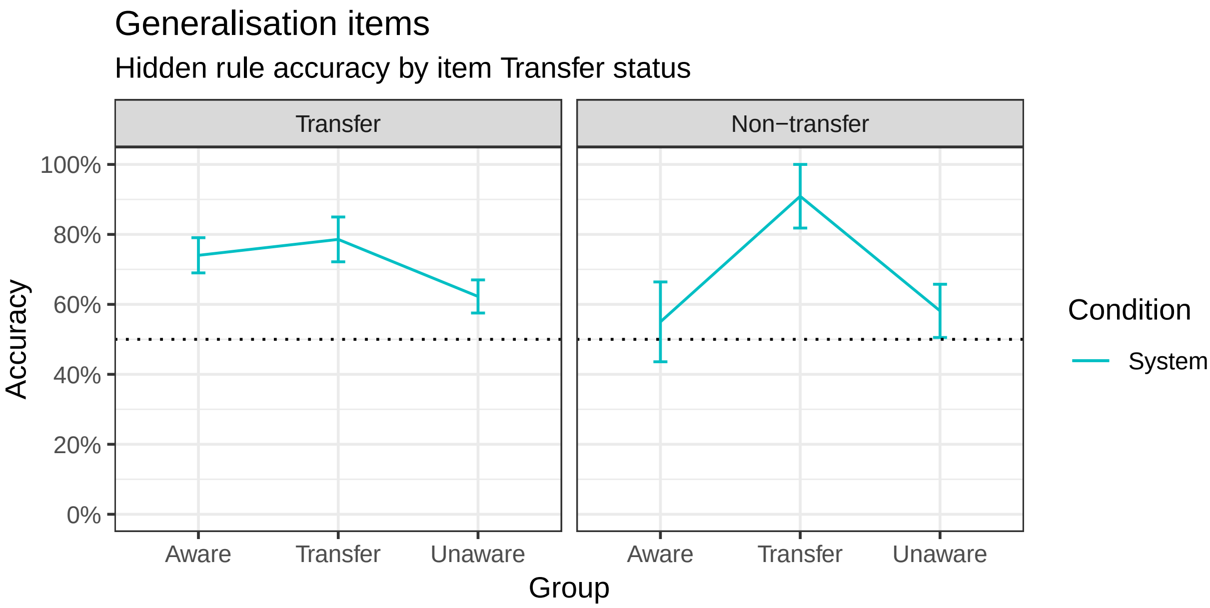


Figure 4.15: Average Hidden rule generalisation accuracy by Transfer status for Aware, Transfer and Unaware subjects. Error bars represent SE of the mean, dotted line marks 50% chance level.

- familiar pictures used to form novel picture-character combinations in the System condition, as a way to probe for implicit knowledge of the Hidden rule, which may be otherwise masked by reliance on item memory. Analysing performance on these items, we found evidence of productive generalisation of the Hidden rule for both groups: both the Aware (accuracy: 74%) and Unaware group (accuracy: 61%) were significantly above chance in their Hidden rule accuracy when producing novel System items. For the Unaware, average accuracy on Generalisation items was higher than that for System items in the recall task (57% in Recall, 61% in Generalisation. Even though the difference between the two averages is not statistically significant, it is compatible with the hypothesis that item memory may create "noise" which impairs the use of implicit rule knowledge, meaning generalisation items created more favourable conditions for the expression of implicit knowledge. Nevertheless, recall accuracy for System items significantly correlated with generalisation accuracy for the Unaware group as well as the Aware, suggesting that it too was driven by sensitivity to the Hidden rule. Finally, performance on Random items was at chance for both groups, suggesting that participants had little or no memory for specific items.

We also looked at the effect of L1 transfer on performance in the production task. Overall, transfer of L1 categories had no significant impact on accuracy, even though it may have caused some facilitation. When we divided the Aware groups into Aware and Transfer subgroups based on the rule they had reported (the exact Hidden rule, or the English *in/on* distinction), we found that the Transfer group performed very similarly to the Aware group, and significantly better than the Unaware group in recall. In generalisation, they numerically outperformed the Aware group, even though the effect was not statistically significant. However, their performance was not significantly affected by item Transfer status, either - in fact, in generalisation there was a trend towards better performance for non-transfer items, and they appeared to perform better even than the Aware group. It is therefore possible that the Transfer subgroup had developed strong implicit knowledge of the Hidden rule, but that they were unable to verbalise it in the questionnaire and as a result produced the closest alternation of which they had explicit knowledge, namely the English *in/on* rule.

4.3.5.2 Development of comprehension skills

In the recognition memory task, we exposed participants to two kinds of items: Old items (previously seen in training), and New items, which were composed of familiar pictures in novel combinations. New items in the System condition were designed to be ungrammatical (using the wrong System preposition for the type of place), so as to assess the effect of rule violation on recognition.

When asked whether they recognised items previously encountered in the Training phase, both groups were significantly more likely to endorse System items compared to Random ones, which suggests they had developed sensitivity to the Hidden rule. However, the endorsement bias for System items may also be the consequence of a general preference or sense of cognitive fluency for these items - for instance, the fact that the character and the picture are closer together may make the items easier to perceive and encode (this and other possible reasons for System advantage across experiments will be discussed in greater detail in Chapter 8). In order to probe the possible development of receptive rule knowledge, the recognition memory task also included new ungrammatical System items, which we compared to new Random items. If participants were sensitive to the Hidden rule, we predicted that they should be more likely to reject the new System items, which violated the rule, compared to new Random items. While we saw a trend in this direction for the Aware group, however, it was not significant. Sensitivity to Hidden rule violation, however, emerged when using a more sensitive measure *d'*, which accounted for the bias towards endorsing System items by comparing items within conditions.

In the Unaware group, we saw a tendency to endorse System items over Random items even when ungrammatical. The direct comparison of System to Random items suggests that there may be a general preference for System items, which drives endorsement patterns for old items and even partially counteracts the effect of explicit knowledge in the Aware in endorsing violation items. It could be the case that, due to the more compact visual representation, System items are easier to encode or simply have higher "cognitive fluency" than Random ones. In an AGL study, Kinder et al. (2003) found that the visual fluency of stimuli during training (manipulated by masking them with visual noise) affected how likely participants were to judge them as grammatical in a subsequent judgement task. However, in the case of the Unaware group we found no difference between conditions even when comparing *d'* scores for Seen vs. New items, which suggests that they had no sensitivity to Hidden rule violation. Overall, we found no evidence of implicit knowledge transfer from the production to the comprehension task. We see higher endorsement rates for old System items compared to Random ones, but no evidence of sensitivity to violation in the Unaware. However, it is possible that implicit knowledge may have emerged if participants had been tested on new grammatical System items, instead of just ungrammatical ones: to address this issue, the recognition memory task in the next experiment will include separate violation and generalisation items, in order to better assess the extent of any receptive implicit knowledge participants may have developed.

Chapter 5

Experiment 3: Modified training paradigm

5.1 Introduction

In this chapter, we report the findings of a new experiment (Experiment 3), in which we adopted a different method for generating memory load, relative to Experiment 2. We no longer used pseudowords to replace of place nouns: apart from the novel prepositions, all the words participants heard were in English. Instead, we inserted questions after each item presentation (Figure 5.1, as described in Section 5.2.3.1). The goal of this manipulation was to increase the effectiveness of the training paradigm, in two ways. The first question was aimed at drawing participants' attention to the physical properties of the place, which are crucial to the Hidden rule. We hypothesised that this would strengthen the association between the place's characteristics (open vs. enclosed space) and the sentence participants had just heard, promoting acquisition of the Hidden rule. The second question was aimed at promoting knowledge of the Overt rule, by probing participants' understanding of the relation between the preposition used and the character's position relative to the picture on screen. Since the Hidden rule is actually a sub-rule of the Overt one - it is necessary to get the Overt rule right in order to get the Hidden rule right too - higher accuracy on the Overt rule would result in a greater number of trials in which the Hidden rule could be applied.

As in Experiment 2, both a long-term recall task and generalisation items were included in the testing phase (Section 5.2.3.2). We also made changes to the types of items included in the comprehension task (outlined in Section 5.2.3.2), in order to avoid confounds that emerged in Experiment 2 and clearly separate the effect of novelty from that of rule violation.

As in the previous experiment, we would expect participants to have better recall for System items and to be be above chance in producing the correct prepositions for new System items, if they had acquired knowledge of the Hidden rule. If changes to the training paradigm had the intended effect, we would also expect accuracy for both rules to be higher than in Experiment 2. Finally, in the recognition memory task, we would expect to see higher endorsement rates for grammatical System items relative to Random ones, both old and new ones, but lower endorsement rates for ungrammatical System items.

5.2 Experiment 3

5.2.1 Participants

42 native English speakers (39 females, mean age 20.7 years) from the University of Cambridge and surrounding community took part in the experiment, receiving £6 as compensation. The foreign languages most commonly spoken by participants were French (n = 15), Spanish (n = 10) and German (n = 7), followed by Russian (n = 2), Mandarin (n = 2), Malay (n = 2), Swedish (n = 2), Italian (n = 1), Turkish (n = 1) and Kurdish (n = 1).

5.2.2 Materials

We used the same items used for Experiment 2. We only made a change the order in which recall and generalisation items were presented during the test (see Section 5.2.3.2 for details).

5.2.3 Procedure

The general structure of the experiment was very close to that used of Experiment 2, with a training phase followed by production and comprehension testing. The main differences were: changes to the training procedure (Section 5.2.3.1), item sequencing during the Production task (Section 5.2.3.2), and item types used in the recognition memory task at the end of the experiment (Table 5.1).

5.2.3.1 Training phase

The items used during the training phase were the same used in Experiment 2. Participants were exposed to 112 sentences, made up of the 64 matched items used in Experiment 1 and 2 (32 unique nouns appearing once in each condition), and the 48 were the unmatched items also used in Experiment 2. Generalisation items were nouns which appeared in only one condition during training, counterbalanced across participants. The training phase followed the same basic procedure as Experiment 2, with one exception: all the place nouns were in English, as done previously in Experiment 1, and participants were instead asked questions after each item (Fig. 5.1), to which they replied by button press. For instance, after the item "Harry is *gi* desert", they would see the following question: "How much do you like this place?", to which they would reply on a scale from 1 to 5 (1 = Not at all, 5 = Very much). After giving their response, the next question would appear: "Has Harry reached the desert yet?" (Y/N). In his case, the correct answer would be "Yes" if Harry were shown superimposed onto the picture (System condition), and "No" if he were shown beside the picture (Random condition). The question would remain on screen until participants gave the correct answer, which acted as feedback on their response. The next item was then presented, followed by questions. Every two trials, participants were asked to recall the previous two items using picture cues, as done in Experiment 2.

5.2.3.2 Testing phase

The testing phase followed the same procedure as Experiment 2: participants were simply shown recall cue screens (picture + character combinations) and asked to produce the corresponding sentence, without any further exposure. Un-

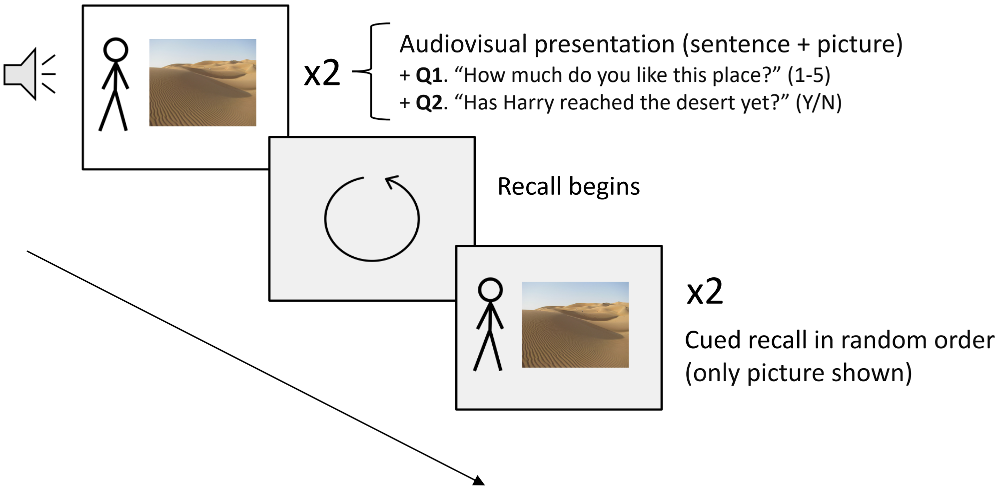


Figure 5.1: Training phase procedure for Experiment 3

known to them, some of the items presented were not from the training phase, but were novel (generalisation) items. However, unlike Experiment 2, generalisation items were presented in a block at the end of the task instead of being randomly intermixed with recall items (Fig. 5.2). The reason for intermixing Generalisation items together with Recall items in Experiment 2 had been a concern that participants may become aware of the manipulation; this was deemed less likely to occur if Generalisation items were "concealed" among Recall ones. However, the drawback of this solution was that it introduced a confounding factor of time: by spreading out a relatively small amount of Generalisation items over the entire testing phase, we increased the potential for variation due to extraneous factors, such as fatigue. Since participants in Experiment 2 did not show any awareness of the manipulation (which is supported by at-chance recall accuracy on Random items, suggesting very limited item memory), in this experiment we decided that it was safe to present Generalisation items as a block at the of the testing phase.

Long-term recall We used the same 64 matched items as Experiment 2 for long-term recall trials.

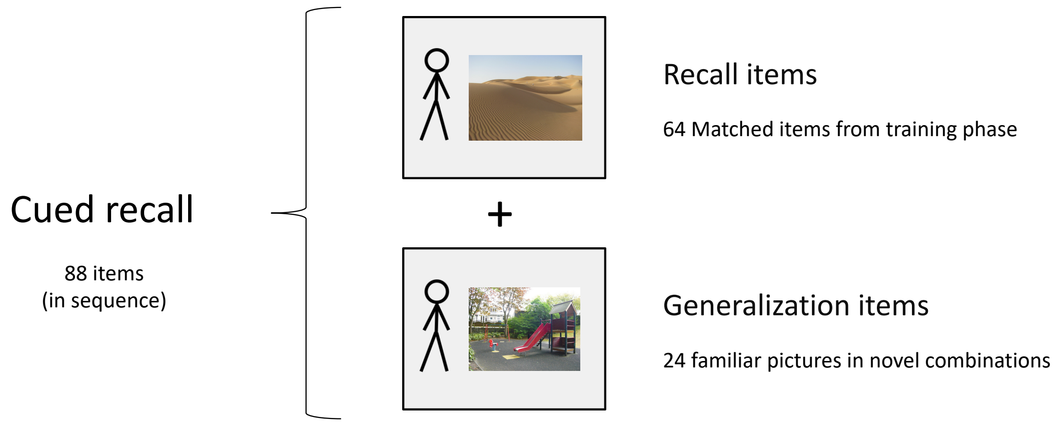


Figure 5.2: Production task procedure for Experiment 3

Generalisation We used the same 24 generalisation items as Experiment 2 for generalisation trials, but presented them all at the end of the Production task, instead of intermixing them with long-term recall trials.

Recognition memory task The comprehension task followed the same procedure as Experiment 2, but with the addition of Generalisation (new, grammatical items) as well as Violation (new, ungrammatical) ones in the System condition. Generalisation items were derived from the other half (24 items) of the 48 unmatched items from the Training phase, presented in the opposite condition. Violation items were ungrammatical versions of a subset of the matched items from training (the same ones used for Long-term recall trials), modified by inverting their preposition assignment, which made System items ungrammatical. Finally, Seen items (old items) were taken from another subset of the 64 matched items from training. The way in which the new items were created meant that they were not matched for picture familiarity. Violation items were made from pictures that had already been seen twice during training (once per condition), while Generalisation items were made from pictures that had already been seen once (Table 5.1).

	System	**Random**	Times picture seen in training:
Seen	Grammatical	Neutral	2
Violation	Ungrammatical	Neutral	2
Generalisation	Grammatical	Neutral	1

Table 5.1: Summary of item types used in Recognition memory task and their grammaticality.

5.2.3.3 Debriefing questionnaire

Participants were debriefed using the same questionnaire used in Experiments 1 and 2 (Appendix A) to assess awareness of the Overt and Hidden rules.

5.2.4 Results

5.2.4.1 Task scoring

We scored the production task following the same procedure followed in Experiment 1 and 2, and the comprehension task following the same procedure used in Experiment 2.

5.2.4.2 Rule awareness

We assessed awareness of the Overt and Hidden rule based on the debriefing questionnaire (Appendix A). We followed the scoring procedure already used for Experiment 2 (Section 4.3.4.2, p. 88), which included dividing participants into Aware, Unaware and Transfer for the purposes of Hidden rule awareness.

5.2.4.3 Analysis

One participant was excluded for failing to become aware of the Overt rule. A further two participants were excluded for technical reasons (one for failing to perform the task correctly, one due to missing data). A total of 37 subjects were included in the analysis (25 females, mean age 21 years, SD = 4.98). Of these, 8 reported awareness of the Hidden rule and a further 8 explicitly transferred the English *in* vs. *on* distinction; following Experiment 2, both groups were classed

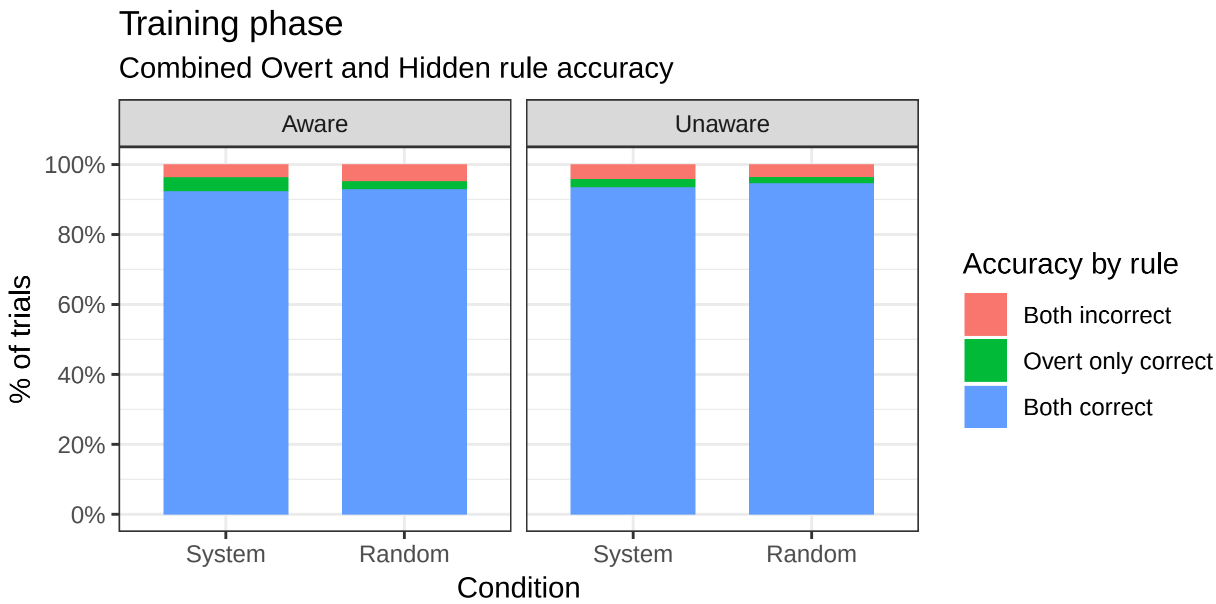


Figure 5.3: Combined Overt and Hidden rule accuracy by Group and Condition during the training phase.

as *Aware* (n = 16). The remaining 21 participants were included in the *Unaware* group.

Accuracy data was analysed with mixed-effect logistic models (GLMERs), following the same procedures for model selection and statistical inference used in Experiment 1 and 2. Comprehension (Likert scale) data was analysed with cumulative link mixed models for ordinal regression (CLMMs), following the procedures used in Experiment 2 for the same task.

5.2.4.4 Training phase

Overt rule accuracy (Fig. 5.3) was very high for both groups (Aware: System 96%, Random 95%; Unaware: System 96%, Random 97%); Hidden rule accuracy was very high, too (Aware: System 96%, Random 98%; Unaware: System 97%, Random 98%). We fitted a logistic mixed-effects model with random intercepts for subjects and correlated random slopes for subjects by condition, with Group, Condition and Block as fixed effects together with all their interactions. None of the factors entered in the model had a significant effect on Hidden rule accuracy in this phase of the experiment.

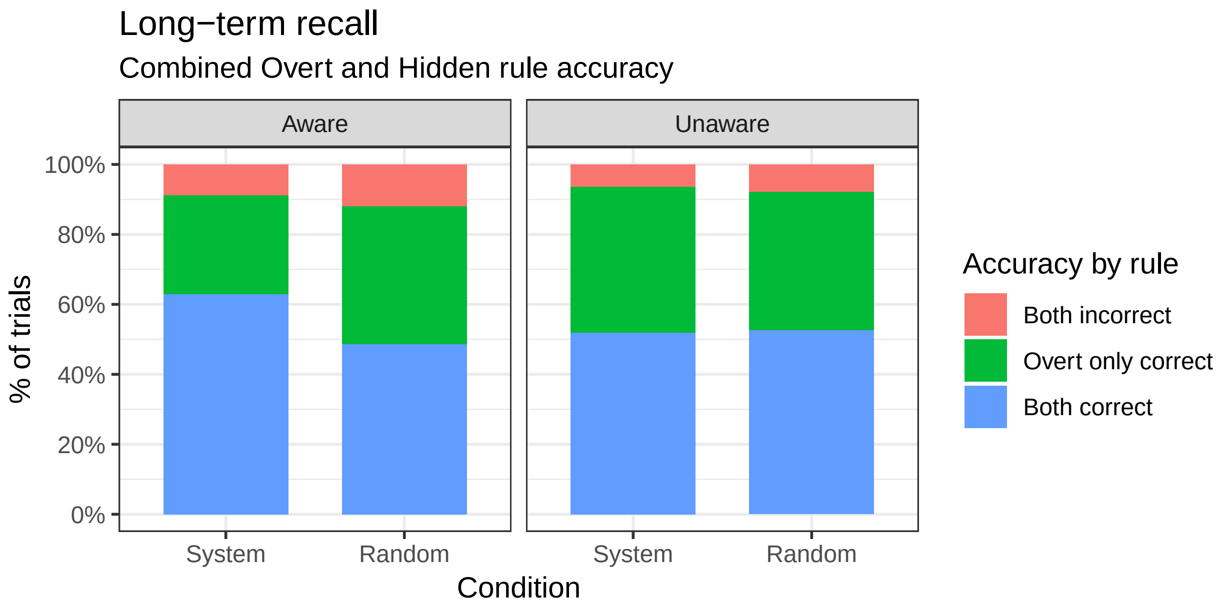


Figure 5.4: Combined Overt and Hidden rule accuracy by Group and Condition during long-term recall (Production task).

5.2.4.5 Testing phase

Long-term recall Overt Rule (Fig. 5.4) remained relatively high during Long-term recall (Aware: System 91%, Random 88%; Unaware: System 93%, Random 92%), while Hidden rule accuracy (Fig. 5.5 & 5.6) decreased for both groups (Aware: System 69%, Random 55%; Unaware: System 56%, Random 57%). We fitted a logistic mixed-effects model with random intercepts for subjects and items and correlated random slopes for subjects by condition, with Group, Condition and Block as fixed effects together with all their interactions (Table 5.2). The model revealed significant interactions between Group and Condition ($\chi^2(1) = 7.84$, $p = .005$) and between Condition and Block ($\chi^2(1) = 3.87$, $p = .049$)(Fig. 5.6). Post-hoc comparisons of the Group x Condition interaction using the Bonferroni correction showed that the effect of Condition was significant for the Aware group ($\chi^2(1) = 9.36$, $p = 0.004$) but not for the Unaware ($\chi^2(1) = 0.48$, $p = 0.979$)

Generalisation As in Experiment 2, Generalisation items were novel System items derived from pictures previously seen in training as part of Random items. Average Hidden rule accuracy rates for Generalisation items (Fig. 5.7) were 69%

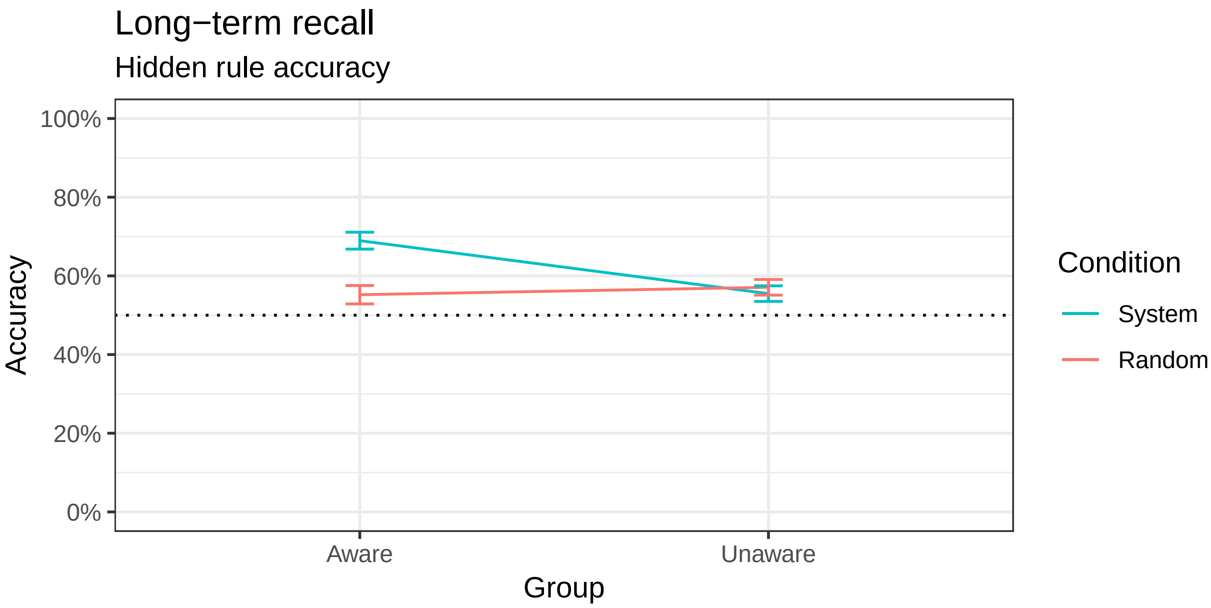


Figure 5.5: Average Hidden rule accuracy by Group and Condition during Long-term recall. Error bars represent SE of the mean, dotted line marks 50% chance level.

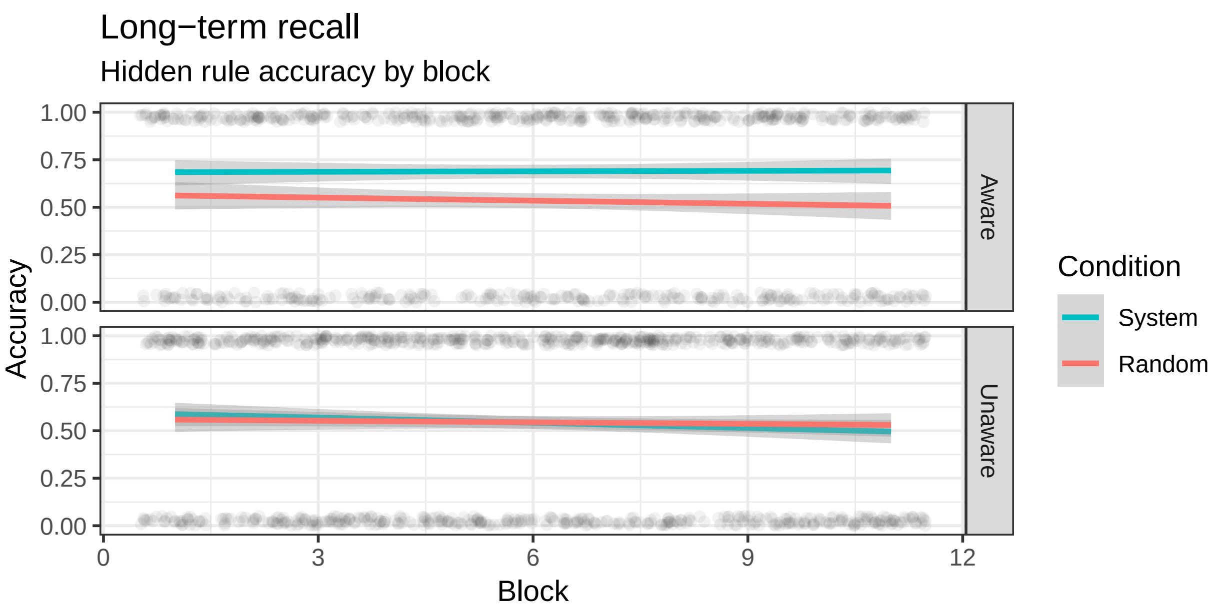


Figure 5.6: Hidden rule accuracy by block during the Production task, with fitted linear trend.

Hidden rule accuracy, Long-term recall

	Coefficient	SE	χ^2	df	p	
Group	-0.700	0.256	0.27	1	0.605	
Condition	-0.694	0.225	2.14	1	0.143	
Block	0.026	0.091	0.78	1	0.376	
Group x Condition	0.821	0.294	7.84	1	0.005	**
Group x Block	-0.115	0.117	0.39	1	0.530	
Condition x Block	-0.031	0.124	3.87	1	0.049	*
Group x Condition x Block	0.317	0.162	3.84	1	0.050	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 5.2: Summary of GLMER model for Hidden rule accuracy, Long-term recall.

for the Aware, and 51% for the Unaware group. Only the Aware group performed above chance in a one-sample t-test againt chance level ($t(15) = 3.3$, $p = .005$), while the Unaware group were at chance ($t(20) = 0.3$, $p = .75$).

Average accuracy on generalisation items positively correlated with recall accuracy for System items in the Long-term recall task in the Aware group (Pearson's $R = .85$, $p < .001$), but not in the Unaware (Pearson's $R = .11$, $p = .60$) (Fig. 5.8).

Recognition memory task Average ratings for Seen items (Fig. 5.9) were generally higher for System than for Random ones (Aware: System 4.8, Random 4.2; Unaware: System 4.6, Random 4.3). We constructed a CLMM with random intercepts for subjects and items and correlated random slopes for subjects by condition, entering Group and Condition as fixed factors. Stepwise model simplification revealed a main effect of Condition ($\chi^2(1) = 14.4$, $p < .001$), which was significant at the group level for both Aware ($t(15) = -3.8$, $p = .002$) and Unaware ($t(20) = -2.7$, $p = .02$).

Average ratings for Generalisation items (Fig. 5.9) were generally higher for System than Random ones (Aware: System 4.1, Random 3.8; Unaware: System 3.8, Random 3.6). We constructed a CLMM with random intercepts for subjects and items and correlated random slopes for subjects by condition, entering Group

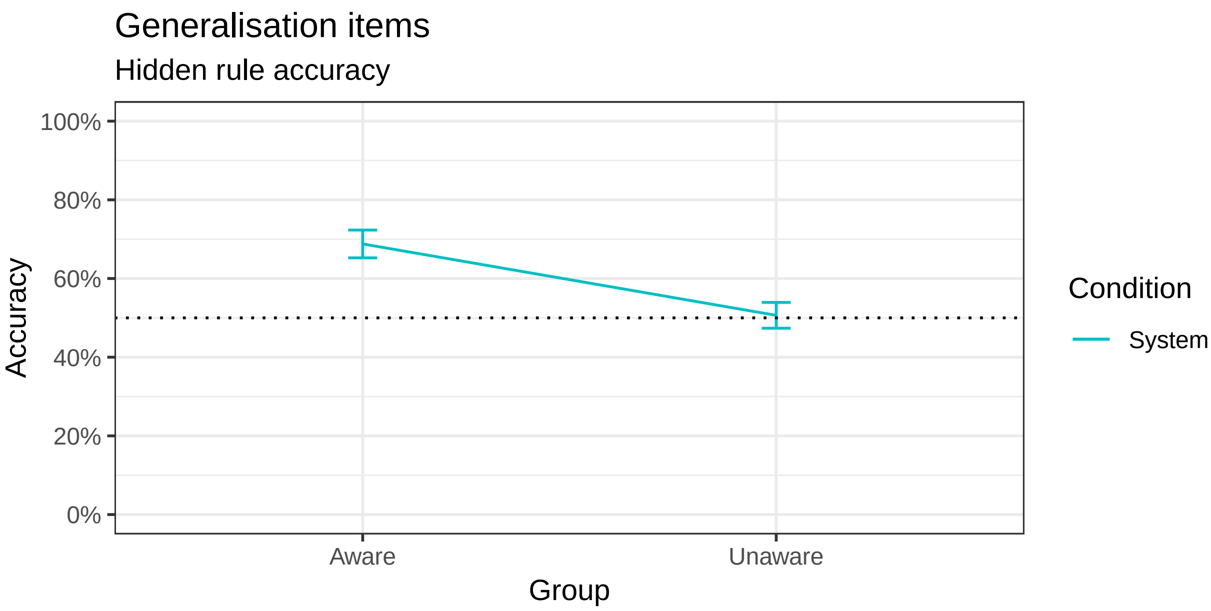


Figure 5.7: Average Hidden rule accuracy by Group for Generalisation items. Error bars represent SE of the mean, dotted line marks 50% chance level.

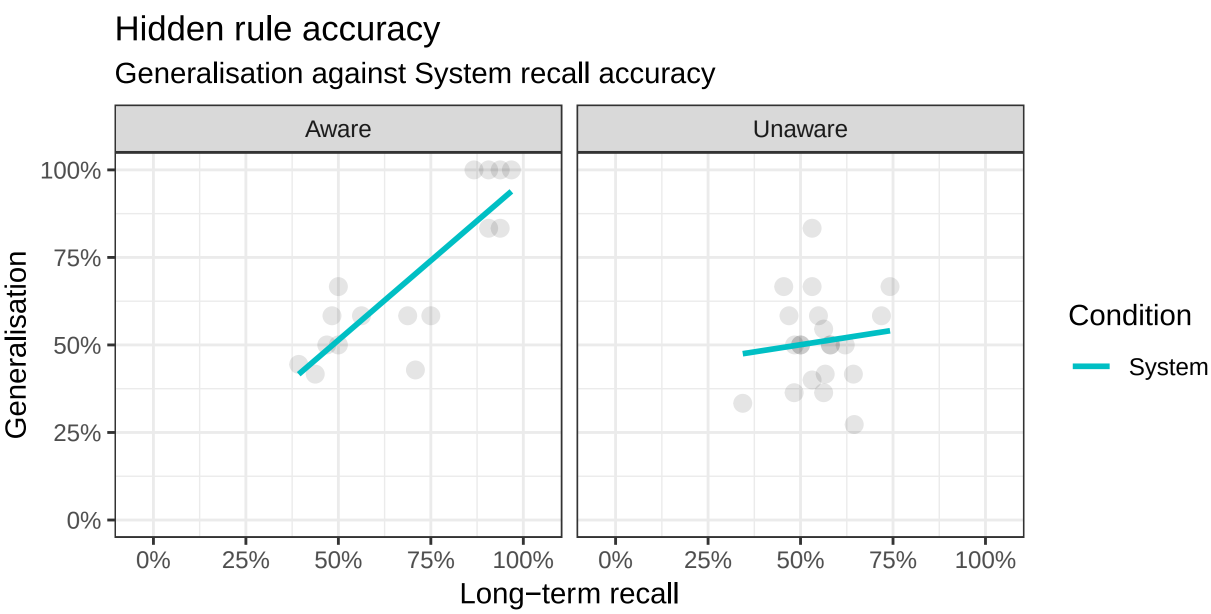


Figure 5.8: Average Hidden rule accuracy during Generalisation plotted against average accuracy for System items during Long-term recall task.

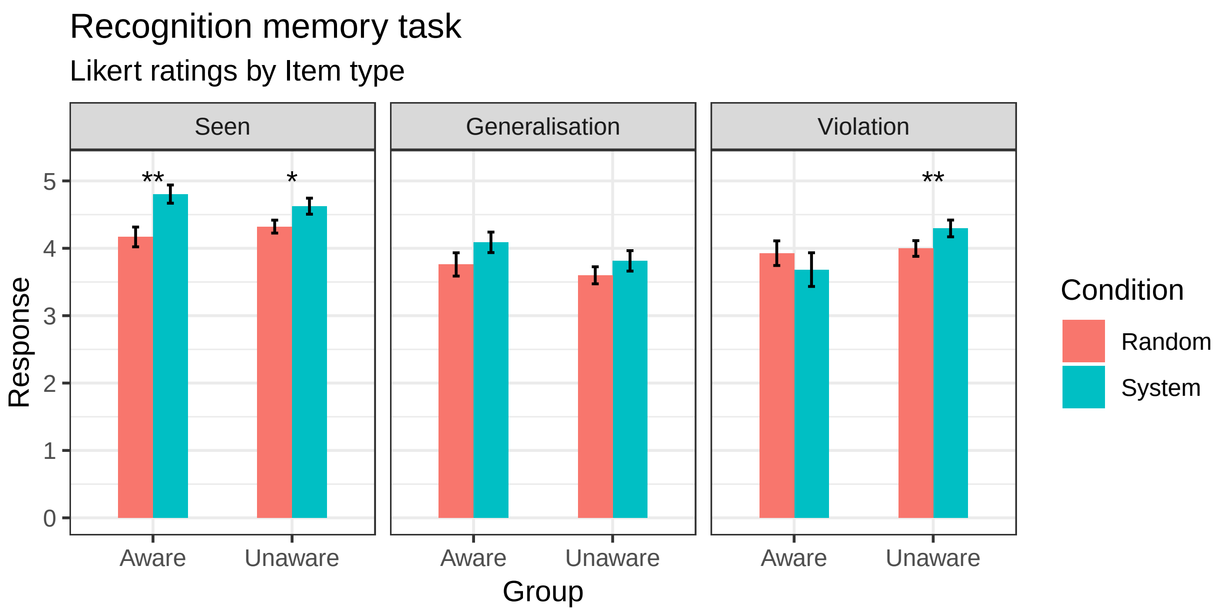


Figure 5.9: Mean ratings in the Recognition memory task by Item type, Group and Condition.

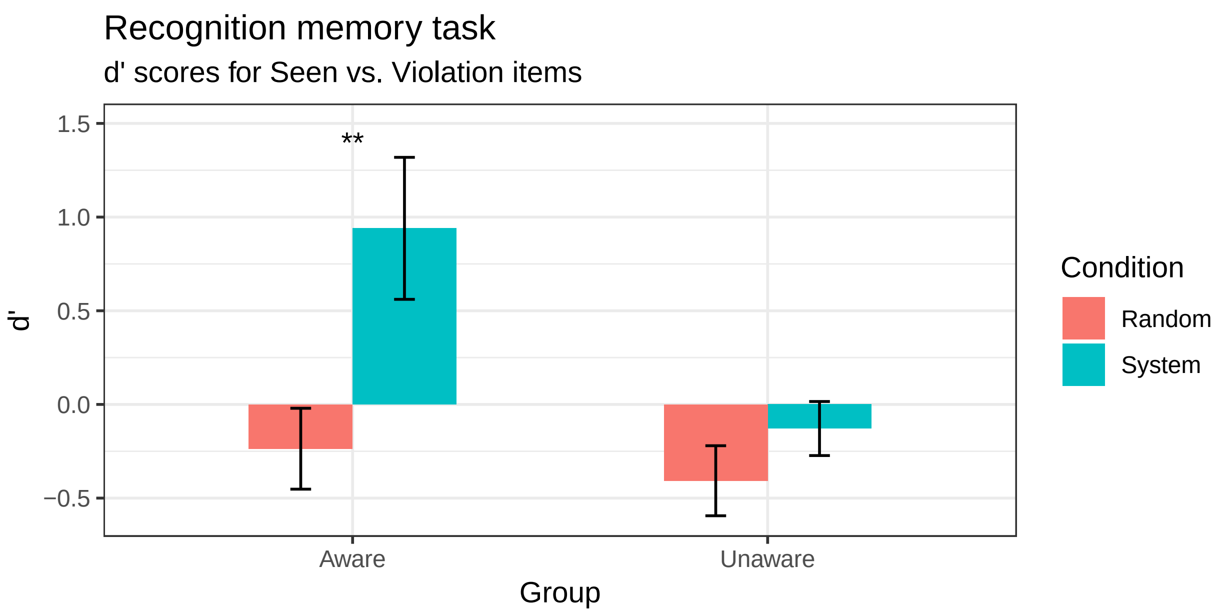


Figure 5.10: Mean d-prime scores for Seen vs. Violation items, by Group and Condition.

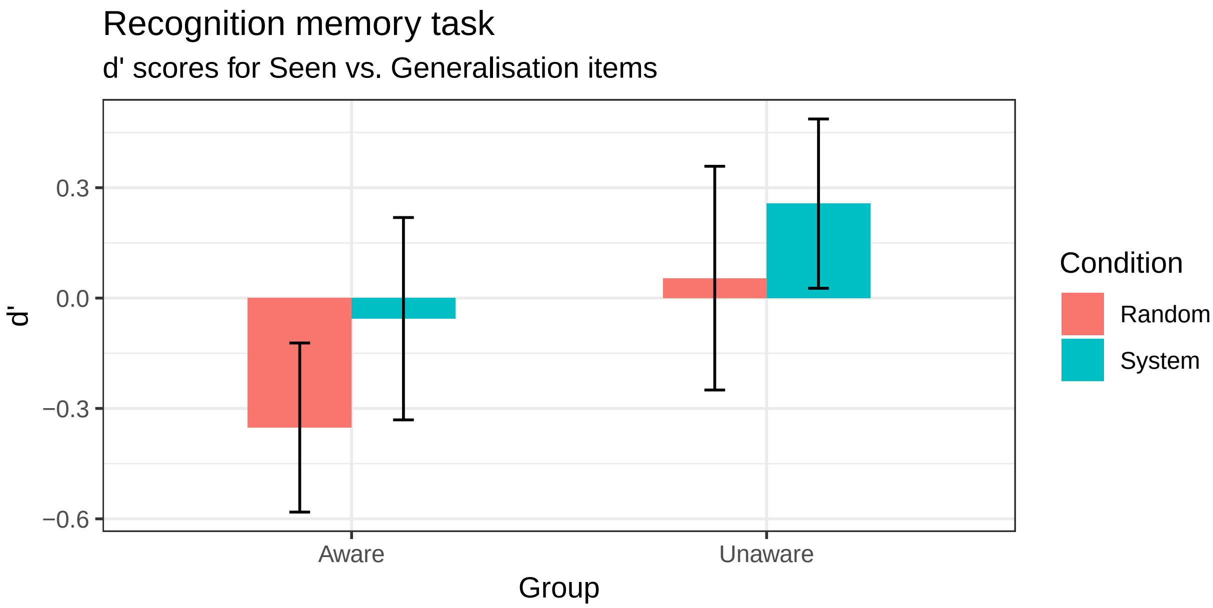


Figure 5.11: Mean d-prime scores for Seen vs. Generalisation items, by Group and Condition.

and Condition as fixed factors. Stepwise model simplification showed a main effect of Condition ($\chi^2(1) = 9.52$, p = .002). However, the difference between conditions was not significant at the group level for either Aware ($t(15)$ = -2, p = .07) or Unaware ($t(20)$ = -1.8, p = .09).

Average ratings for Violation items (old item with reversed preposition assignment) varied as a function of Group and Condition (Aware: System 3.7, Random 3.9; Unaware: System 4.3, Random 4.0) (Fig. 5.9). We constructed a CLMM with random intercepts for subjects and items and correlated random slopes for subjects by condition, entering Group and Condition as fixed factors. Stepwise model simplification revealed a significant interaction between Group and Condition ($\chi^2(1) = 6.92$, p = .009). Against predictions, Unaware participants were actually more likely to endorse System items than Random ones, while the opposite trend was found for Aware participants (Fig. 5.9). The difference was only significant for the Unaware group ($t(20)$ = -3, p = .007), not for the Aware ($t(15)$ = 1.4, p = .18).

To explore the extent to which participants were sensitive to the rule in comprehension, we carried out a comparison of Seen vs. Violation items (Fig. 5.10).

Both items types were composed of pictures seen twice in training, which were equally familiar to subjects. However, only Seen items had appeared in training; any difference in endorsement between Seen and Violation items should be due to the change in preposition. If participants had developed sensitivity to the Hidden rule, we expected them to show greater sensitivity to preposition change in the System condition, where it would have made items ungrammatical, than in the Random condition. Following the same procedure as Experiment 2, we compared response to Seen and Violation items using signal detection measures (*d'*) (Fig. 5.10). We entered the *d'* scores into a mixed ANOVA with Group as between-subject factors and Condition as within-subject factor, which revealed main effects of Group ($F(1{,}35) = 5.7$, $p = .023$), Condition ($F(1{,}35) = 11.0$, $p = .002$), and an interaction between Group and Condition ($F(1{,}35) = 4.9$, $p = .034$). Post-hoc comparisons using the Bonferroni correction showed that the effect of Condition was only significant for the Aware group ($\chi^2(1) = 11.1$, $p = .003$). We also calculated *d'* scores for Seen vs. Generalisation items, but found no significant effects of either Group or Condition (Fig. 5.11).

5.2.4.6 Additional analyses

L1 transfer As done previously for Experiment 2, we carried out further analyses to determine whether L1 transfer had affected participants' performance. We entered data in for recall and generalisation in separate glmers with Group and Transfer status as fixed predictors. In recall, Aware participants appeared to perform better on Transfer items (Fig. 5.12), but there was no statistically significant effect of Transfer, either as main effect ($\chi^2(1) = .56$, $p = .45$) or in interaction with Group ($\chi^2(1) = 1.47$, $p = .23$) . In generalisation, too, we found no significant effect of Transfer; this is supported by visual inspection of the data showing minimal differences between Transfer and Non-transfer items (Fig. 5.13).

Breakdown of Aware group We ran a further analysis of the data by separating participants from the Aware group into those that could verbalise the exact Hidden rule (Aware, n = 8) and those who transferred the English *in/on* distinction (Transfer, also n = 8). For recall, we again found a significant interaction between Group and Condition ($\chi^2(1) = .8.64$, $p = .013$). Post-hoc comparisons

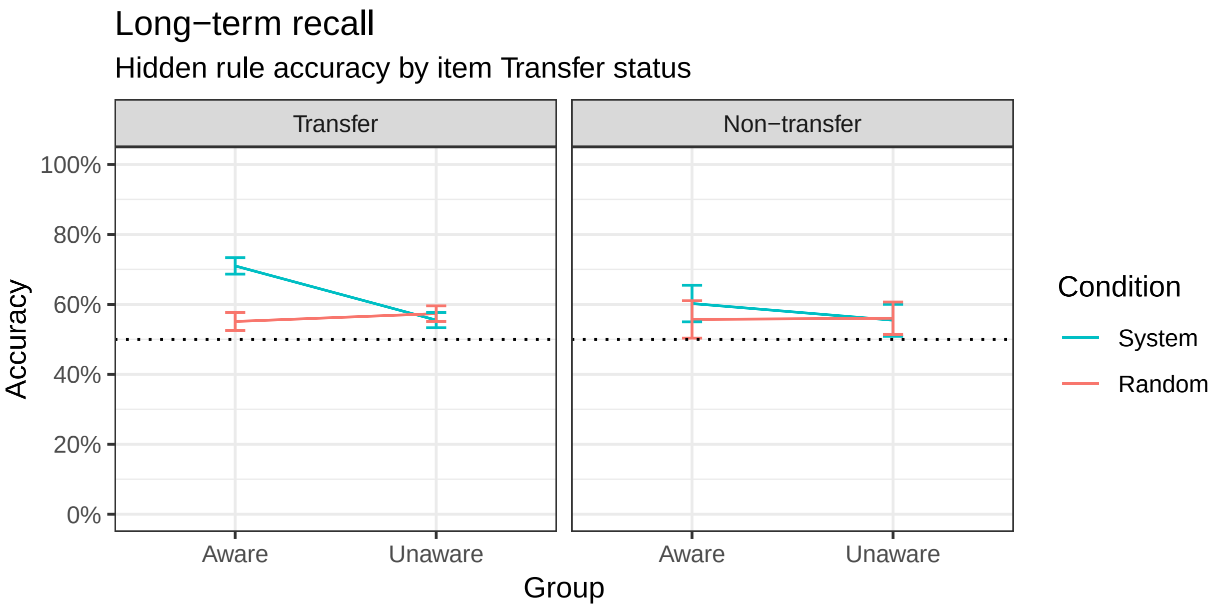


Figure 5.12: Average Hidden rule accuracy by Group, Condition and Transfer status during Long-term recall. Error bars represent SE of the mean, dotted line marks 50% chance level.

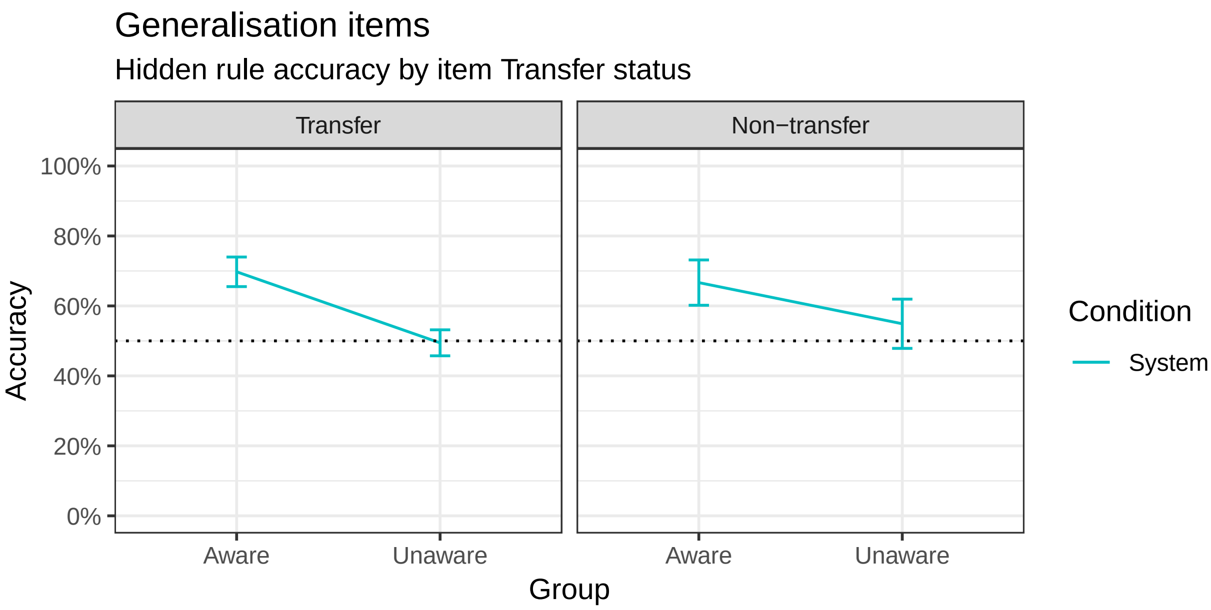


Figure 5.13: Average Hidden rule accuracy by Group and Transfer status for Generalisation items. Error bars represent SE of the mean, dotted line marks 50% chance level.

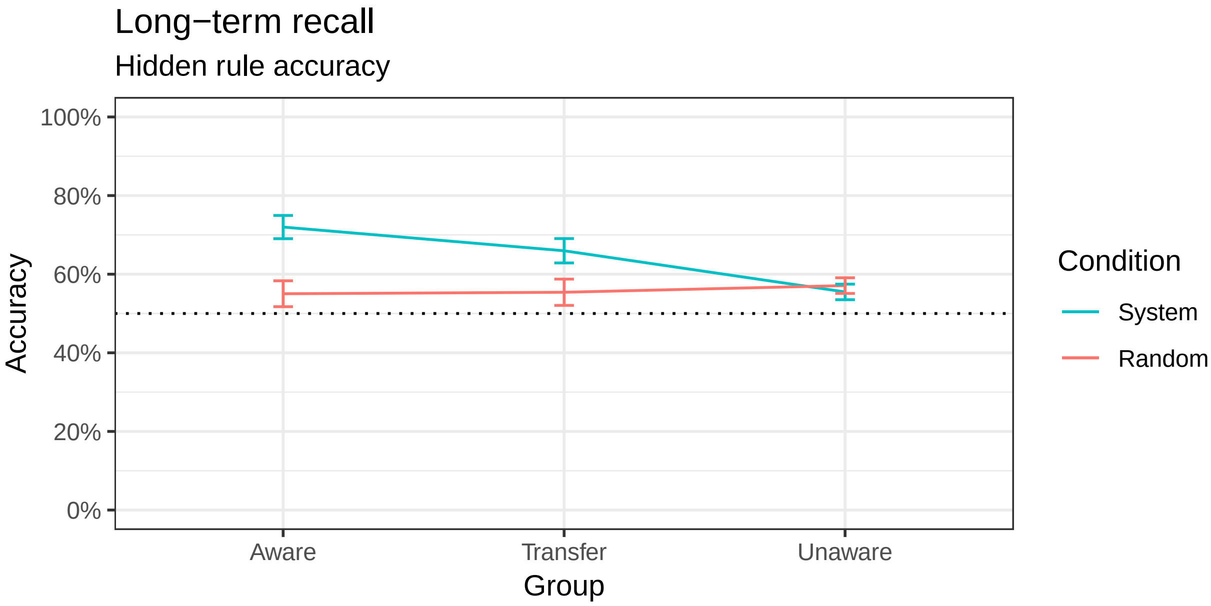


Figure 5.14: Average Hidden rule accuracy by Condition for Aware, Transfer and Unaware subjects during Long-term recall. Error bars represent SE of the mean, dotted line marks 50% chance level.

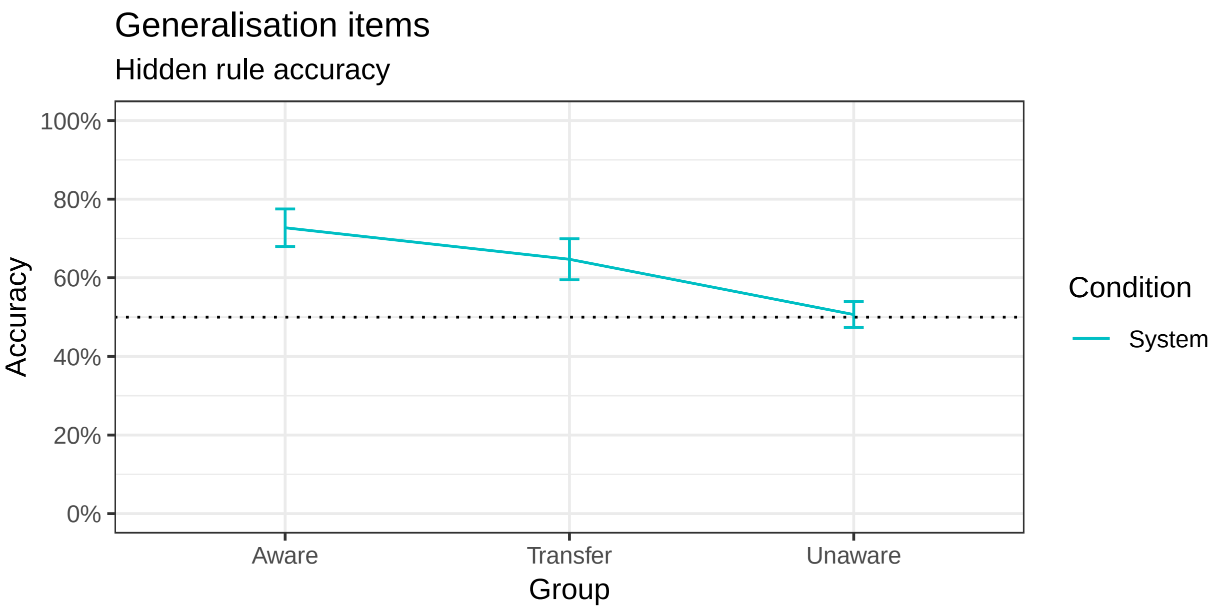


Figure 5.15: Average Hidden generalisation rule accuracy for Aware, Transfer and Unaware subjects. Error bars represent SE of the mean, dotted line marks 50% chance level.

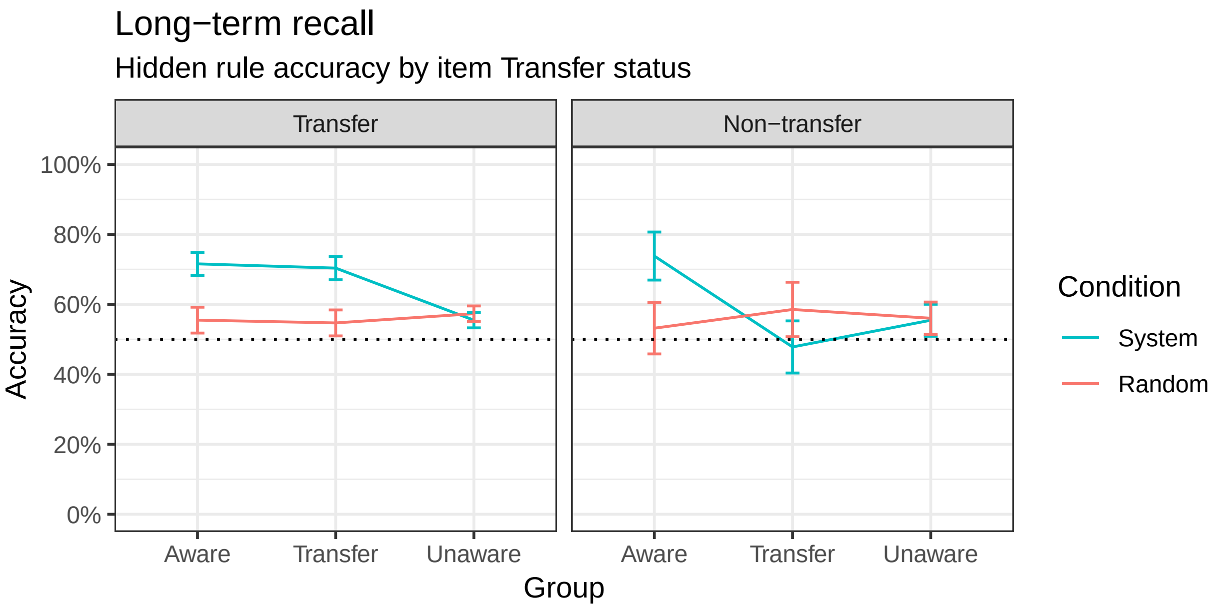


Figure 5.16: Average Hidden rule accuracy by Condition and Transfer status during Long-term recall for Aware, Transfer and Unaware subjects. Error bars represent SE of the mean, dotted line marks 50% chance level.

showed that there was a significant difference between Aware and Unaware participants in the System condition ($\chi^2(1) = 7.77$, $p = .032$), while the Transfer group was not significantly different from either the Aware ($\chi^2(1) = .95$, $p = 1$) or Unaware group ($\chi^2(1) = 2.70$, $p = .60$) (Fig. 5.14). For generalisation items, there was a main effect of Group ($\chi^2(1) = 10.5$, $p = .005$). Post-hoc comparisons showed that while the Aware and Unaware groups differed significantly from each other ($\chi^2(1) = 9.17$, $p = .007$), again the Transfer group did not significantly differ from either the Aware ($\chi^2(1) = 1$, $p = .95$) or Unaware group ($\chi^2(1) = 3.53$, $p = .18$).

Finally, we again looked at the effect of Transfer and Non-transfer items after breaking down the Aware group into the relevant subgroups (Figs. 5.16 & 5.17). For System item recall, we found a significant main interaction of Transfer with Group ($\chi^2(1) = 12.83$, $p = .002$). Post-hoc comparisons showed that the effect of Transfer was significant for the Transfer group only ($\chi^2(1) = 12.68$, $p = .001$). For generalisation items, we found an interaction between Group and Transfer ($\chi^2(1) = 6.22$, $p = .045$). Post-hoc comparisons with the Bonferroni correction

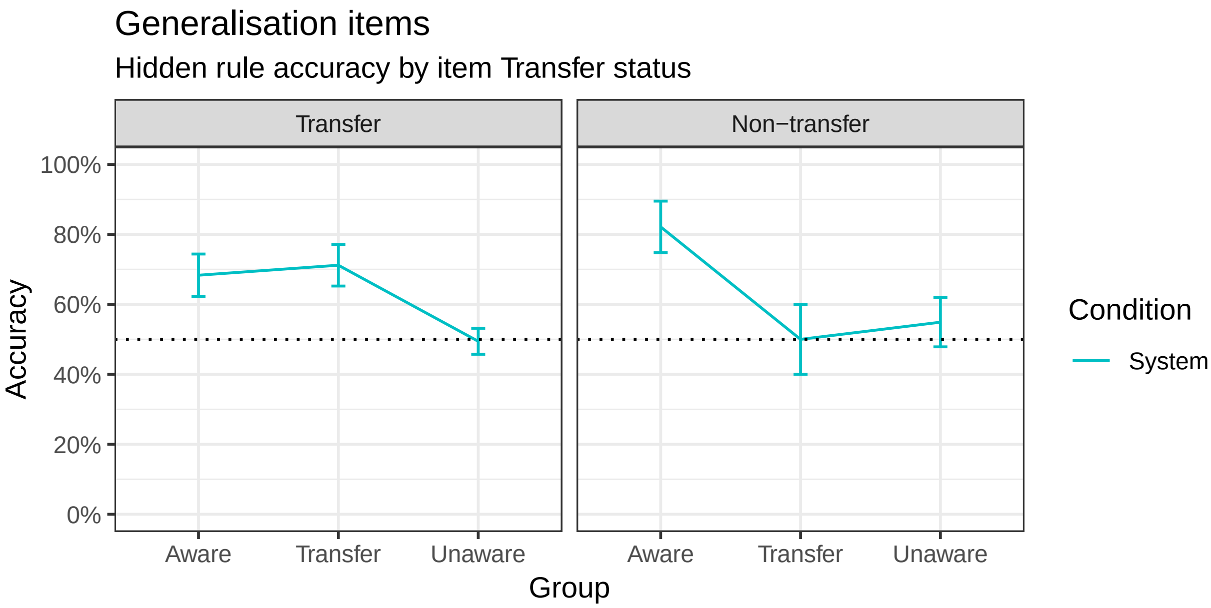


Figure 5.17: Average Hidden rule generalisation accuracy by Transfer status for Aware, Transfer and Unaware subjects. Error bars represent SE of the mean, dotted line marks 50% chance level.

failed to reach significance, possibly because of the small number of Non-transfer generalisation items. However, visual inspection of the data (5.17), shows a pattern similar to that we found in recall data (5.16): the Transfer group performed above chance - on par with the Aware group - on Transfer items, but was exactly at chance level on Non-transfer items, unlike the Aware group.

5.2.5 Discussion

5.2.5.1 Development of productive knowledge

If participants had developed productive knowledge of the Hidden rule, we expected recall accuracy to be higher for System items than for Random ones, and we also expected performance on Generalisation items to be above chance. Our predictions were only confirmed for the Aware group: Aware participants were significantly more accurate in their recall of System items, and above chance in their production of Generalisation items. By contrast, the Unaware group showed no difference between conditions in long-term recall and were at chance level in their production of Generalisation items. Likewise, the Aware group exhibited a significant correlation between System recall accuracy and generalisation, as previously observed in Experiment 2, but no such correlation was observed for the Unaware group. This suggests that participants in this experiment did not develop any implicit productive knowledge of the Hidden rule, unlike what we observed in Experiment 2, and that participants in the Aware group relied on explicit knowledge to do the task.

As done for Experiment 2, we analysed the effect of L1 transfer on accuracy in the production task. Again, there was no overall statistical effect L1 transfer. However, when we divided the Aware group into Aware and Transfer subgroups - based on the rule they verbalised in the questionnaire - we observed a different pattern from that of Experiment 2. Overall, performance by the Transfer group was intermediate between that of the Aware group and the Unaware group, both in recall and generalisation, and did not differ significantly from either. When we further broke down the data by item Transfer status, we saw that the Transfer group was the only that benefitted from L1 transfer: their performance on Non-transfer items was at chance level in both recall and generalisation. This suggests that they had developed no knowledge of the Hidden rule, and that their responses were entirely driven by explicit knowledge of the English *in/on* alternation.

5.2.5.2 Development of comprehension skills

When presented with items they had already encountered during training, both groups were more likely to endorse System items compared to Random ones, as already observed in Experiment 2. The fact that this was the case for the Unaware participants, too, who otherwise showed no evidence of having acquired the Hidden rule in this experiment, supports the hypothesis that the endorsement bias for System items was due at least in part to a general preference for these items, rather than to rule knowledge.

When presented with novel, grammatical items, there was an overall tendency for participants to endorse System items over Random ones; however, it would seem that this, too, was due to an endorsement bias for System items. Signal detection measures revealed no evidence of generalisation: when comparing *d'* scores for Seen vs. Generalisation items, there was no difference between conditions for either group. If participants had acquired receptive rule knowledge and were able to generalise it, we would have expected *d'* scores for Seen vs. Generalisation items to be lower in the System condition, since the presence of the rule in may make participants more likely to believe they had already encountered generalisation items, as previously observed in similar studies (Paciorek, 2012).

Finally, for Violation items (items from training, with reversed preposition assignment), we expected participants to be more sensitive to the preposition change in the case of System items, and therefore be less likely to endorse them. Again, direct comparison between condition was skewed by an endorsement bias for System items in both groups, which meant that we observed no difference between conditions in the Aware, and we actually found higher endorsement rates for Systems in the Unaware, despite grammatical violations. When comparing *d'* scores for Seen vs. Violation items, however, we found a significant difference between conditions for the Aware but not for the Unaware, indicating that only aware participants were sensitive to violation of the Hidden rule.

5.2.5.3 Comparison of Experiments 2 and 3

In these last two experiments, we investigated the development of productive generalisation of a novel form-meaning association. Specifically, we investigated whether the Hidden rule, which we had used during Experiment 1, could be acquired by participants through a recall-based paradigm and then used productively to generate novel phrases. Given the findings from Experiment 1, we modified both training and testing paradigm in order to promote the acquisition of implicit rule knowledge, and have better means of detecting it. In the training phase, we reduced the number of items to be retained for short-term recall from four to two, adding different kinds of memory load instead: pseudowords replacing a subset of place nouns (Experiment 2) or questions probing Overt rule knowledge and drawing attention to the physical characteristics of the place pictured (Experiment 3). In the testing phase, we increased the difficulty of the recall task, which now involved cued long-term recall of 88 items from training in random sequence, without any further exposure. This was done in order to minimise memory for specific items, in the hope that doing so would facilitate the emergence of abstract knowledge in productive behaviour. Likewise, we introduced generalisation items - novel picture-character combinations disguised as recall cues - in order to assess whether true productive knowledge had been developed.

In Experiment 2, we found a learning effect for both Aware and Unaware, which suggests that participants had acquired implicit knowledge of the Hidden rule and could use it productively. Specifically, we found a significant effect in both recall and generalisation for the Aware; among the Unaware, we found a significant effect in generalisation, with a tendency towards an effect in recall, too. In Experiment 3, we found no learning effect for the Unaware - either in recall or generalisation; we still found significant effects in both recall and generalisation among the Aware, albeit with lower overall accuracy rates compared to Experiment 2. The effect size (Cohen's d) for Aware participants in Experiment 2 was 0.89, which dropped to 0.49 in Experiment 3, showing a significant reduction in effect size. Similarly, the effect size for Unaware in Experiment 2 was 0.49, dropping to as low as -0.12 in Experiment 3.

The diminished effect of Condition in Experiment 3 was not due not only to lower performance on System items (Exp. 2: Aware 77%, Unaware 57%; Exp. 3: Aware 69%, Unaware 56%) but also to improved recall accuracy for Random items (Exp. 2: Aware 51%, Unaware 51%; Exp. 3: Aware 55%, Unaware 57%). Unlike Experiment 2, both groups in this study were significantly above chance in their Hidden rule recall accuracy for System items (Aware: $t(20) = 3.6$, $p = .003$, Unaware: $t(20) = 2.6$, $p = .02$), but also, at least in the case of the Unaware, for Random ones ($t(20) = 3.2$, $p = .004$), with a tendency in the same direction for the Aware group, too ($t(20) = 2.1$, $p = .06$). The higher recall accuracy for Random items, in particular, suggests that participants had better item memory than in Experiment 2.

In generalisation trials, too, we saw an overall decrease in Hidden rule accuracy in Experiment 3 relative to Experiment 2, which affected both groups. For the Unaware group, the effect disappeared entirely, and participants were at chance levels in their responses. Following Dienes (2014), we calculated the Bayes factor for Unaware accuracy in the Generalisation trials of Experiment 3, to check whether our failure to replicate the effect for the Unaware group should be taken as support for the null hypothesis (i.e. that participants really were at chance level), or simply as inconclusive. All factors were obtained using the calculator developed by Dienes (2008). As a rule, a Bayes factor (B) below 1/3 is considered to support the null hypotheses, $B > 3$ supports the theory, while a B between 1/3 and 3 is inconclusive (Dienes, 2014). We obtained a Bayes factor (B) of 0.29, which supports the null hypothesis - i.e. that the Hidden rule did not have any effect on performance by Unaware participants in Experiment 3 (by contrast, the Aware group, who performed above chance in Experiment 3 despite not being as accurate as in Experiment 2, had $B = 5.18$).

To test for the development of receptive knowledge, in Experiments 2 and 3 we used a recognition memory task instead of the listening task previously employed. In both experiments, participants from all groups showed evidence a general endorsement bias for System items, which meant that directly comparing endorsement rates for items in the System and Random condition did not provide an accurate measure of their receptive knowledge. When using signal detection measures to account for this bias, we found evidence of sensitivity to Hidden rule violation in the Aware in both Experiment 2 and 3, but not in the Unaware, which would seem to indicate that this violation detection was driven by explicit knowledge. In Experiment 3, we introduced generalisation items, but found no evidence of generalisation for either group: the presence of the Hidden rule in new items did not have any effect on familiarity judgements, which would seem to indicate a lack of rule generalisation. Overall, the results of the recognition memory task from these to experiment show little evidence of implicit knowledge being transferred from production to comprehension.

5.3 Conclusion

In Experiment 2, we found evidence of implicit learning of the Hidden rule, which resulted in productive generalisation of the rule even for participants who had not become aware of the rule. In Experiment 3, we did not replicate this finding. The lower overall accuracy and smaller effect size that we observed, for both groups, in this experiment suggests that the failure to replicate may be due to the different training paradigm employed. Bayes factor analysis suggests that the null result in Experiment 3 was not simply a failure to replicate Experiment 2, but that it provides support for a lack of effect (null hypothesis). Possible reasons for the difference will be explored in greater detail in Chapter 7. Since the two experiments differed in training paradigm but used the exact same materials for production training and testing, this confirms that the unconscious transfer of English categories was not primarily responsible for the advantage on System items in Experiment 2 (although it may have had a facilitating effect). It also suggests that the recall advantage found for System items in Experiment 2 could not simply be due to item salience. Findings from the recognition memory task, on the other hand, suggest that there was no transfer of implicit knowledge from production to comprehension.

Chapter 6

Experiment 4: The role of working memory and overnight consolidation

6.1 Introduction

The outcome of Experiment 2 suggests that it was possible for participants to develop implicit knowledge of the Hidden rule through elicited recall - regardless of whether they had become aware of its content or not - and to generalise it to new instances in spoken production. Our use of memorisation as a learning tool was partly based on the assumption underlying the Oral Imitation task, which is that linguistic material exceeding phonological working memory capacity will force a listener to rely on linguistic knowledge to successfully reproduce it. Following work by Radulescu and colleagues, we also hypothesised that our recall paradigm would promote the extraction of regularities by overloading participants' memory capacity (Radulescu, 2014; Radulescu et al., 2019). However, our previous experiments did not not test that hypothesis. Therefore, in order to explore the effects of working memory on the learning process, we decided to run a new experiment which would also include a test of phonological working memory.

Additionally, our previous experiments all consisted of a single session, which did not provide an opportunity for overnight memory consolidation. As pointed out by DeKeyser (2003), however, language learning studies of short duration are biased in favour of explicit knowledge. Adding an opportunity for delayed testing, while still resulting in a relatively short study, may help us tap into some of the processes involved in more prolonged, naturalistic implicit learning, such as sleep consolidation. It could help us to confirm whether the effects we observed in previous experiments were short-lived priming effects, or whether they constituted true learning of new form-meaning connections, which could become integrated in the knowledge system. In the latter case, we may expect to see an improvement in test performance after the opportunity for overnight consolidation, as has been observed in previous research (Tamminen et al., 2012, 2015).

To address both these points, we decided to run a fourth experiment which would use the same training paradigm as Experiment 2, with the addition of delayed generalisation testing to investigate overnight consolidation, while also incorporating a measure of working memory. The following sections (6.1.1 & 6.1.2) provide a brief overview of the roles of working memory and consolidation in implicit learning, while Section 6.2 reports the findings of Experiment 4.

6.1.1 The role of working memory in implicit L2 learning

Working memory (WM) is a type of short-term memory that allows us to temporarily store and manipulate information in our mind while performing complex cognitive tasks (Baddeley, 1992). In the multi-component model of WM developed by Baddeley and Hitch (1974) and its subsequent developments, WM includes separate storage for phonological and visuospatial information, as well as an executive function component which controls attention (Baddeley, 2000, 2015; Baddeley et al., 1998). Working memory, particularly its phonological component, is known to correlate with L1 proficiency (Baddeley, 2003; Kidd et al., 2018), and there is a growing body of research showing that it correlates with L2 proficiency, too (Miyake and Friedman, 1998; Van den Noort et al., 2006; Williams, 2011). Phonological WM has been shown to support acquisition of both L2 vocabulary (Baddeley et al., 1998; French, 2006; Williams and Lovatt, 2003) and grammar (French and O'Brien, 2008; Speciale et al., 2004).

The available experimental evidence suggests that the effect of WM on L2 learning may depend on the training paradigm used (Grey et al., 2015; Robinson, 2005; Tagarelli et al., 2011); however, the exact nature of this relationship is unclear. In a study on the acquisition of a semi-artificial language based on German syntax, Tagarelli et al. (2011) found that WM predicted learning outcomes under explicit, but not under incidental, conditions. Likewise, Grey et al. (2015) found no effect of WM under incidental learning conditions of a semi-artificial language combining English lexicon with Japanese syntax. Further support for a link between WM and explicit learning comes from research showing that WM correlates with the ability to use corrective feedback (Fyfe et al., 2015; Goo, 2012). On the other hand, Robinson (2005) found a correlation between WM (measured by reading span) and learning of Samoan under incidental learning conditions, where participants had to process sentences for meaning without any prior grammar instruction. The testing phase used by Robinson (2005) differed from Tagarelli et al. (2011) and Grey et al. (2015) in that it incorporated a production measure, which is particularly relevant to the present study. Specifically, Robinson (2005) found a positive correlation between WM and performance on a guided sentence production task (where participants were given words to arrange in the correct order to form correct sentences), as well as an aural GJT; by contrast, Tagarelli et al. (2011) used an aural GJT with judgement source attribution, while Grey et al. (2015) used a visual GJT and a picture-matching task. Robinson's (2005) findings are compatible with research suggesting that WM is particularly implicated in L2 speech production (Skehan, 2015). Since the present study is specifically concerned with examining production abilities as a consequence of implicit learning, we may expect WM to have an effect on our outcome measures, as it did in Robinson (2005). However, Robinson (2005) did not test learning under explicit conditions; given the evidence of a link between WM and explicit learning (Fyfe et al., 2015; Goo, 2012), we may also expect to see a correlation between WM and rates of Hidden rule awareness in our study, on the assumption that high WM capacity will afford participants greater resources for making and testing hypotheses on possible rules underlying the input. Finally, as previously mentioned, including a measure of phonological WM in our study will allow us to test the assumptions behind our recall paradigm, namely that it promotes rule learning by pushing memory capacity. If that is the case, we may expect to see a non-linear relationship between WM and rule accuracy: high-WM participants for whom the memory load is too small may benefit less from training than those with lower WM; but it also conceivable that those at the lowest end of the WM scale will not benefit as much, if the load becomes too great. Therefore, we may expect to find that there is an "optimal" level of WM which allows participants to draw the greatest benefit from the training task, while those with higher or lower WM capacity will show smaller learning effects.

6.1.2 Memory consolidation and generalisation

In addition to the role played by short-term memory, L2 learning is of course dependent on long-term memory. We know that sleep promotes the encoding of new representations in long-term memory, in a process known as consolidation (Burnham, 1903; Gais and Born, 2004; Winocur and Moscovitch, 2011); this occurs in a variety of situations, from learning word lists (Ficca et al., 2000) to the acquisition of motor skills (Brashers-Krug et al., 1996). It has also been suggested that consolidation during sleep is needed in order to integrate novel memories into existing knowledge without disrupting the latter, allowing for the gradual formation of abstract generalisations from individual episodic memories. Dual-mechanism models such as the Complementary Learning Systems model (McClelland et al., 1995; O'Reilly and Norman, 2002; Schapiro et al., 2017) hypothesise that this is carried out in the hippocampus and the neocortex, which play complementary roles in handling the transition from individual representations to more abstract general knowledge: the hippocampus is responsible for the fast encoding of new episodic memories, while the neocortex slowly integrates them into existing knowledge. This is required in order to avoid "catastrophic interference" (French, 1999; McClelland et al., 1995): if new representations were fed directly into the knowledge system, there would be a radical restructuring of the system each time any new memories are formed which do not conform to familiar patterns.

Generalisation of a rule to new instances, such as is required in our study, should therefore benefit from consolidation, since it involves a shift away from individual representations to more abstract knowledge. Indeed, in language learning, consolidation has been shown to support both the integration of new vocabulary into the lexicon (Davis and Gaskell, 2009; Dumay and Gaskell, 2007; Gaskell and Dumay, 2003; Mirković and Gaskell, 2016), the abstraction of phonological categories (Fenn et al., 2003) and phonological constraints in speech production (Gaskell et al., 2014), as well as the generalisation of novel morphology to new lexical items (Tamminen et al., 2012, 2015). In their study on the acquisition of novel derivational morphology, Tamminen et al. (2015) tested for the integration of novel affixes into the lexicon using an implicit measure (semantic priming in a sentence congruity task) and an explicit one (a sentence congruity judgement). In the implicit task, they found no evidence of integration immediately after training; however, when they tested participants again after a 1-week interval, they did find evidence of integration into the lexicon (provided that the training phase contained enough unique exemplars), suggesting that consolidation was necessary for it to emerge. Conversely, generalisation on the explicit measure could already be observed immediately after training. Further evidence of the role played by consolidation comes from a study by Morgan-Short et al. (2012a), which explored the long-term retention of an artificial language learned under either explicit or incidental conditions. Having previously shown that only incidental training led to native-like processing, as evidenced by ERP responses to grammatical violations (Morgan-Short et al., 2012b), Morgan-Short and colleagues found the same native-like responses even after months of no exposure. Strikingly, however, they also found that both groups showed more native-like processing than they had in the original study, suggesting that the time lapse had had a beneficial effect for both the explicit and implicit training group. Taken together, these findings suggest that consolidation favours the development of generalisation and automatic knowledge in L2. Therefore, we may expect to see an improvement in generalisation performance after an opportunity for overnight consolidation.

6.2 Experiment 4

The aim of this experiment was to explore the contribution of working memory to the learning process, as well as testing the effect of overnight consolidation on implicit knowledge acquired during the training phase. In order to obtain a larger sample of participants, data collection was carried out online instead of in a laboratory. This is an increasingly common method of carrying out behavioural research and has been shown to be provide generally reliable data (Clifford and Jerit, 2014; Crump et al., 2013), including for implicit learning studies (Kerz et al., 2017). Based on findings from Experiments 2 and 3, we reverted to the training and testing paradigm used in Experiment 2, which had proved successful at generating implicit rule knowledge, unlike the one we used in Experiment 3. Doing so would also allow us to test our hypothesis with regards to the different outcomes of Experiment 2 and Experiment 3, which we attributed to the change in training methodology. A replication of results from Experiment 2 would provide validation for the paradigm used. We therefore used the same training and testing methodology as Experiment 2, with the addition of a second generalisation task, carried out one day after training, and a measure of WM (a forward digit span task). In order to avoid making the experiment too long, given the introduction of additional consolidation and WM tasks, we chose not to include a comprehension component in this experiment, focusing on production only.

If participants had acquired implicit knowledge of the Hidden rule (and the training paradigm from Experiment 2 was indeed better suited for generating implicit knowledge), we would expect to see higher recall rates for System items and above-chance performance on generalisation items, as in previous experiments. We would also expect performance on the generalisation task to improve in the second testing session, as a consequence of overnight consolidation. With regards to the role of WM, several different predictions could be made, based on the existing literature. Following Robinson (2005), we may expect there to be a positive linear relationship between WM and rule accuracy, with high-WM subjects performing better than low-WM ones. On the other hand, if our recall paradigm worked by loading working memory, we may expect to see an inverted U-shape relationship between WM and Hidden rule accuracy: subjects whose WM span is optimally suited to the task load should show the greatest learning effect, while those at the opposite ends of the range, for whom the load is either too small or too large, should show a smaller effect. Finally, given the evidence of a link between WM and explicit learning (Fyfe et al., 2015; Goo, 2012), we may expect WM to correlate with discovery of the Hidden rule, as high-WM subjects would have more resources available for hypothesis testing and rule search.

6.2.1 Participants

We tested 42 native English speakers (25 females, mean age = 30.9) recruited through online platform Prolific (www.prolific.ac). Participants were from the United Kingdom (N = 29), USA (N = 7), Australia (N = 3), New Zealand (N = 2) and Canada (N = 1). Knowledge of foreign languages was limited compared to previous samples; only one participant reported being fluent in a language other than English (Spanish).

6.2.2 Materials

We used the same materials previously used for Experiments 2 and 3: 80 unique place nouns, from which we derived a total of 160 unique items. Of these, 112 were used as training items (64 matched + 48 unmatched), while the remaining 48 served as generalisation items.

6.2.3 Procedure

Unlike previous experiment, the experiment was not carried out in a laboratory but was instead deployed online, recruiting participants through online platform Prolific (www.prolific.ac). We recreated the design of previous experiments using JavaScript library jsPsych (De Leeuw, 2015), which allowed participants to run the experiment in a web browser on their own computers. The experiment was composed of a training phase followed by a production task which included long-term recall and two blocks of generalisation items, one on the day of training (Day 1) and one after a 24-hour interval (Day 2). No comprehension tasks were included in the experiment.

6.2.3.1 Training phase

The training phase (Fig. 6.1) followed the same procedure used in Experiment 2. Participants were exposed to 112 sentences: of these, 64 were the same matched

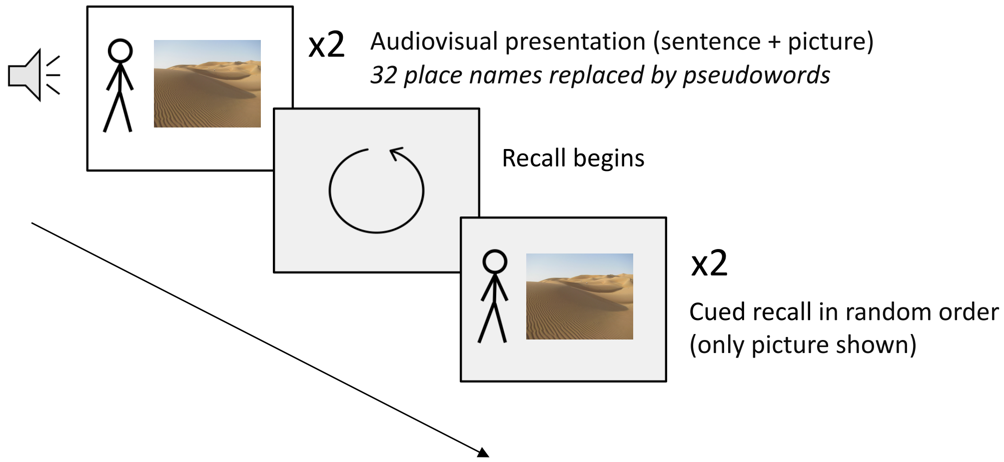


Figure 6.1: Training phase procedure for Experiment 4

items used in previous experiments (32 unique nouns appearing once in each condition), while the remaining 48 were unmatched items, nouns appearing only in one condition during training (counterbalanced across participants). Nouns in half of the matched items (32 nouns) were replaced by pseudowrds. Recall was every two items.

6.2.3.2 Testing phase

On Day 1, the testing phase (Fig. 6.2) followed the same procedure as Experiment 3: a long-term recall task in which participants had to recall the 64 matched items using the corresponding pictures as cue, followed by a block of 24 generalisation items (in this respect, we deviated from Experiment 2 by presenting recall and generalisation items in sequence rather than intermixing them, because previous experiments had revealed that it was not necessary to disguise generalisation items to make ensure that participants would not become aware of the manipulation). On Day 2, a second set of 24 generalisation items was administered, using the same procedure (Day 1 and Day 2 generalisation items were counterbalanced across participants).

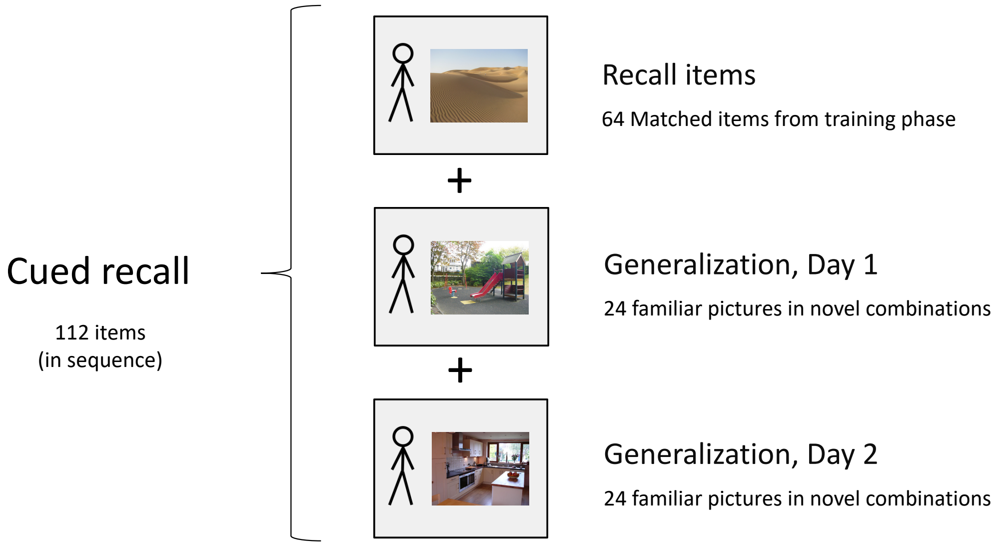


Figure 6.2: Production task procedure for Experiment 4

6.2.3.3 Digit Span task

At the end of Day 2, we administered a forward Digit Span task to measure phonological working memory. Participants were exposed to digit sequences of increasing length, starting from two digits. Digits were flashed in rapid succession on screen: each digit was displayed for 800ms, with a 200ms interval between digits. After the digit sequence, a digital number pad appeared on screen, and participants were required to enter the sequence they had just seen by clicking on the appropriate keys. They had the option to erase entered digits if they made any mistakes. Participants did five trials for each sequence length, before moving on to the next level. If they got more three or more trials wrong on any given level, the task was terminated.

6.2.3.4 Debriefing questionnaire

At the end of the experiment, awareness of the Overt and Hidden rules was assessed using a debriefing questionnaire (Appendix A), as done in previous experiments.

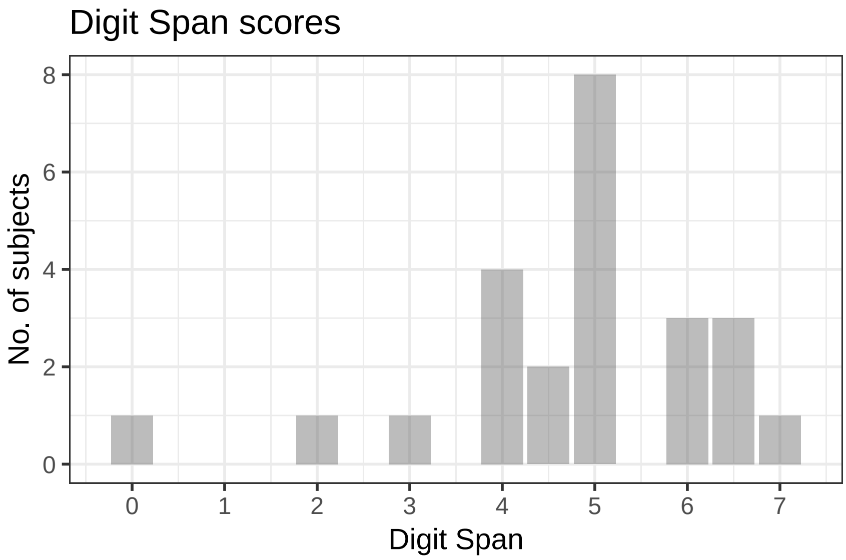


Figure 6.3: Distribution of digit span scores for Experiment 4 participants.

6.2.4 Results

6.2.4.1 Task scoring

Production data was transcribed and scored following the same procedure as Experiments 1-3. The digit span task was scored as follows: each level with 3 correct trials out of 5 counted as 0.5 points, while levels with 4 or 5 correct counted as 1 point.

6.2.4.2 Rule awareness

We assessed awareness of the Overt and Hidden rule based on the debriefing questionnaire (Appendix A), following the scoring procedure already used for Experiments 2 and 3.

6.2.4.3 Analysis

Due to variation in the recording devices used by participants, audio quality was not consistently high. Three subjects were excluded from the analysis due to low overall audio recording quality; a further four were excluded for failing to complete all parts of the experiment. Finally, 11 subjects failed to become aware of the Overt rule and were therefore not included in the analysis, following the procedure from previous experiments. A total of 24 subjects were included in the analysis (16 females, mean age 28 years, SD = 6.53). Trials from remaining participants that were unintelligible due to poor audio quality (3% of total data) were eliminated.

Awareness rates for both Overt and Hidden rule were lower than in previous experiments: as noted above, 11 subjects failed to become aware of either rule and were excluded (for comparison, the corresponding figure was 6 subjects in Experiment 2, and only 1 in Experiment 3). Furthermore, of those included in the analysis, only 4 subjects became aware of the Hidden rule. No subjects were classed as Transfer.

Despite the very small number of participants in the Aware group (N = 4) they were entered in the analysis as a separate group alongside the Unaware group (N = 20). Since the analysis was carried out on a by-trial basis, this yielded a sufficient number of observations even for the Unaware group (255 obs. in Long-term recall) given an analysis with Group, Condition and Digit Span as independent variables.

We analysed production accuracy data using generalized mixed-effect models (GLMERs) for logistic regression, as done in previous experiments. We no longer included Block as a factor in these models, as it did not have any effect on accuracy in either Experiment 2 or 3. Where fitting random effects was not possible due to low variability (Long-term recall phase) we ran a binomial logistic regression with Condition and Digit Span as predictors. The distribution of Digit Span scores in the data was not normal (Fig. 6.3); therefore, logistic regression was a fitting choice of statistical analysis for our data, since it does not assume normality for independent variables.

6.2.4.4 Training phase

While there was some variation in Overt rule accuracy (Aware: System 89%, Random 97%, Unaware: System 87%, Random 85%), Hidden rule accuracy during the training phase was almost at ceiling for both groups (Aware: System 98%, Random 97%, Unaware: System 96%, Random 96%) (Fig. 6.4). We found no

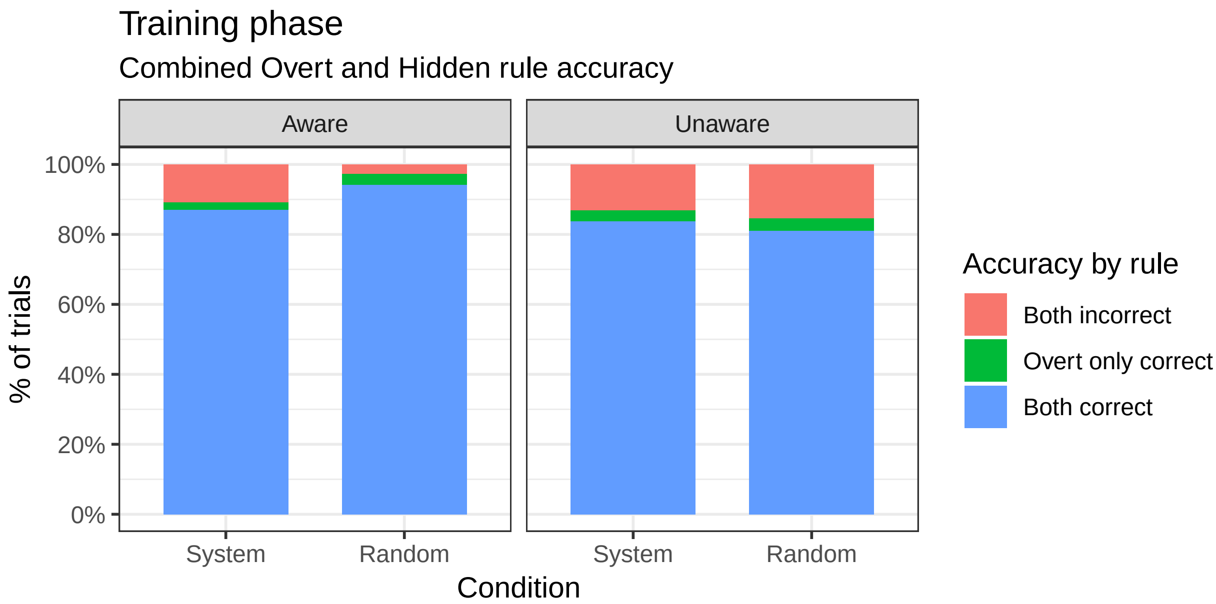


Figure 6.4: Combined Overt and Hidden rule accuracy by Group and Condition during the training phase.

significant effects of either Group, Condition or Digit Span in a GLMER with random intercepts for subjects and items, random slopes for subjects by block and for items by condition.

6.2.4.5 Testing phase

Long-term recall In long-term recall, Overt rule accuracy was lower than in the training phase, particularly for the Unaware (Aware: System 76%, Random 83%; Unaware: System 50%, Random 45%) (Fig. 6.5). Hidden Rule accuracy was also lower among the Unaware group (Aware: System 71%, Random 47%, Unaware: System 57%, Random 49%) (Fig. 6.6). To analyse Hidden rule accuracy rates, we initially constructed a GLMER with Group, Condition and Digit Span as predictors and random intercepts for subjects and items; however, we had to eliminate all random effects as each of the specified models had singular fit, due to low variability among subjects and items. We therefore ran a binomial logistic regression with Group, Condition and Digit Span as predictors. There was no effect of Digit Span on accuracy, but there was a significant main effect of Condition ($\chi^2(1) = 11.384$, $p = .001$) and an interaction between Group and

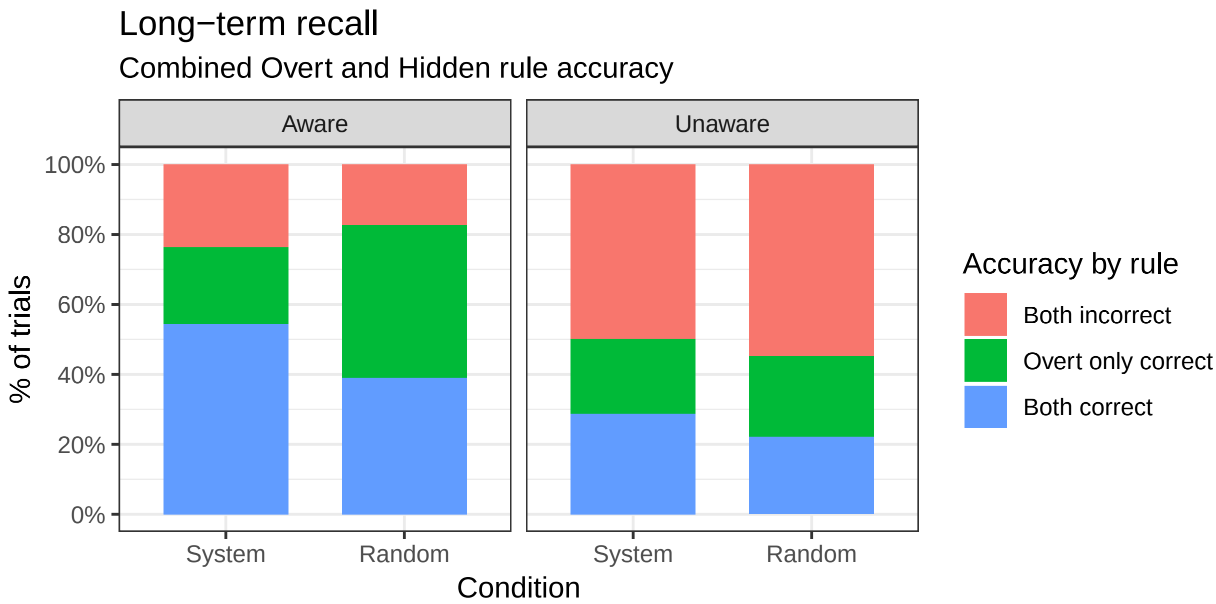


Figure 6.5: Combined Overt and Hidden rule accuracy by Group and Condition during Long-term recall.

Condition ($\chi^2(1) = 4.772$, $p = .029$) (Table 6.1). Post-hoc comparisons showed that the effect of Condition was highly significant for the Aware group ($\chi^2(1) = 14.030$, $p < .001$) but only approached statistical significance for the Unaware ($\chi^2(1) = 3.775$, $p = .052$).

Generalisation Generalisation items were novel System items derived from pictures previously seen in training as part of Random items. Hidden rule accuracy scores increased from Day 1 (Aware: 65%, Unaware: 46%) to Day 2 (Aware: 72%, Unaware: 56%)(Fig. 6.7). We entered Hidden accuracy scores for both days in a GLMER with Group, Day and Digit Span as fixed factors and random intercepts for subjects, which showed a significant main effect of Group (Table 6.2). As in previous tasks, there was no correlation between Digit Span and accuracy. While the overall increase from Day 1 to Day 2 was not significant, a one-sample t-test against chance level (50%) showed that the Unaware group was significantly above chance on Day 2 ($t(16)$=2.53, p=.02), despite being below chance on Day 1 ($t(16)$=-2.134, p= .05). Due to its small sample size (n = 4), the Aware group fails to meet statistical significance; however, their average accuracy on Day 2

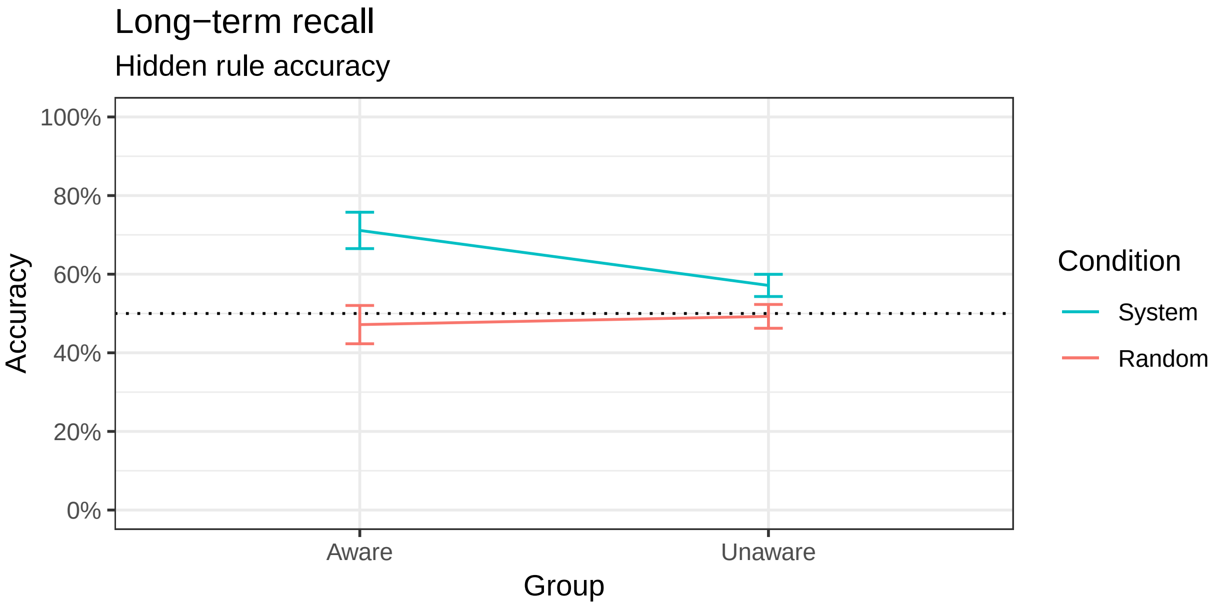


Figure 6.6: Average Hidden rule accuracy by Group and Condition during Long-term recall. Error bars represent SE of the mean, dotted line marks 50% chance level.

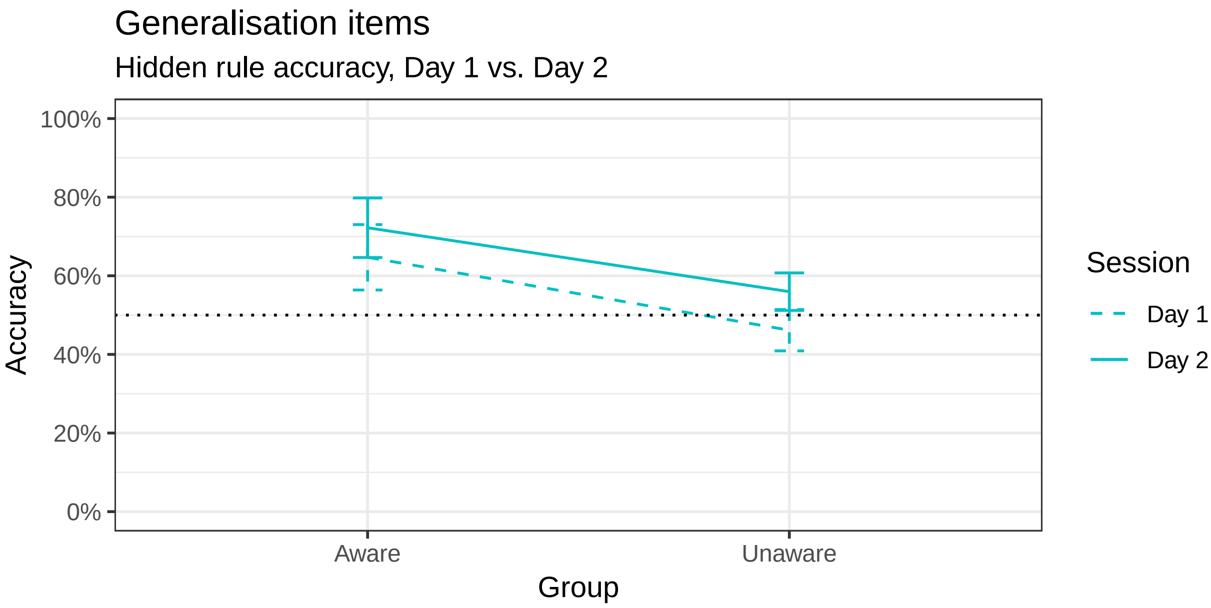


Figure 6.7: Average Hidden rule accuracy by Group for Day 1 and Day 2 Generalisation items. Error bars represent SE of the mean, dotted line marks 50% chance level.

Hidden rule accuracy, Long-term recall

	Coefficient	SE	χ^2	df	p	
Group	0.039	0.021	3.244	1	0.072	
Condition	0.088	0.021	11.384	1	0.001	***
Digit Span	0.025	0.017	2.709	1	0.100	
Group x Condition	0.047	0.021	4.772	1	0.029	*
Group x Digit Span	0.015	0.017	1.004	1	0.316	
Condition x Digit Span	0.014	0.017	1.525	1	0.217	
Group x Condition x Digit Span	0.023	0.017	1.837	1	0.175	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 6.1: Summary of GLMER model for Hidden rule accuracy, Long-term recall.

(72%) is comparable to that of Aware participants in Experiment 2 (74%). For Unaware participants, there was no correlation between generalisation accuracy (Day 1 or Day 2) and recall accuracy for System items (Day 1: Pearson's R = -0.06, p = .8, Day 2: Pearson's R = .11, p = .67). Correlations for the Aware group were not statistically significant on account of its small sample size, although they showed the same trend observed in Experiment 2 (Day 1: Pearson's R = .79, p = .20, Day 2: Pearson's R = .73, p = .26).

6.2.5 Discussion

In this experiment, we further investigated the development of productive implicit knowledge of the Hidden rule, using the training paradigm we already adopted in Experiment 2. As in previous experiments, we predicted that, if participants had acquired productive knowledge of the rule, we should see an advantage for System items in long-term recall, as well as above-chance accuracy when producing novel (generalisation) System items. In addition to this, we added a measure of working memory, to test our assumptions that the training paradigm promoted rule learning by loading participants' memory capacity. Finally, we tested participants on a second set of generalisation items after the opportunity for overnight consolidation, predicting that they should display higher accuracy compared to immediate generalisation testing. Unlike our previous experiments, Experiment 4

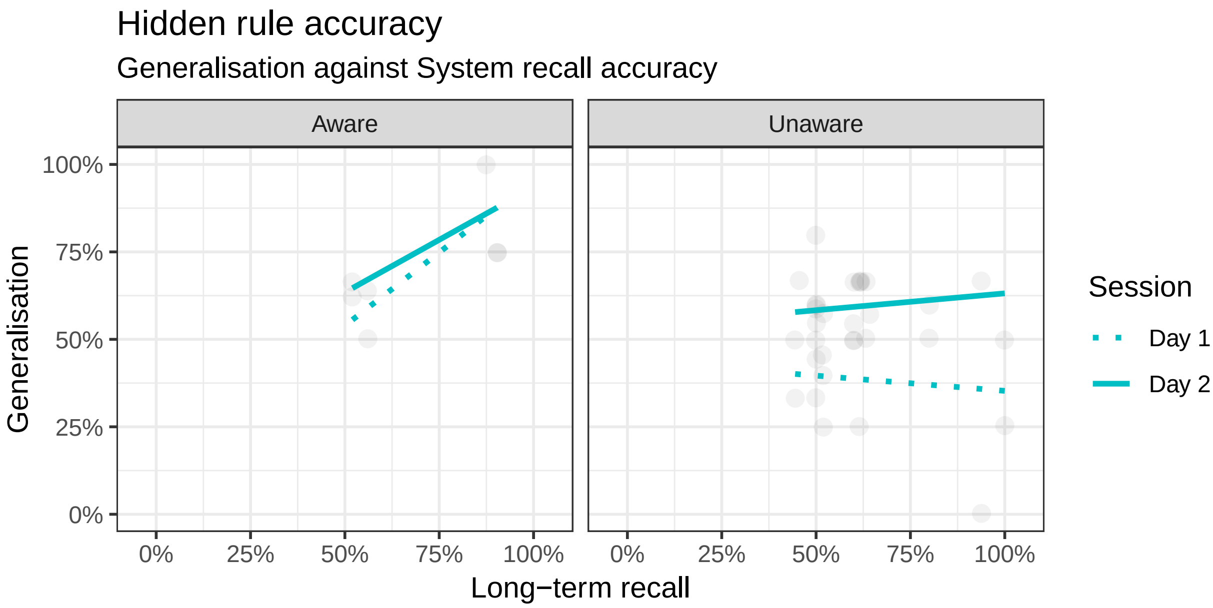


Figure 6.8: Average Hidden rule accuracy during Generalisation plotted against average accuracy for System items during Long-term recall task.

was not laboratory-based: we instead carried out data collection on the internet, recruiting participants through an online research platform.

With respect to productive knowledge in long-term recall, the pattern of results we observed was very similar to that of Experiment 2. The Aware group showed a significant difference between System and Random items, with higher recall accuracy for System ones. Among Unaware participants, there was a visible trend towards an effect of Condition, but it failed to reach statistical significance. One of our aims when switching to online data collection had been to collect a larger participant sample, in order to have more statistical power than Experiment 2; we hoped that this would help to shed more light on the nature of the trend we observed for Unaware participants. However, even though the initial sample was bigger, Experiment 4 suffered from higher attrition rates than previous experiments due to both technical issues and lower levels of compliance from participants, meaning that the resulting sample of Unaware subjects (n = 20) was comparable to those of previous experiments (Experiment 2: 17 Unaware, Experiment 3: 21 Unaware). Therefore, it is not possible to determine whether the trend we see reflects an underlying difference which would be statistically

Hidden rule accuracy, Generalisation

	Coefficient	SE	χ^2	df	p	
Group	0.714	0.252	7.738	1	0.005	**
Day	-0.170	0.130	2.400	1	0.121	
Digit Span	0.099	0.191	0.213	1	0.644	
Group x Day	-0.008	0.130	0.000	1	0.987	
Group x Digit Span	0.103	0.191	0.305	1	0.580	
Day x Digit Span	0.048	0.101	0.212	1	0.645	
Group x Day x Digit Span	-0.030	0.101	0.089	1	0.766	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 6.2: Summary of GLMER model for Hidden rule accuracy, Generalisation items.

significant, given a larger sample.

In generalisation trials, the pattern of results we observed was partially different from that of Experiment 2. In this experiment, we included two blocks of generalisation items, administering the first one immediately after recall testing (Day 1) and the second one after a 24-hour delay (Day 2). On Day 1, only the Aware performed above chance, while the Unaware were at chance level in their production of generalisation items. This differed from what we observed in Experiment 2, where both groups were above chance in generalisation immediately after training. On Day 2, however, we saw an improvement in performance for both Aware and Unaware: even though the difference between Day 1 and 2 was not statistically significant, which means our predictions on consolidation were only partially supported, the increase meant that both groups were now performing above chance. Finally, we did not find a significant correlation between System recall and generalisation accuracy in the Unaware group, even on Day 2, which again deviates from what we observed in Experiment 2. Nevertheless, while it does not fully replicate the outcome of Experiment 2, this pattern of results indicates that participants had acquired implicit knowledge of the Hidden rule, and could use it productively.

With regards to the role of working memory in the learning process, our predictions were not confirmed. We had expected to see a relation between working memory and Hidden rule accuracy, either in the shape of a positive linear correlation between the two, or as a U-shaped relationship, on the assumption that the recall paradigm worked by loading participants' memory capacity to an optimal level. However, we did not find any correlation between the two. Our prediction of a connection between working memory and awareness of the Hidden rule could not be confirmed either: due the extremely small number of participants who became aware of the rule (n = 4), we could not detect any statistical relationship between working memory and Hidden rule awareness. In order to fully explore the possible effects of working memory on the learning process, we tested for an effect of WM on Overt rule accuracy, too; we did not find any effect of working memory on Overt rule accuracy in either the training phase ($\chi^2(1) = 2.637$, $p = .10$) or the long-term recall task ($\chi^2(1) = 0.876$, $p = .35$). The lack of correlation between WM and recall accuracy which we observed could be due a variety of factors: it is possible that the specific measure of WM we used (forward digit span task) could be less sensitive to linguistic processes than a reading span task, which was the measure of WM used in Robinson (2005). It is possible that WM as a construct is not particularly relevant to implicit learning, and that focusing on the distinction between procedural and declarative memory instead (Antoniou et al., 2016; Ettlinger et al., 2014; Morgan-Short et al., 2014) may have yielded a more informative picture. These possibilities will be explored in greater detail in the general discussion (Chapter 8).

Finally, in the the production phase of this experiment we also observed very low rates of Overt rule accuracy in the Unaware group (50% for System items, 45% for Random items), relative to previous experiments. Even though all participants in the experiment were aware of the Overt rule, it appears that Unaware subjects in this experiments were not able to reliably used this rule in spoken production. Since the Hidden rule has been defined as a "sub-rule" of the Overt rule, this would seem to pose a problem when claiming that participants had acquired the Hidden rule. However, chance-level performance on the Overt rule is not incompatible with knowledge of the Hidden rule. In its essence, the Hidden rule is an association between particular prepositions and particular place types. In the context of this experiment, the Hidden rule should only be applied, and the relevant prepositions used, when the character was in a particular position on screen, as dictated by the Overt rule. Therefore, failure to use the Overt rule correctly did not imply that the Hidden rule association had not been acquired: it only indicates that subjects could not reliably determine when they should apply the Hidden rule (i.e. when the character was in one position on on screen, and not in the other).

6.3 Conclusion

In Experiment 4, we reverted to the training paradigm used in Experiment 2, with the addition of further generalisation testing one day after training. We found evidence of implicit learning of the Hidden rule and productive use of the rule by both Aware and Unaware participants, with above-chance performance in generalisation tasks in both groups, and an effect of condition on recall in the Aware. The experiment mostly replicated the findings of Experiment 2, with one notable difference: the effect only emerged after overnight consolidation in the Unaware group, who were above chance on the second day of generalisation testing only. Nevertheless, the similarities between Experiment 2 and 4, despite the fact that the two studies were delivered in different settings and were sampling from different populations, suggest that the training paradigm had a significant impact on learning outcomes, as we had hypothesised. This issue will be explored in greater detail in the next chapter, which carries out a systematic comparison of Experiments 2, 3 and 4. Finally, in this experiment we also included measures of working memory, in order to test our assumptions with regards to the role of memory in our training paradigm. However, we did not find any effect of digit span on accuracy, either for Hidden or Overt rule. Possible reasons for this lack of correlation will be explored in the general discussion (Chapter 8).

Chapter 7

Effects of training paradigm

7.1 Introduction

The last three experiments we conducted all produced evidence of rule learning: however, only in two of them (Experiments 2 and 4) did participants acquire implicit, productive knowledge of the Hidden rule. Experiment 3 failed to replicate the results of Experiment 2 with respect to Unaware participants, despite being similar in design, using the same materials and recruiting participants from the same community. By contrast, Experiment 4, which sampled participants from a very different community and was conducted on the web instead of in a laboratory, used the exact same training paradigm as Experiment 2 and partially replicated its findings. Therefore, we hypothesized that the experimental paradigm used - specifically, the training procedure used - may have been responsible for the differences we observed between Experiments 2 and 3. In the following sections, we will compare the two versions of the training procedure which we used in the last three experiments and examine the differences between them, in order to establish whether they may have affected learning outcomes.

7.2 Comparison of Experiments 2, 3 and 4

7.2.1 Procedure

Experiments 2, 3 and 4 all used one of two variations of the same training paradigm, which involved presentation of two items in sequence, followed by short-term recall. Experiments 2 and 4 used the same procedure, with recall every two items and pseudowords replacing half of place nouns in matched items, for a total of 32 nouns (Fig. 4.1, p. 86). In Experiment 3, recall was also every two items but all place nouns were in English, and participants had to answer questions after each item presentation (Fig. 5.1, p. 108).

All participants recruited for our experiments were native English speakers, but they differed in other respects. Participants in Experiments 2 and 3 were recruited from the the same pool (University of Cambridge students and surrounding community) and were broadly similar in terms of language background, educational level and age (Exp. 2: mean age 20.5, Exp. 3: mean age 20.7). Participants for Experiment 4 where recruited online through platform Prolific Academic, mostly from the United Kingdom; only 14 out of 42 subjects in the experiment reported being in education (either full- or part-time) at the time of the study. Most were either in full-time (N = 14) or part-time (N = 14) employment; a further 10 subjects (3 of which were students) were not in employment, while 4 did not report their employment status. Participants in Experiment 4 were also significantly older than the sample used in previous experiments (mean age 28.2) (Table 7.1).

Design summary, Experiments 2 - 4

	Setting	Mean age	Design feature
Exp. 2	Lab-based	20.5 (SD = 2.09)	Pseudowords
Exp. 3	Lab-based	20.7 (SD = 4.98)	Questions
Exp. 4	Online	28.2 (SD = 6.53)	Pseudowords

Table 7.1: Summary table of Experiments 2, 3 and 4.

7.2.2 Results

In Experiment 2, we found a significant effect of Condition on Hidden rule recall for the Aware, with higher accuracy for System items, and a trend in the same direction for the Unaware. Additionally, both groups in this experiment were significantly above chance in generalisation. In Experiment 3, the effect of Condition on Hidden recall shrank for both groups: compared to Experiment 2, recall accuracy decreased for System items, but increased for Random ones. Only the Aware group in this experiment was significantly above chance in generalisation. In Experiment 4, we found the same pattern of long-term recall that we had observed in Experiment 2, with a significant effect for the Aware group, and a trend in the same direction for the Unaware. In generalisation, Aware participants were above chance on both Day 1 and Day 2; the Unaware were above chance on Day 2 only (Table 7.2).

Hidden rule accuracy, Experiments 2 - 4

		Long-term recall			
	Group	System	Random	Cohen's d	Generalisation
Exp. 2	Aware (n = 16)	77%	51%	0.89	74%
	Unaware (n = 17)	57%	51%	0.49	61%
Exp. 3	Aware (n = 16)	69%	55%	0.49	69%
	Unaware (n = 21)	56%	57%	-0.12	51%
Exp. 4	Aware (n = 4)	71%	47%	0.89	t_1: 65%, t_2: 72%
	Unaware (n = 20)	57%	49%	0.59	t_1: 46%, t_2: 56%

Table 7.2: Hidden rule accuracy during production tasks by group and experiment, Experiments 2 to 4.

We see similarities between the findings of Experiment 2 and 4, despite different demographics and modes of delivery, which set them apart from Experiment 3. Experiment 3 differs from the other two studies in several ways. The effect size for Condition in Long-term recall is smaller than in the other two experiments: this is due to both lower accuracy on System items, both in recall and generalisation, and higher recall accuracy for Random ones. Overall, we found no evidence of implicit learning of the Hidden rule in Experiment 3, unlike what we observed the other two experiments.

There are two ways in which the training paradigm used could have contributed to creating this pattern of results. Firstly, the combination of higher recall accuracy for Random items, and lower accuracy for System ones, suggests that the training procedure used in Experiment 3 may have encouraged retention of memories for individual items, while impairing rule generalisation. The paradigm used in Experiment 3 differed from the other version in that it lacked pseudowords: all place nouns were in English. We therefore hypothesised that the use of pseudowords in Experiments 2 and 4 may have made it harder for subjects to retain individual items, thus encouraging rule abstraction. We conducted a post-hoc analysis of the data from Experiments 2 and 4 to assess the effect of pseudowords in these studies, in order to determine whether the lack of pseudowords could explain the outcome of Experiment 3. The results of the analysis are presented in Section 7.3.

Secondly, we hypothesised that differences in learning outcomes may have been due to different attentional demands created by the two training paradigms. In Experiment 3, participants were asked questions which drew their attention to the physical characteristics of the places pictured by asking them to rate their attractiveness, and which probed their knowledge of the Overt rule. Given the crucial role played by attention in implicit learning, it is possible that these questions may in fact have been detrimental to the acquisition of the Hidden rule, by directing attention away from the form-meaning connections underlying it. Specifically, questions asking participants to rate the places pictured may have drawn their attention to the places themselves, rather than their association with specific prepositions. Similarly, questions probing Overt rule knowledge could have been directing participants' attention to the Overt rule at the expense of the Hidden rule. To test these hypotheses, we first conducted a post-hoc analysis of data from Experiment 3, to determine whether the rating given to an item during training had any impact on recall accuracy for that item during testing (Section 7.4.1). We then conducted a further analysis to test whether there was a trade-off in our studies between Overt and Hidden rule learning, by examining the effect of Overt rule accuracy in training across all three experiments (Section

7.4.2).

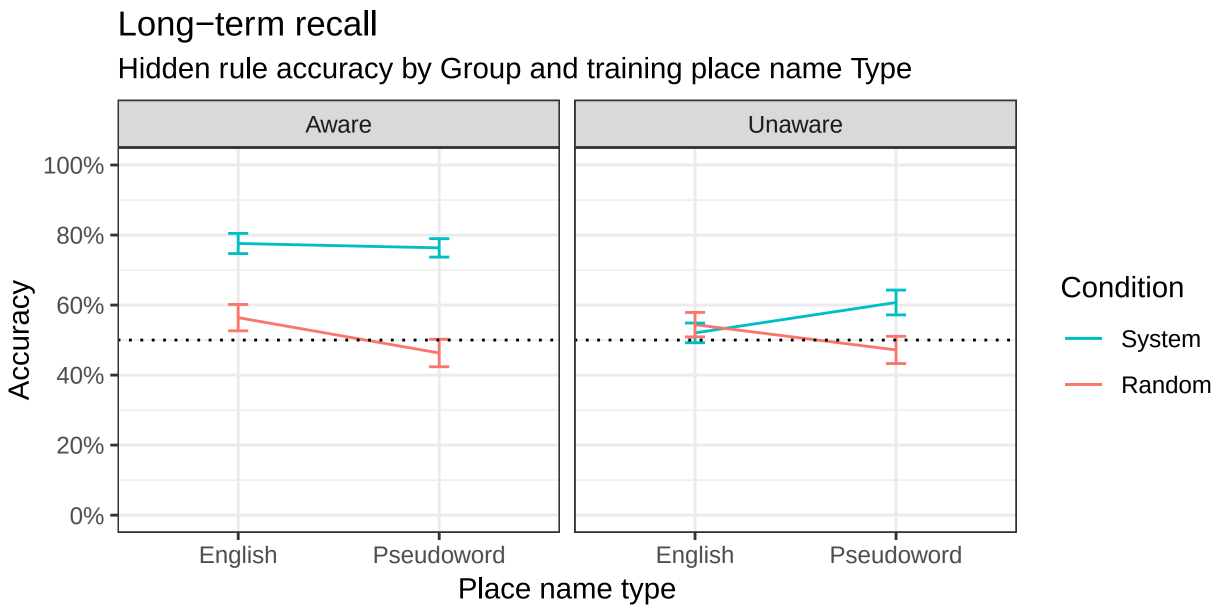
7.3 Effect of pseudowords

In order to determine whether the use of pseudowords in Experiments 2 and 4 had impaired the formation of specific item memories, we conducted a post-hoc analysis of Hidden rule accuracy scores during Long-term recall, based on whether the items had been presented with an English place noun or a pseudoword during training.

In Experiment 2, we discovered that the use of pseudowords during training had an effect on Hidden rule recall accuracy for those items during Long-term recall (Fig. 7.1). The effect was modulated by condition: there was a significant interaction between place name Type (English vs. Pseudoword) and Condition ($\chi^2(1)$ = 5.22, p = .02) in a GLMER with Condition, Group and place name Type as fixed effects and random intercepts for subjects and items. Post-hoc comparisons showed that if a Random item had its place noun replaced by a pseudoword during training, subjects were significantly less likely to recall it correctly in the Long-term recall task ($t(63)$ = 2.1, p = .04). For System items, we observed the opposite trend - participants were on average more accurate when recalling System items presented with pseudowords, although the difference fell short of statistical significance ($t(63)$ = -2, p = .06).

When performing the same analysis on Experiment 4 data, we did not initially observe the same pattern: averaging across participants by group, we found no effect of place name Type on Hidden rule accuracy in Long-term recall (Fig. 7.1). However, a breakdown of the data at the subject level, with Digit Span score included as factor, revealed an interaction between Digit Span and place name Type (Fig. 7.2). We built a GLMER with Group, Condition, place name Type and Digit Span as predictors, and random intercepts for subjects together with correlated random slopes for subjects by condition. The model revealed a significant 4-way interaction between Group, Condition, Type and Digit Span ($\chi^2(1)$ = 4.355, p = .037) as well as main effects of Group ($\chi^2(1)$ = 10.135, p = .001), Condition ($\chi^2(1)$ = 6.182, p = .013) and place name Type ($\chi^2(1)$ = 8.849, p = .003).

Experiment 2



Experiment 4

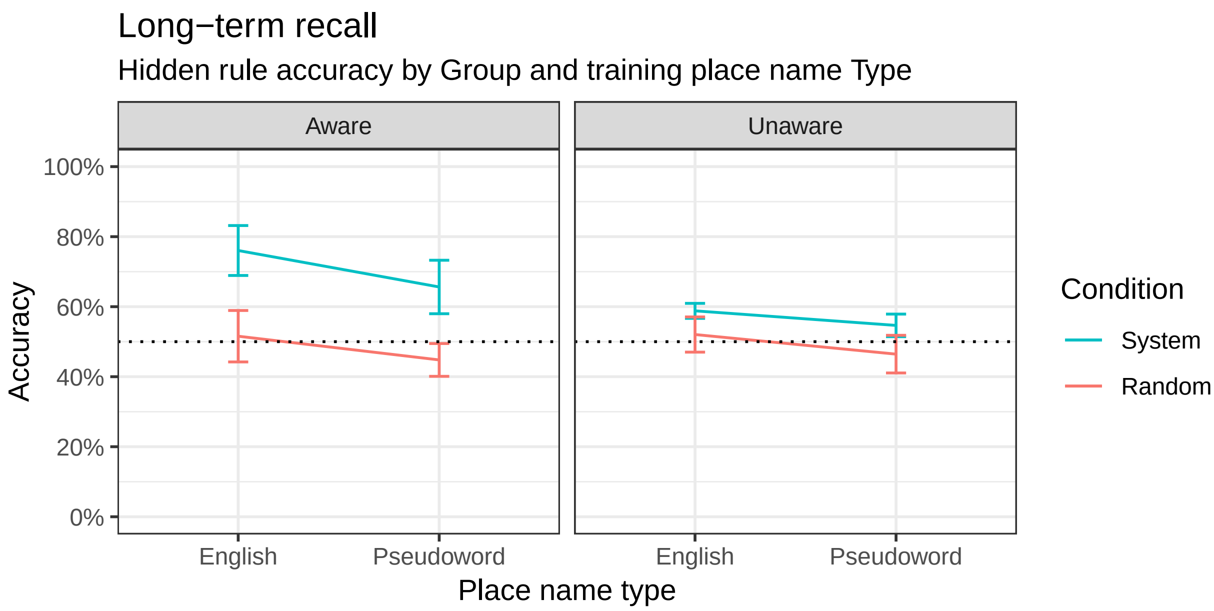


Figure 7.1: Average Hidden rule accuracy by Group in the Long-term recall task (Experiments 2 and 4), based on version of place name used for item presentation during training. Error bars represent SE of the mean, dotted line marks 50% chance level.

Experiment 4

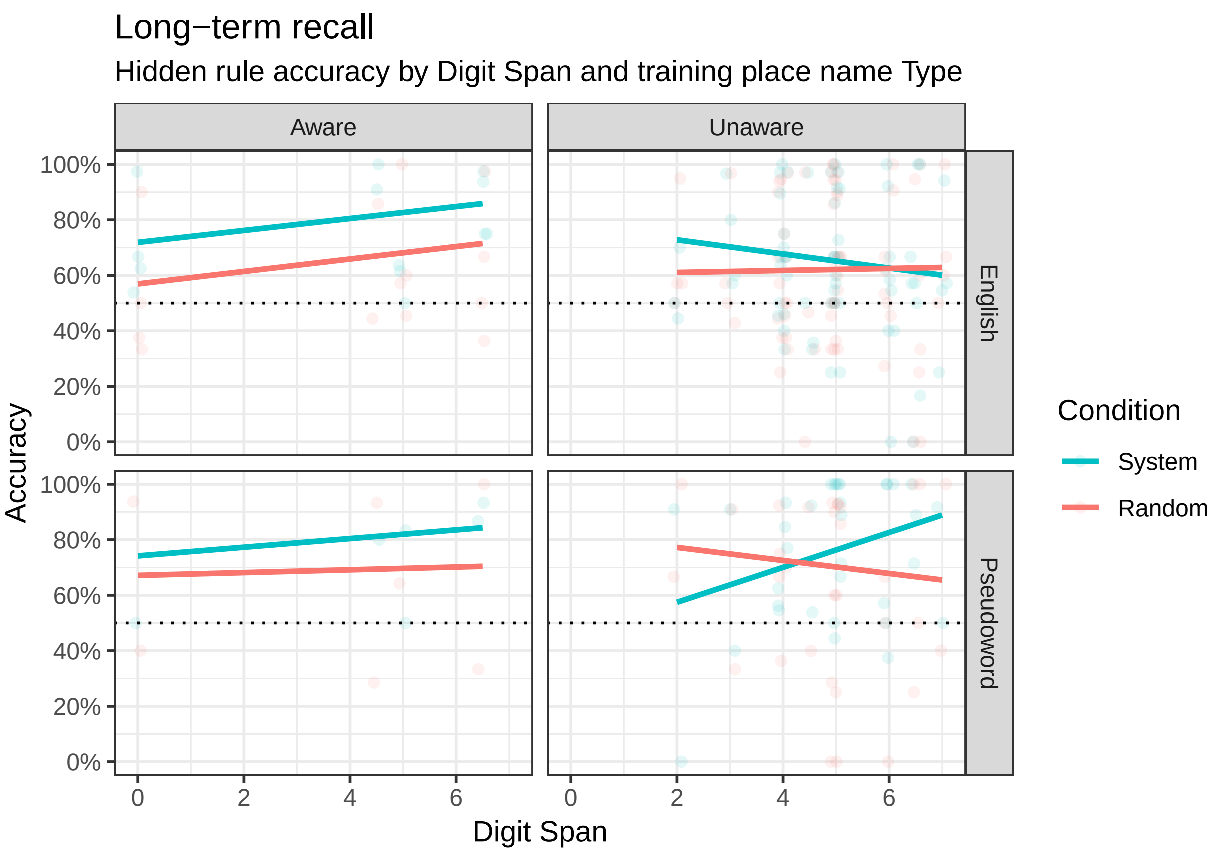


Figure 7.2: Effect of place name Type (noun vs. pseudoword) in Experiment 4 Long-term recall task by Group, Condition and Digit Span score. Solid lines represent fitted linear models.

Experiment 2, Pseudoword effect

	Coefficient	SE	χ^2	df	p	
Condition	-1.161	0.24	40.11	1	0.000	***
Type	-0.073	0.25	0.20	1	0.654	
Group	-1.254	0.30	4.11	1	0.043	*
Condition x Type	-0.260	0.34	5.22	1	0.022	*
Condition x Group	1.246	0.31	22.06	1	0.000	***
Type x Group	0.497	0.32	1.55	1	0.214	
Condition x Type x Group	-0.427	0.45	0.92	1	0.338	

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 7.3: Summary of GLMER model for Hidden rule accuracy, Long-term recall, with place name Type (English vs. Pseudoword) added as factor.

Visual inspection of the data (Fig. 7.2) shows that the 4-way interaction is due to an effect of pseudowords in the Unaware group, which is modulated by Digit Span score. Specifically, there is a trend for Unaware subjects at the high end of the Digit Span scale to perform similarly to Unaware subjects in Experiment 2: for training items which contain an English noun, there is no difference in recall between System and Random conditions; conversely, when training items contain a pseudoword, accuracy on those items in Long-term recall is higher if they are in the System condition, and lower if they are in the Random condition.

7.4 Use of questions

Besides the use of pseudowords in Experiments 2 and 4, the second potential source of difference we surveyed was the insertion of questions after each item in Experiment 3. In Experiment 3, all place nouns were presented in English; no pseudowords were used in the experiment apart from the spatial prepositions *gi*, *ro*, *wa* and *ne*. Instead, participants saw two questions: one asking them to provide a rating of the place pictured in the item, as a way of drawing attention to its physical characteristics, and one probing their knowledge of the Overt rule. In the following sections, we will examine the effects of item rating (Section 7.4.1) and Overt rule accuracy during training (Section 7.4.2) on Hidden rule accuracy

Experiment 4, Pseudoword effect

	Coefficient	SE	χ^2	df	p	
Condition	0.320	0.129	4.898	1	0.027	*
DS	0.148	0.139	0.863	1	0.353	
Type	0.241	0.081	6.981	1	0.008	**
Group	0.590	0.185	9.437	1	0.002	**
Condition x DS	0.122	0.097	1.216	1	0.270	
Condition x Type	0.075	0.081	0.099	1	0.753	
DS x Type	-0.045	0.062	0.651	1	0.420	
Condition x Group	0.152	0.129	0.913	1	0.339	
DS x Group	0.028	0.139	0.050	1	0.822	
Type x Group	0.088	0.081	0.590	1	0.443	
Condition x DS x Type	-0.086	0.062	1.615	1	0.204	
Condition x DS x Group	0.088	0.097	0.926	1	0.336	
Condition x Type x Group	0.044	0.081	0.042	1	0.838	
DS x Type x Group	0.087	0.062	1.947	1	0.163	
Condition x DS x Type x Group	0.129	0.062	4.355	1	0.037	*

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

Table 7.4: Summary of GLMER model for Hidden rule accuracy, Long-term recall, with place name Type (English vs. Pseudoword) added as factor.

in the Long-term recall task.

7.4.1 Item rating

In the training phase of Experiment 3, participants were asked to rate the attractiveness of each place pictured as a way to draw attention to its physical properties, on the assumption that they would come to associate those properties with the specific preposition they heard as part of that item. However, it is also possible that this manipulation merely focused participants' attention on the picture itself and how pleasant they found it, thus favouring the retention of specific items, rather than rule generalisation as hoped. If that were the case, we might expect to see a correlation between how highly participants rated an item during training, and how accurate they were when recalling it in the testing phase. To test this hypothesis, we broke down Hidden accuracy scores in the

Long-term recall task of Experiment 3 by the rating given to items during the training phase. We built a GLMER with Group, Condition and Item rating as predictors and random intercepts for subjects and items. There was no effect of Item rating on accuracy ($\chi^2(1) = 1.52$, $p = .218$), and no interactions of Item rating with other factors, either, indicating that the first question participants were asked did not have any effect on learning outcomes.

7.4.2 Overt rule accuracy

Overt rule accuracy, Experiments 2 - 4

		Training phase		Long-term recall	
	Group	System	Random	System	Random
Exp. 2	Aware (n = 16)	97%	96%	80%	74%
	Unaware (n = 17)	95%	91%	79%	76%
Exp. 3	Aware (n = 16)	96%	95%	91%	88%
	Unaware (n = 21)	96%	97%	93%	92%
Exp. 4	Aware (n = 4)	89%	97%	76%	83%
	Unaware (n = 20)	87%	85%	50%	45%

Table 7.5: Overt rule accuracy by group and experiment, Experiments 2 to 4.

The second question participants were asked, after each item presentation in the training phase of Experiment 3, tested their knowledge of the Overt rule. They were asked to indicate the character's position relative to the place pictured (at the place, or far from it), and received indirect feedback on their answer: the experiment would only resume once they had given the correct answer. This manipulation did seem to improve the acquisition of the Overt rule in Experiment 3, resulting in higher Overt rule accuracy scores during the Long-term recall task compared to other experiments (Table 7.5).

However, it also possible that, by drawing attention to the Overt rule, this paradigm may have been drawing attentional resources away from the Hidden rule, hindering its acquisition. If participants only had a finite amount of attention that they could allocate to either rule during the training phase, there may

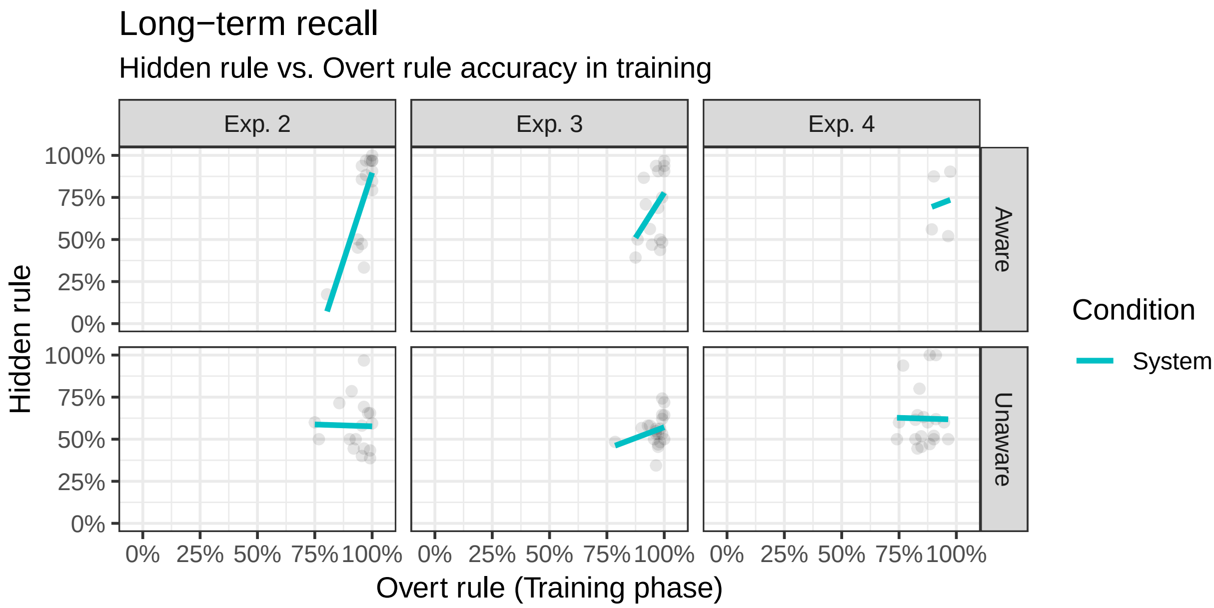


Figure 7.3: Average Hidden rule accuracy for System items in Long-term recall, plotted against mean Overt rule accuracy during training (for both System and Random items) with fitted linear models.

have been a trade-off between the two rules: the more attention was given to the Overt rule during training, the less would be available to the Hidden rule, which would make subjects less likely to acquire it. If that were the case, we may expect to see a negative correlation between Overt rule accuracy during the training phase (as a proxy of the amount of attention devoted to the rule) and Hidden rule accuracy in the testing phase. Specifically, we looked at Hidden rule accuracy for System items only, in order to isolate a measure of Hidden rule knowledge from the effect of item memory, which is what would drive recall accuracy for Random items.

To test our hypothesis, we constructed separate GLMER models for data from Experiments 2, 3 and 4, entering Group, Condition and Overt rule accuracy during training as predictors of Hidden rule accuracy for System items in the Long-term recall task (Fig. 7.3). In Experiment 2, we found an interaction of training Overt rule accuracy with Group ($\chi^2(1) = 15.76$, $p < .0001$), with post-hoc comparisons showing a strong positive correlation for the Aware group ($\chi^2(1) = 21.3$, $p < .0001$), but none for the Unaware. In Experiment 3, there was a significant main effect of training Overt rule accuracy ($\chi^2(1) = 5.20$, $p = .022$); however, Bonferroni-corrected post-hoc comparisons showed that it was only significant at the group level for the Aware group ($\chi^2(1) = 6.00$, $p = .029$) and not for the Unaware. Finally, in Experiment 4, we found no effect of training Overt rule accuracy ($\chi^2(1) = 0.008$, $p = .929$). This is not surprising: given the pattern observed in previous experiments, we may have expected an effect to be driven by the Aware group; however, the group of aware participants in this experiment was extremely small (n = 4), which would be insufficient to detect statistical trends.

7.5 Discussion

In order to shed light on the pattern of results we observed in Experiments 2, 3 and 4, we carried out a post-hoc analysis of the features characterising each of the two training paradigms we used: pseudowords replacing a subset of place nouns (Experiments 2 and 4), and questions after each item presentation (Experiment 3).

With regards to the use of pseudowords in Experiments 2 and 4, we hypothesised that it may have been responsible for the higher recall rates observed for System items in these experiments, together with lower Random recall rates. Our findings confirmed this hypothesis. In Experiment 2, we found that the use of pseudwords replacing place nouns in the training phase affected recall for those items. In the Unaware group, it led to lower recall accuracy for Random items and higher accuracy for System items. In fact, it seems to have been responsible for the learning effect observed in Unaware participants: for training items which used the English place noun, no difference could be detected in recall between System and Random items. For Aware participants, the use of pseudowords had the same effect on Random items, decreasing recall accuracy, but not on System items, for which accuracy was consistently high. This may have been due to the effect of explicit rule knowledge: if Aware participants were relying preferentially on explicit rule knowledge to perform the recall task, this could have masked any variation in implicit knowledge as a consequence of the pseudoword manipulation. In Experiment 4, we initially did not observe the same effect of pseudoword use, when averaging across participants by group. However, breaking down the data by Digit Span score, we saw that there was an effect in Experiment 4, too, which interacted with Digit Span. Specifically, Unaware participants at the high end of the Digit Span score displayed the same effect as those in Experiment 2: the use of pseudowords in training items led to better recall in testing if they were in the System condition, and worse if they were in the Random condition. By contrast, if the training items contained an English place noun, there was no difference in Long-term recall accuracy between conditions. These findings suggest that the use of pseudowords did indeed affect item memory and facilitate implicit rule learning in Experiments 2 and 4: possible explanations for this effect will be further explored in the general discussion (Chapter 8).

With regards to the use of questions in Experiment 3, our findings suggest that the questions did not have any measurable direct consequences for the development of implicit Hidden rule knowledge. The first question asked participants to provide ratings for the places they saw: we hypothesised that this may hinder learning by focusing participants' attention on the places themselves, rather than the association between places and prepositions. However, the post-hoc analysis showed that this was not the case: there was no correlation in either group between the rating given to an item in training, and recall accuracy for that item in the testing phase. The second question probed participants' knowledge of the Overt rule: we hypothesised that it may have been hindering acquisition of the Hidden rule, by drawing attentional resources away from it and focusing them on the Overt rule. Against our predictions, however, we did not find a negative correlation between Overt rule accuracy during training and Hidden rule accuracy in the testing phase. On the contrary, in the Aware group we found a positive correlation between the two (in Experiments 2 and 3); in the Unaware group there was no correlation whatsoever, either positive or negative. This suggests that performance on the Overt rule in training was not implicated in the development of implicit Hidden rule knowledge. Rather, it may have been connected with the ability to use explicit knowledge deliberately, as suggested by the positive correlation with Hidden rule recall we found in the Aware group.

Finally, even though we did not find a direct correlation between specific questions and lowered performance in testing, it is possible that the insertion of questions led to an overall decrement in learning because it forced participants to do the training procedure under dual-task conditions, which has been shown to be detrimental for implicit learning (Hendricks et al., 2013; Jiménez and A Vázquez, 2005; van Schagen, 2017). In an extension of Radulescu et al. (2019), van Schagen (2017) tried overloading subjects' working memory with a dual task instead of increasing entropy: not only did the experimental manipulation not improve learning, but participants in the dual-task condition actually performed worse than those in a single-task control condition (van Schagen, 2017). In a series of AGL experiments, Hendricks et al. (2013) found that dual-task requirements did not affect implicit learning of an artificial grammar under standard conditions, but that they did when the testing phase required transfer of the rules onto a novel set of letters. Hendricks et al. hypothesised that transfer to new letters, where perceptual similarity would be diminished, required some amount of explicit, intentional processing, unlike simple AGL learning. It is possible that the type of testing used in our experiment, which involved cued retrieval and production rather than simple pattern recognition, may have posed similar demands on subjects, requiring a degree of intentional processing. Therefore, learning in Experiment 3 could have been hindered by dual-task requirements, as it was in Hendricks et al. (2013).

7.6 Conclusion

The results of this analysis support our general hypothesis that learning outcomes in Experiments 2, 3 and 4 varied as a function of the training paradigm used. Admittedly, this was not designed as a between-groups study, and the fact that we performed the analysis post-hoc means that we could not control for variables such as participants' age and educational background. Crucially, we did not have measures of Digit Span for participants in Experiments 2 and 3. Therefore, any conclusions we can draw on the differences between the two experiments can only remain at the level of speculation. Nevertheless, the pattern of results emerging from our analysis strongly suggests that changes made to the training paradigm had a measurable impact on learning outcomes, in both Aware and Unaware participants, and plausibly determined the extent to which subjects could acquire implicit knowledge of the Hidden rule. This highlights the importance of using appropriate training procedures in implicit learning research: mere exposure is not enough, and even subtle changes to the way in which participants attend to the input can have an impact on their ability to extract patterns from it.

Chapter 8

General discussion

The main goal of this study was to investigate the development of implicit, productive knowledge of novel linguistic regularities. Our primary research question was whether it would be possible for learners to develop implicit knowledge of a novel linguistic rule and use it correctly in production. We found that, through recall-based activities, participants could develop implicit knowledge of a novel linguistic regularity, and use it correctly in spoken production without becoming aware of its content. Specifically, participants acquired implicit productive knowledge of a pair of novel spatial prepositions and their usage, which was based on a distinction between open and enclosed spaces found in natural language. Even though this rule bears some similarity to a rule found in participants' L1 (the *in/on* alternation in English), our analysis showed that participants' responses were not significantly influenced by transfer from L1 (even though it may have had a facilitating effect), indicating that they had acquired a novel distinction. However, we also found that learning outcomes were highly sensitive to the type of training paradigm used. The use of pseudowords replacing place nouns appeared to affect both recall and generalisation accuracy on the rule-based items. Our findings also suggest that this manipulation affected participants differently depending on working memory. Possible explanations for the observed effect of pseudowords, and its interaction with working memory, are discussed in Section 8.1.3.

Our second research question concerned the relationship between production and comprehension skills. We asked whether developing implicit knowledge through a production task would also lead to learning gains in comprehension. The data which we collected from two different types of receptive tasks suggests that there was minimal transfer of implicit knowledge from production to comprehension, supporting a skill-specific account of implicit knowledge (DeKeyser and Sokalski, 1996). However, there are a number of methodological issues which may have confounded these results, which means we can only draw tentative conclusions from them. Details of the comprehension tasks and of methodological problems affecting them can be found in Sections 8.2 and 8.5.

Our training paradigm was based on elicited recall, which was intended to promote rule learning by loading participants' memory capacity. We therefore hypothesised that working memory would predict rule learning. Against our predictions, we did not find a direct relationship between working memory and learning outcomes. Again, there are potential issues with the way memory load and working memory were conceptualised in this study, which may explain our findings. We did, however, find that working memory interacted with the effect of learning paradigm discussed above, potentially explaining differences between experiments in this respect. Based on the existing literature, we also predicted that overnight consolidation would aid the development of implicit generalisation. Our prediction was supported by the data; this is compatible with an emergentist account of language acquisition, where rules are acquired as a consequence of abstractions from stored exemplars. The effects of working memory, consolidation and potential individual differences on learning and awareness are discussed in Section 8.3.

As could be expected, participants who had become aware of the rule had higher accuracy than those who remained unaware, presumably because they could rely on explicit knowledge do to the task. However, even performance by aware participants was affected by changes to the training paradigm which we assumed had en effect on the development of implicit knowledge. Therefore, it is likely that aware participants were relying on a mixture of explicit and implicit knowledge when performing the task. Issues pertaining rule awareness and the development of explicit knowledge among participants are discussed in Section 8.4.

8.1 Development of productive rule knowledge

8.1.1 Main findings

In Experiment 1, we began to explore the use of cued recall as a way to detect implicit knowledge of a rule participants had been exposed to during training (the Hidden rule). In the testing phase, participants were exposed to items from training once again; every eighth item, a recall cycle began, in which they were asked to repeat the last eight items as cued by the corresponding pictures, which were presented in random order. We compared recall accuracy on System items (which were constructed following the Hidden rule) to that on Random items (which sampled randomly for a different pair of prepositions) on the assumption that if participants had acquired sensitivity to the Hidden rule, they should display higher recall accuracy for prepositions in the System condition, where they would be able to supplement item memory with rule knowledge. That is what we found for the Aware group, but not for the Unaware group. In Experiments 2, 3 and 4 the testing phase changed to long-term recall: participants received no further exposure to the items and were instead shown all the pictures corresponding to items from training, again in random order, and were asked to recall the sentence corresponding to each picture. In all experiments, participants who had become aware of the Hidden rule had significantly higher accuracy for System items, indicating that they were using their rule knowledge when doing the task. Unaware participants, by contrast showed no difference in accuracy between conditions in two of the experiments (Experiments 1 and 4), but showed a trend in that direction in the other two experiments (Experiments 2 and 4), which came close to being statistically significant. Our hypothesis was that recall accuracy would provide a measure of implicit rule knowledge: however, our findings suggest that was not the case. The evidence suggests that the Aware were relying on explicit rule knowledge when doing the recall task, while performance by the Unaware was still mainly driven by item memory. Reasons for this conclusion are laid out in Section 8.1.2.

In Experiments 2, 3 and 4, we also introduced generalisation items as part of the recall task. Generalisation items were made of familiar place pictures arranged into new combinations in the System condition. They were presented as recall items, but they had not actually appeared during training: therefore, producing the correct sentence associated with them would require productive use of the Hidden rule. As could be expected, the Aware group were significantly above chance when producing generalisation items in all three experiments, again indicating that they could rely on explicit knowledge of the rule to perform the task. Unaware participants, too, performed significantly above chance in this task in Experiments 2 and 4 (although not as well as the Aware group), which suggests that they had developed implicit knowledge of the Hidden rule and were using it productively to generate new items in this task. In Experiment 3, on the other hand, they did not perform above chance, suggesting they had not developed implicit knowledge in this experiment. As previously suggested, this was likely due to differences in the training paradigm, which are further discussed in Section 8.1.3.

The Hidden rule partially overlapped with a rule found in the participants' L1 (the *in/on* distinction in English. However, our additional analyses showed that participants did not just map the novel forms onto the English rule, but had acquired a novel distinction alongside the novel forms. While there was a numerical advantage in accuracy rates for items that could be transferred from English, the difference was not significant. Therefore, even though L1 transfer may have aided participants, it could not by itself explain their performance (with the exception of the Transfer subgroup in Experiment 3, which is compatible with our conclusion that participants in that experiment did not develop any implicit knowledge of the Hidden rule).

8.1.2 Comparison of recall and generalisation: are they indexing the same factor?

In Experiment 1, we used short-term recall as a way to investigate implicit rule knowledge: however, we found that only participants who had become aware of the rule showed a learning effect, meaning that the task did not reveal any evidence of implicit knowledge. Furthermore, analysing performance by the Unaware group across experiments 2, 3 and 4, we see a consistent discrepancy between

System item recall and generalisation. For this group, we observed above-chance generalisation accuracy in two of the experiments (Experiments 2 and 4), but we never found a clear effect of Hidden rule knowledge on recall. We see a trend towards an effect in recall in these experiments, but it always falls just short of statistical significance. If both generalisation accuracy and recall accuracy for System items were driven by implicit knowledge of the Hidden rule, we would expect them to correlate with each other: that was indeed the case for the Aware group, where we found a strong correlation between System item recall and generalisation across all three experiments. For the Unaware group, on the other hand, the connection between between System recall and generalisation accuracy was weaker, with only a statistically significant correlation in Experiment 2.

Our interpretation for these findings is that System recall accuracy in the Unaware group was driven mostly by better memory for System items, rather than productive rule knowledge. In generalisation, on the other hand, Unaware subjects had to rely on productive rule knowledge to perform the task, since no item memory was available. This explains how it could be possible for the Unaware group to be above chance in generalisation, but not show a significant effect of the rule in recall: in this group, there was an imperfect correlation between recall for System items and generalisation, because they relied on different, although related, sources. By contrast, System recall accuracy in the Aware group could be supported by explicit productive rule use, as generalisation was, leading to a significant effect for both measures and a high correlation between the two. The Aware group had explicit knowledge of the rule, which they could use deliberately during both recall and generalisation trials. The Unaware group, on the other hand, only had implicit knowledge, which cannot be used deliberately (Cleeremans and Jiménez, 2002): therefore it is likely that, for this group, if implicit rule knowledge emerged at all during the recall task it would only "fill the gaps" left by item memory.

There is, of course, a close connection between recall accuracy for System items and the development of implicit rule knowledge: we assume that rule knowledge emerged as a consequence of generalisation from individual items stored in memory. As we have seen, however, the opposite is not true: recall accuracy for individual System items in the Unaware group was driven by memory, not by implicit rule knowledge. How, then, should we explain the trend towards an advantage for System items that the Unaware exhibited in recall? It is possible that better memory for System items was a direct consequence of the Hidden rule, which had the effect of increasing their semantic consistency. The fewer unrelated meanings a lexeme has, the more semantically consistent it is; semantic consistency in lexical items has been shown to aid both word recognition and word learning (Rodd et al., 2002, 2012). System items were based on the Hidden rule, which assigned each System preposition to either enclosed or open spaces only; therefore, System prepositions were semantically more consistent than Random ones, because they were always used with the same kinds of noun. This may have affected the ease with which System items could be memorised and recalled, even prior to the development of productive rule knowledge. Further support for this interpretation comes from the observed effect of training paradigm on recall and generalisation, which is discussed in Section 8.1.3, as well as data from comprehension tasks (Section 8.2).

8.1.3 The role of training paradigm

Unlike our first experiment, Experiments 2, 3 and 4 all used variations of the same procedure, which included recall every two items in training, and long-term recall together with generalisation in testing. As previously mentioned, we found differences in Hidden rule recall and generalisation accuracy between these three experiments, which we attributed to differences between the specific versions of the training procedure used. Generalisation was only observed in Experiments 2 and 4, which used a training paradigm in which a subset of place nouns were replaced by pseudowords, unlike Experiment 3; these experiments also yielded higher recall rates for System items. The analysis we performed in Chapter 7 showed that, in Experiment 2, the introduction of pseudowords had the effect of improving memory for System items in the Unaware group, while it led to worse recall for Random items for both groups. This effect supports our hypothesis that System recall accuracy among the Unaware was driven by item memory, rather than rule knowledge. If System item recall depended exclusively on rule knowledge, then we should see no difference based on how the item was presented in training. That is indeed what we observed for the Aware group, who showed an effect of pseudoword use on recall for Random items, but not for System ones. Their recall accuracy for System items was consistently high regardless of noun type, because, as argued earlier, it was driven by explicit rule knowledge rather than item memory.

If, as previously argued, semantic consistency is what was driving higher recall rates for System items, this suggests that the use of pseudowords replacing English place nouns in training items enhanced the effect of semantic consistency. In fact, it is possible that the manipulation was increasing the effect of semantic consistency by forcing learners to pay more attention to the relevant properties of the places pictured. They had to do so because a potential aid to recall (the English label for the place) had been removed: therefore, to correctly recall the preposition, they had to associate it with the visual depiction of the place, which would lead to greater activation of features relevant to the rule (open vs. enclosed space). For System items, this had a beneficial effect, because the same prepositions were always associated with similar kinds of pictures, thus making it easier to encode and recall them. The introduction of the pseudoword, however, also increased task demands, since more material now needed to be encoded in order to be able to recall the sentence: the preposition, the pseudoword replacing the place noun, and visual information concerning the place picture (because the pseudoword, unlike the English place noun, did not provide any information in that sense). This would increase memory load, which should make recall harder. Even though separate stores exist in working memory for visual and verbal information, evidence suggests that thy share the same central capacity limit, due to limited attentional resources (Cowan, 2001, 2015), and that when both verbal and visual information have to be retained at the same time, there is a trade-off between the two (Saults and Cowan, 2007). In System items, it appears that the increased memory load was offset by the beneficial effect of semantic consistency, leading to higher recall rates despite increased task demands. For Random items, on other hand, the introduction of pseudowords in training led to lower recall rates: since these items did not benefit from semantic consistency, their recall deteriorated as a consequence of greater task demands. This interpretation is also compatible with our finding, in Experiment 4, that the effect of pseudowords on item recall was modulated by working memory capacity: only high-WM participants in this experiment showed the effect observed in Experiment 2, with a recall advantage for System items. It would appear that, in order to reap the benefits of semantic consistency for System items, participants needed to have enough working memory capacity to be able to store visual information about the place, alongside the phonological material.

8.2 Transfer from production to comprehension

The findings of the generalisation task supported our main experimental hypothesis: participants could develop implicit knowledge of the Hidden rule and use it productively, given the right circumstances. Our second research question concerned the nature of this knowledge: we asked whether the knowledge developed through a production task, which could be detected in productive behaviour, would also translate into sensitivity to the rule in a comprehension task. By contrasting comprehension measures with the outcome of the production tasks, we could attempt to gain insight into the nature of the knowledge acquired. We hypothesised that, if the knowledge acquired in training amounted to abstract implicit rule knowledge, we should see evidence of it in a comprehension test, too. On the other hand, if it is the case that implicit knowledge requires modality-specific proceduralisation to emerge, as argued by DeKeyser (DeKeyser and Sokalski, 1996; Li and DeKeyser, 2017), we would expect to see a dissociation between the results of the production and the comprehension test. We found some evidence for a proceduralisation account: in the listening task, sensitivity to the rule seemed to emerge over the course of the task, more so for participants who had explicit knowledge of the rule. However, this sensitivity only developed for items which had already been encountered during training, not to generalisation or violation items, which is what we would have expected if participants had developed knowledge of the rule. We found a similar pattern in the recognition memory task, with a general preference for previously encountered System items, but no evidence of generalisation; sensitivity to violation was limited to Aware participants. This suggests that participants, overall, had developed little or no receptive implicit knowledge of the rule, which would support a skill-specific account. However, several methodological issues may have confounded the outcomes of the tasks we used, which are discussed below in greater detail.

8.2.1 Listening comprehension

In Experiment 1, we used a listening task with timed responses to test for comprehension skills. Participants heard the sentences as a pair of pictures appeared on screen (one depicting an open space, the other an enclosed one); their task was to select the picture matching the noun as quickly as possible. If participants were sensitive to the Hidden rule in comprehension, we predicted that they would respond faster to System items, where the preposition offered a cue to the type of upcoming noun. What we did find was that, for Training items, a difference between conditions was not visible at the start, but it gradually emerged over the course of the task. The trend was visible for both groups, but it was only statistically significant for the Aware group, where it was more pronounced. This finding would seem to indicate that they were relying on explicit declarative knowledge to begin with and that this began to be proceduralised over the course of the RT task, leading to a gradual increase in processing speed. This is compatible with DeKeyser and Sokalski's (1996) account of implicit knowledge in comprehension and production as separate skills, which require specific practice to develop.

However, several factors speak against this interpretation of the emerging difference between System and Random items observed in the Aware group. Firstly, a trend in the same direction, while not statistically significant, was also observed in the Unaware group, who ought to have no explicit knowledge of the rule. Secondly, results from the rest of the listening task are not compatible with what we would have expected to see, had the participants developed true rule knowledge. After Training items, the task also included a set of new items conforming to the training pattern (Generalisation items) and another new set in which System items were altered to make them ungrammatical (Violation items). For Generalisation items, we were expecting to observe the same pattern we had observed for Training items, with faster responses for System items. For Violation items, we were expecting to see a reversal of the pattern, with a slowdown for System items. Neither of these predictions were confirmed: the effect of Condition disappeared shortly after the beginning of the Generalisation block, and no further changes were detected. This suggests that participants were not predicting upcoming content in the first phase of the listening task, but were simply faster at recognising the items: they were relying on memory rather than rule. A comprehension-specific skill did develop, but it was a recognition skill, rather than rule-based. To summarise, the result of the listening task showed no development of receptive implicit knowledge in the Unaware group. In the Aware group, we saw emerging sensitivity to item type recognition, but no generalisation or sensitivity to violation. However, it is still possible that they had acquired receptive rule knowledge but could not use in this task, because it was requiring them to use it predictively: literature on L2 processing shows that early L2 learners lack the ability to anticipate upcoming content based on morphosyntactic dependencies, unlike L1 speakers (Lew-Williams and Fernald, 2010; Martin et al., 2013). To address this point, in subsequent experiments we used a different comprehension task, which did not require subjects to formulate predictions.

8.2.2 Recognition memory

In Experiments 2 and 3, we tested comprehension using a recognition memory task. Participants were presented with a mixture of new and old items, and asked whether they thought they had already encountered those items during training, giving them a rating on a scale from 1 ("Definitely no") to 6 ("Definitely yes"). In both experiments, when presented with items already encountered during training, both aware and unaware participants were significantly more likely to endorse System items.

Initially, we compared System and Random items to obtain a measure of Hidden rule sensitivity, as we did in production. However, as previously discussed, our findings from the production task showed that the comparison between conditions in recall was not, as we had hoped, tapping into implicit rule knowledge: rather, it was sensitive to item memory in the Unaware, and to explicit rule knowledge in the Aware. Likewise, direct comparison of conditions in comprehension suffered from a general endorsement bias for System items. Even in the Aware group, sensitivity to Hidden rule violation only emerged when using signal detection measures and comparing *d'* scores, rather than directly comparing System to Random items. However, even using signal detection measures we did not find any evidence of generalisation in either the Aware or Unaware group. This suggests that subjects did not have implicit receptive knowledge, and that sensitivity to rule violation in the Aware was driven by explicit knowledge.

As previously mentioned, in both recognition memory tasks we found an endorsement bias for System items, but only for those items that had already been seen during training. This seems to stem from a general preference for System items rather than sensitivity to the rule: it also appeared in situations where Unaware participants otherwise showed no sensitivity to the rule (Experiment 3), and even caused them to endorse ungrammatical System items over Random ones, suggesting that it was due to visual features rather than the verbal content. However, it could not just be a general preference for the visual features of System items: if that were that case, we should see a preference for System items in generalisation trials, too, which is not what we found. Rather, there seemed to be an ease of recognition for the visual representations of System items from training, possibly due to visual features which made them easier to encode than Random items. For instance, the character in the System condition was superimposed onto the picture, rather than next to it as in the Random condition: this meant that participants needed to attend to a smaller screen area, which may have made it easier to encode the items. This would be compatible with previous findings by Kinder et al. (2003) on the effect of cognitive fluency in implicit learning: in an AGL task, they manipulated the ease of recognition for items during the training phase by adding visual noise, and found that items which were more easily perceived were also more likely to be judged as grammatical in the testing phase later (Kinder et al., 2003).

To summarise, in comprehension tasks we observed an advantage for training items in the System condition, as we did in the recall task, but we found no evidence of generalisation in either group, or of sensitivity to violation in the Unaware group. We only found evidence of sensitivity to violation in the Aware group; however, the fact that this group did not show any evidence of generalisation as well suggests that they were mainly relying on their explicit knowledge to detect rule violation. This implies that participants could not transfer implicit knowledge from the production task to the comprehension task, supporting a skill-specific account of implicit knowledge.

8.3 Cognitive factors and individual differences

8.3.1 Working memory

In using elicited recall as our training paradigm, we hypothesised that it would promote rule extraction by loading participants' short-term memory capacity. This was based on the assumption underlying the Elicited Oral Imitation task (Slobin and Welsh, 1973), on which our paradigm was based, that elicited sentence recall involves a process of decoding and re-encoding of linguistic material (Potter and Lombardi, 1990). We also followed previous work on rule induction (Radulescu, 2014; Radulescu et al., 2019) which showed that loading participants' memory capacity could promote the extraction of regularities from complex syllable strings. Therefore, in Experiment 4 we introduced a measure of phonological WM (digit span task), expecting to find a correlation between WM capacity and learning outcomes. Against our predictions, we found no direct effect of WM on either recall or generalisation accuracy in the production task.

One possible reason for this finding is that the measure of WM we used did not capture aspects of this construct relevant to L2 learning. In a study on the incidental learning of Samoan by Robinson (2005), WM correlated with performance on a production task in the testing phase. However, Robinson did not measure WM with a digit span task, as we did, but used a reading span task. This task requires participants to maintain words in mind while simultaneously processing full sentences, which involves semantic and morphosyntactic processing as well as loading phonological short-term memory; accordingly, it has been shown to correlate with reading comprehension scores, unlike the digit span task (Daneman and Carpenter, 1980). It may be the case that what was driving the correlation found in Robinson (2005) was the linguistic component indexed by the reading span task, which may also play a role in language learning. The digit span task, while not entirely free from linguistic influence either (Schmidt et al., 2019), arguably constitutes a "purer" measure of phonological WM than the reading span task. The lack of correlation of digit span scores with learning outcomes in our study is compatible with findings by Grey et al. (2015), who found no effect of phonological WM (assessed by non-word repetition task) on learning outcomes in a study of the implicit learning of a semi-artificial language.

However, we did find an interaction of WM with the pseudoword effect. This suggests that the measure of working memory we used captured the factor which allowed subjects to take advantage of the presence of pseudowords in items during training, enabling them to memorise visual information about the place pictured together with the preposition and the pseudoword. Therefore, it is possible that working memory was responsible for the amount of material participants could encode upon stimulus presentation, but that it was not implicated in their ability to draw implicit generalisations from that material. It has been suggested that WM may be a better predictor of learning under explicit, rather than implicit, conditions (Tagarelli et al 2011, Grey et al 2015), and that measures of declarative and procedural memory may be more relevant for implicit learning (Antoniou et al., 2016; Ettlinger et al., 2014; Morgan-Short et al., 2014). In Radulescu (2014) and Radulescu et al. (2019), memory load was achieved by increasing entropy (input complexity) rather than the amount of material to be memorised, which may have been a critical feature: in an extension of Radulescu (2014), van Schagen (2017) found that working memory capacity, measured by digit span task, was not a predictor of learning outcomes in their paradigm. All this suggests that working memory may not have been a relevant construct to our experiment, which investigated the development of implicit knowledge and its productive generalisation, while measures of procedural and declarative memory may have been more relevant.

8.3.2 Rule extraction and memory consolidation

In our last experiment, we also added the opportunity for overnight consolidation, including a second block of generalisation testing 24 hours after the training session. Following the literature on consolidation in language learning (e.g. Gaskell and Dumay, 2003; Gaskell et al., 2014; Morgan-Short et al., 2012a; Tamminen et al., 2012, 2015) we expected participants' scores on the generalisation task to improve after sleep, as a consequence of integration and abstraction from individual items. Compatibly with our hypothesis, we did observe an improvement in generalisation accuracy on Day 2, even though the difference was not statistically significant. In the Unaware group, however, the improvement meant that they performed significantly above chance in generalisation on the second day of testing, while they showed no evidence of generalisation on the first day.

The fact that generalisation in the Unaware group only emerged on Day 2, as a consequence of memory consolidation, is compatible with previous research on the integration of novel morphology in L2 (Tamminen et al., 2012, 2015): Tamminen et al. (2015) found evidence of integration on implicit measures only after overnight consolidation, while explicit measures showed an effect already immediately after training. The fact that generalisation benefitted from memory consolidation (and therefore, strengthening of item memory traces) also suggests that it may be driven by patterns of similarity between items, rather than the application of abstract categorical rules, which is compatible with computational modelling of results from similar implicit learning experiments (Alikaniotis, 2018). In Williams (2005), participants acquired implicit knowledge of the rule governing a pair of novel determiners, which alternated based on noun animacy. Applying a distributional semantics analysis to the items used in Williams (2005), Alikaniotis (2018) found that accuracy in generalisation was sensitive to collocational patterns for training and testing nouns in natural language, suggesting that semantic relatedness, rather than the application of an abstract categorical rule, was driving participants' responses.

In Experiment 4, the fact that generalisation only emerged after consolidation for the Unaware may have also contributed to the weak correlation between recall and generalisation scores, by introducing an additional source of individual variability to the transition from individual representations to abstract rules may vary between individuals. Research on memory consolidation shows that there is a limit to the amount of material which can be consolidated overnight (Feld et al., 2016), which entails that memory consolidation is selective. A range of extraneous factors can affect what is consolidated: for instance, emotionally charged material is preferentially consolidated over neutral material (Payne et al., 2008), and so is information which is relevant to a subject's future plans and expectations, compared to irrelevant information (Fischer and Born, 2009; Wilhelm et al., 2011).

While overnight consolidation may be helpful, however, evidence from IL research shows that it is not strictly necessary for implicit generalisation to emerge: implicit learning studies, going as far back as Reber (1967), often take place over the course of one session only, showing immediate learning effects. In SLA, implicit generalisation of novel form-meaning connections can emerge and be detectable immediately after training (Leung and Williams, 2011; Paciorek and Williams, 2015; Williams, 2005). Similarly, in our study, we found evidence of generalisation in Unaware participants immediately after training in one of our experiments (Experiment 2). Therefore, there is a discrepancy in our study: in Experiment 2, we found implicit generalisation already on the day of training, while in Experiment 4, it only emerged after overnight consolidation. The reasons for this discrepancy may be connected to both differences between populations and the development of rule awareness, which will be discussed in the following sections.

8.3.3 Differences between populations

Participants in Experiments 1, 2, and 3 were recruited from among students at the University of Cambridge and surrounding community. The use of university students as subjects, while very common in psychological research, has been criticised for sampling from a restricted, specific group that is in many ways not representative of the population as a whole (Henrich et al., 2010). By contrast, participants in Experiment 4 were older, less likely to speak foreign languages and to be in education than the subjects we used for our first three experiments. In general terms, Experiment 4 replicated the findings of Experiment 2, with above-chance generalisation among the Unaware indicating that they had developed implicit knowledge of the Hidden rule, and could use it productively. However, there were also differences: most notably, generalisation only emerged for the Unaware on Day 2, and accuracy on the Overt rule in long-term recall was far lower in this experiment than in Experiment 2, despite the fact that all participants were aware of the rule. Subjects were also less likely to become aware of the Hidden rule, relative to previous experiments. Some of these discrepancies may be due to individual differences which we did not account for, resulting from the fact that we sampled from two different distributions: university students for the first three experiments, and general population for the fourth one.

The lower accuracy on the Overt rule observed among Unaware subjects in Experiment 4, for instance, suggests that they were less able to use explicit knowledge strategically under time-pressured conditions, which could be explained by individual differences such as executive function. Individual cognitive differences might also explain lower rates of Hidden rule awareness. The effect of pseudwords which we found in Experiment 2 was reproduced in Experiment 4, but only among participants with high digit score spans. Due to the lack of working memory measures in previous experiments, we cannot say for certain, but it is possible that the distribution of working memory scores in the student population we surveyed in our first three experiments was skewed towards the higher end of the scale, relative to the general population: it has been shown that educational attainment correlates with higher performance on a range of cognitive measures, including working memory (Guerra-Carrillo et al., 2017; Ritchie et al., 2015). It is also likely that the different level of foreign language skills and exposure played a role. Apart from one, none of the subjects of Experiment 4 reported speaking a foreign language, while the vast majority of subjects from previous experiments spoke at least one. Previous research on the implicit learning of novel form-meaning connections showed that experience of learning a foreign language with gendered determiners facilitated the acquisition of novel determiners, even if they encoded animacy instead of gender (Williams, 2005); this is likely because subjects were used to paying attention to determiner forms as a potential source of information. More generally, prior experience of language, both L1 and L2, shapes attention patterns, which affects the kind of information that learners are able extract from the input (e.g. N. Ellis and Sagarra, 2010).

Finally, the specificities of Experiment 4 may also be due to contingent factors related to the way the experiment was run, rather than underlying differences in the population. Unlike previous experiments, Experiment 4 was run online, allowing participants to take part remotely using their own desktop computers. The advantage of online data collection is that it gives researchers the opportunity to collect large amounts of data, over a short period of time, sampling from a large population. A possible disadvantage of running experiments online is that it does not afford as much control over participant behaviour as a laboratory setting would. Despite this, research on the topic suggests that the results of internet-based data collection are mostly comparable with lab-based studies (Clifford and Jerit, 2014; Crump et al., 2013): implicit learning studies have also been successfully run online, including a replication of Williams (2005) using a crowdsourcing platform (Kerz et al., 2017). In our study, it is highly unlikely that participants could have significantly altered outcomes through "cheating" behaviour, such as writing down items for later recall, due to both the speed of presentation and the nature of the task (since recall was cued by pictures). Furthermore, this kind of strategy might have been possible for recall items but not for generalisation ones, which were the crucial measure by which we established whether participants had acquired productive implicit knowledge of the rule. On the other hand, lack of engagement with the task may have a constituted a bigger problem. The higher proportion of subjects who did not become aware of the Overt rule and had to be eliminated from analysis suggests that participants, on average, were not as attentive in this experiment as they would have been in a laboratory setting. Given that implicit learning, as shown in the literature and in our study, is sensitive to even minor changes in experimental procedure, we cannot rule out the possibility that lower attention had an impact on the outcomes of Experiment 4, possibly explaining the delay in the emergence of generalisation as well as overall lower rates of Hidden rule awareness.

8.4 Implicit and explicit knowledge: the role of awareness

Our comparison of experiments, which shows variation in Hidden rule accuracy in both the Aware and the Unaware groups, gives us an indirect way to assess the development of implicit knowledge, suggesting that it was affecting the performance of both groups. An analysis of performance by Aware participants across experiments may therefore offer some insights in the interaction between implicit and explicit knowledge, as well as the development of the latter.

Aware participants' accuracy on System item recall, like that of the Aware group, varied considerably between experiments. It was considerably higher in Experiment 1 than in the following experiments. This may not seem surprising, since the recall task in Experiment 1 was easier than the one used in the following experiments: participants received additional exposure to each of the items, and only had to recall eight at the time, instead of the whole training set. Therefore, item memory was better in this experiment, as evidence by high recall accuracy for Random items, too. However, if Aware participants were relying primarily on their explicit, categorical knowledge of the Hidden rule when doing the task, we should not see such a big difference; indeed, they should have consistently high accuracy for System items in all experiments. Instead, their accuracy on System items seemed to track variation in accuracy in the Unaware group: they were consistently about 15-50% more accurate than the Unaware group. This suggests that Aware performance on System items was influenced by item memory, as well as rule knowledge. The Aware group also exhibit similar fluctuation in generalisation accuracy between experiments, which again suggests that they were relying on a mixture of explicit and implicit knowledge, and shows that the generalisation task was sensitive to both implicit and explicit knowledge. Therefore, having a reliable measure of awareness is crucial to establish whether implicit knowledge was developed during the task, by ruling out the possibility that Unaware participants, too, may have been relying on explicit knowledge during generalisation testing.

Retrospective verbal reports, which we used to measure rule awareness in our study, have been criticised for potentially not being sensitive enough to conscious knowledge. Specifically, the concern is that there may be underreporting of conscious knowledge, if participants are not confident or unable to accurately report the content of their knowledge (Rebuschat, 2013; Shanks and St. John, 1994). To address this point, we tried to make our questionnaire as sensitive as possible by including multiple indirect questions aimed at probing participants' knowledge of the rule, by asking participants to provide a translation of the novel prepositions as well as providing their impressions on preposition usage (Appendix A).

Additionally, some of the criticism addressed to the use of retrospective verbal report stems form concerns which are not relevant to our particular experimental setup. The original criticism of verbalisation as a measure of awareness concerned its use in AGL experiments such as Reber (1967), where the rule to be learned was usually a complex finite-state grammar, which could only be verbalised as an intricate set of conditional rules. The Hidden rule in our experiment, by contrast, was a simple binary distinction which would be quite easy to formulate in abstract categorical terms, being based on the distinction between open and enclosed spaces. Therefore, we would argue that it is unlikely that if someone had some level of awareness of it they would not be able to express it verbally, either as an abstract rule or when providing translations for the prepositions. Furthermore, concerns about under-reporting of conscious knowledge usually applied to studies which used receptive tasks as measures of rule knowledge, while ours used a production task. It could be argued that the strength of representation required for correctly using a form in production is greater than that required for sensitivity to it in comprehension; therefore, a representation that is strong enough to be causally efficacious in production, but still cannot be verbalised, is likely to be truly unconscious.

Despite this, it is possible that, even though participants in the Unaware group were not yet able to verbalise their knowledge at the time of testing, they may have been in the process of developing explicit knowledge, and that this may have driven their performance in the generalisation task. The fact that we found generalisation already on the day of training, in Experiment 2, is in contrast with previously cited findings by Tamminen and colleagues, which showed that generalisation on implicit measures could only occur after sleep consolidation (Tamminen et al., 2012, 2015), while generalisation on explicit measures could be observed immediately after training. Therefore, it is possible that emerging explicit knowledge was driving generalisation by Unaware participants in Experiment 2 immediately after the training session. We may speculate that Unaware participants in Experiment 2, at the moment of testing, were more "advanced" towards awareness than those of Experiment 4, which would explain both their better performance in generalisation and the fact that more participants had become aware of the Hidden rule by the end of Experiment 2, than in Experiment 4. Had the experiment lasted longer, it is possible that the Unaware group, too, would have become fully aware of the Hidden rule, and eventually become able to verbalise it.

On the other hand, aspects of the Aware group's performance contradict this account of emerging awareness driving performance, suggesting a categorical difference between the two groups instead. One is the different effect of pseudowords on the two groups: in Experiment 2, Aware participants showed the same reduction in Random item recall accuracy as the Unaware did when pseudowords were used, but they did not show a corresponding increase in System accuracy. They were just as accurate on System items regardless of whether the item contained a pseudoword during training, suggesting that their performance was driven by the application of a rule, while that of the Unaware was driven by item memory. Another categorical difference we observed between Aware and Unaware participants is the correlation between Overt rule accuracy in training and Hidden rule accuracy in recall, which we only found among Aware participants. The Overt rule was acquired by subjects through a process of explicit rule search, as prompted by task instructions, and was used deliberately; the correlation with Hidden rule accuracy in recall suggests this process, too, was a consequence of deliberate explicit rule use in the Aware group, but not in the Unaware. Therefore, it appears that Aware and Unaware were in two distinct states with regards to the application of explicit knowledge during the task, which goes against an account of emerging awareness driving performance.

A second possibility is that it was not awareness that was graded, but the strength of the underlying representation: awareness could be discrete and epiphenomenal, a consequence of strong emerging representations. According to graded accounts of awareness, the quality of mental representations, including their strength of activation, determines how likely they are to enter conscious awareness (Cleeremans and Jiménez, 2002; Dehaene et al., 2006). Dehaene et al. (2006) draw a distinction between different kinds of unconscious representations in perception, *subliminal* and *preconscious* ones. Both are characterised by a lack of conscious attention to the stimulus; however, they differ in stimulus strength and consequent degree of activation. Subliminal representations are so weak that the subject would not become aware of them, even in the presence of directed conscious attention. Preconscious representations, on the other hand, are characterised by stronger and more widespread activation, such that they have the potential to be consciously accessed, but temporarily remain unconscious as long as attention is engaged elsewhere. In our case, the relevant representation would be the subjects' own implicit knowledge of the Hidden rule (or more properly, patterns of association between prepositions and broad clusters of nouns). As the strength of the representation increases, going form subliminal to preconscious, participants are increasingly likely to become aware of their implicit structural knowledge. When they do, they develop explicit knowledge, formulating an abstract categorical rule which describes the observed pattern, and which may enter into a synergy with existing implicit knowledge, leading to a positive shift in performance. This is compatible with previous research on the interaction between implicit and explicit knowledge in AGL (Mathews et al., 1989), which found that a process of implicit learning followed by explicit rule discovery yielded better learning outcomes than either implicit or explicit learning alone. In our study, different strength of implicit representations would explain the differences in Unaware performance between Experiments 2 and 4, while the emergence of awareness, and consequent explicit rule knowledge, would explain the differences observed between Aware and Unaware participants across experiments.

8.5 Study limitations and directions for future research

Several of the conclusions we drew from our experimental findings, while they fit the available data, must remain tentative due to a number of methodological limitations in the study. One such limitation is the fact that we did not have measures of cognitive abilities for any of the experiments, apart from Experiment 4. For instance, our interpretation of the interaction between pseudowords and working memory in Experiment 4 rests on the assumption that participants recruited in previous experiments had working memory scores above population average; however, we cannot test the assumption due to the lack of cognitive measures in these experiments. Additionally, the only cognitive measure that we have for our subjects is phonological working memory, which, against our predictions, did not predict learning outcomes in our study. Adding other measures, such as tests of declarative and procedural memory, or executive function, may have helped to shed more light on individual variation in processes involved in acquiring the Hidden rule, such as abstraction from individual representations and the relationship between Overt and Hidden rule accuracy. Therefore, a potential development of the current study would be to replicate our experiments with the addition of more detailed cognitive testing, including measures of procedural and declarative memory, in order to shed light on the individual differences involved in learning thorough elicited recall and in the transition from item memory to generalisation. Additionally, our comparison of training paradigms was carried out post-hoc, which means that factors found to have potential explanatory power (such as the use of pseudowords, or sampling from different populations) were not controlled for in advance. A potential avenue for development in this respect would be to carry out a controlled comparison of the elements which we found to have an effect on learning outcomes, particularly the effect of pseudowords, which, if confirmed, could have significant pedagogical implications. A further potential development connected with individual differences would be to explore the role of foreign language experience. While we collected information on foreign languages spoken, this was primarily used a way to control for potential confounds (e.g. knowledge of Slavonic languages), rather than as a factor of interest in its own right. However, as previously mentioned, this may have been one source of variation contributing to different performance across the population samples we tested. It has been argued that bilinguals have better meta-linguistic awareness (Cummins, 1978; Galambos and Goldin-Meadow, 1990), which can help learners to infer grammatical rules and dissociate form from content. This could have affected performance in our experiment, both in terms of rule acquisition and in the development of rule awareness. Further analysis could therefore focus on the effect of bilingualism as a potential explanatory variable in our data, either as a binary variable (bilingual vs. monolingual) or as a n-ary variable (number of languages spoken). Similarly, gender could also be analysed as a potential factor of interest, as there is evidence that gender may affect L2 learning strategies (Dörnyei, 2005; Oxford et al., 1988).

A second problem area for our methodology lies in the choice of tasks used to test for comprehension skills. The listening task used in Experiment 1 required participants to process input predictively, which is often problematic for early L2 learners. Therefore, this may mean that the task was not sensitive enough to implicit knowledge. Additionally, even Aware participants in this task showed limited receptive knowledge of the rule, with no generalisation to new instances or sensitivity to violation. This may be due to the fact that they were tested immediately after training, without the opportunity for overnight consolidation. Given the literature on consolidation and integration in implicit measure (Tamminen et al., 2012, 2015), it is possible that a second testing session after the opportunity for overnight consolidation would have yielded different results, and that generalisation in comprehension may have emerged. Likewise, in the recognition memory task used in Experiments 2 and 3, we found sensitivity to violation (in the Aware), but no generalisation, which suggests responses were driven by explicit knowledge. It is possible that the integration required for implicit generalisation to show would have required overnight consolidation. Additionally, we only included a measure of generalisation in the recognition memory task of Experiment 3, in which Unaware participants did not develop implicit knowledge of the rule, but not in Experiment 2, where they did develop implicit productive knowledge. Therefore, we cannot know whether they would have performed differently from the Aware, had we included the task in Experiment 2: they might have shown implicit generalisation immediately after training. Further research could address all these open questions by replicating the training paradigm of Experiment 2 with the addition of both immediate and delayed comprehension tasks, including measures of generalisation and well as sensitivity to violation, in order to better explore the interplay between overnight consolidation and the development of receptive knowledge.

A third problem in our study was the visual difference between conditions, which may have created a confound: System items were always represented by a character superimposed onto a picture, while Random items always featured the character and the picture side by side. The comprehension tasks revealed a general bias towards System items, which may have resulted from a visual preference for them. It is also possible that participants preferred System items because the situation they depicted was less ambiguous and more ecologically valid from a linguistic point of view than that encoded by Random items. System sentences could be clearly translated using a preposition such as "in/on/at" in English, while Random items did not have an equally straightforward translation, partially because of the vagueness of the spatial relationship they encoded: they were variously translated by subjects as "next to", "away from" or "near". However, we did not counterbalance the type of visual representation across conditions, precisely in order to maintain ecological validity. In Czech, as in many other natural languages, prepositions used to indicate position in a location are sensitive to the type of location, but those that indicate proximity are not: the Czech equivalents of "near" or "next to" are not sensitive to the open vs. enclosed space distinction, unlike *v* and *na*. Counterbalancing the visual representation would have meant assigning the rule-based condition to a visual depiction encoding proximity, which would have been unnatural from a linguistic point of view. We know that implicit learning can occur with a variety of natural language rules, but that it is less likely to happen when participants are exposed to linguistically unnatural ones, such as novel classifiers encoding relative size (Chen et al., 2011; Leung and Williams, 2014). At any rate, the visual difference between conditions may have constituted a confound for comprehension tasks, and, potentially, for the long-term recall task. It would not have had any consequences for the generalisation task, which was crucial to establishing whether participants had acquired productive implicit knowledge, because a preference for System visual items alone could not have explained participants' ability to generalise the Hidden rule in production.

Besides these considerations, however, it may be argued that the ecological validity of our study was still limited, because we did not use a true natural language rule: while it was derived from natural language, the rule employed in our study was a reduced, simplified version of a natural rule. The *v/na* alternation in Czech, it may be recalled, relies on patterns of prototypicality as well as some morphophonological regularities, while also including idiosyncratic elements. By contrast, the Hidden rule in our study could be captured by a straightforward categorical distinction between open and enclosed spaces, which is the distinction that Aware participants reported discovering. It is unclear whether we would have observed the same learning outcomes if participants had been exposed to the natural Czech rule instead, instead of the simplified version. Further research could address this point by moving away from the categorical rule we used, and instead training participants on a more complex or “fuzzy” one, akin to the real Czech rule, in which the distribution of forms was following probabilistic patterns. Existing literature on studies using artificial language suggests that learners should acquire and reproduce the patterns of statistical distribution characterising probabilistic in the input, when presented with a probabilistic distribution (Wonnacott et al., 2017, 2008). If we were to replicate these findings with the natural Czech rule, it would greatly improve the ecological validity of our findings, which in turn would lend support to implicit learning as a potentially beneficial component in L2 instruction.

8.6 Conclusion

The findings of our study show that it is possible to acquire implicit knowledge of a novel linguistic rule through a production task, and use it in spoken production to generate new instances, while remaining unaware of the contents of the rule. The subjects in our study could successfully produce new sentences following the rule they had learned, regardless of whether they had become aware of it. This was not significantly influenced by L1 transfer: participants learned a novel distinction and could apply it productively, rather than just map the novel forms onto existing L1 distinctions (even though transfer from L1 may have provided some added facilitation). Our findings show that participants could develop productive implicit knowledge through production activities, against the predictions made by Processing Instruction (VanPatten and Cadierno, 1993). However, just as guiding attention is central to PI interventions, so it must be in production-based training tasks. Learning outcomes in our experiments were dependent on the specific training paradigm used, which confirms that implicit learning is very sensitive to the way in which attention is oriented: mere exposure to the stimuli is not enough to obtain implicit learning.

Participants who had developed explicit knowledge of the rule consistently outperformed participants who remained unaware of it. However, they were also, to some extent, affected by changes in the training paradigm in a manner similar to unaware participants, which suggests that their performance was based on a combination of implicit and explicit knowledge. At the same time, performance on the Overt rule - which participants were told about and encouraged to discover - was not consistently high, even though all subjects included in the analysis were aware of its contents. This is compatible with previous findings from AGL research, showing that neither memorisation nor explicit rule search by themselves lead to optimal results, while the best learning outcomes were obtained through a combination of memorisation and explicit learning, arranged in sequence (Mathews et al., 1989).

Our results also suggest that there is a skill-specific aspect to implicit knowledge, supporting DeKeyser's position (DeKeyser, 1997; DeKeyser and Sokalski, 1996). We found little to suggest that the knowledge developed through recall-based training could be used in comprehension tasks to generalise to new instances; the only instances of transfer we found could be attributed to explicit knowledge. This mirrors the findings of studies using comprehension-based IP activities such as VanPatten and Cadierno (1993) and VanPatten and Sanz (1995), which led on improvement on all measures apart from, crucially, production tasks tapping into implicit knowledge. This suggests that a combination of comprehension and production activities, both designed in such a way as to direct participants' attention to the relevant form-meaning connections, may be optimal for the development of fluency and automaticity L2 learning.

Appendix A

Debriefing questionnaire

Participant code:

Debriefing Questionnaire

Age: **Gender:**

1. Did you think the use of words *ro, wa*, *ne* and *gi* was governed by any rules?

☐ Yes ☐ No

2. Did you think it depended on whether the character was inside or outside/near the place pictured?

☐ Yes ☐ No

3. Did you think there were any other rules?

☐ Yes ☐ No

If you answered 'No', go to Question 4. If you answered 'Yes', please describe the rules:

4. When during the experiment did you come to the conclusion that this was a rule?

☐ Part 1 ☐ Part 2 ☐ Part 3

5. How confident were you? (1 = Not confident at all, 5 = Very confident)

1 2 3 4 5

6. Could you give a rough translation of the words *ro, gi, wa* and *ne*?

☐ Yes ☐ No

If you answered YES, please write any translations below:

Thank you for taking part in this experiment.

Appendix B

Instructions to participants

Welcome! Thank you for taking part in this study.

The experiment is composed of three parts, and will last about 35 minutes in total. You are now about to start Part 1: the following set of instructions will show you how to complete it.

Press the space bar to advance through the instructions.

These are Harry and Lucy

Press the space bar to continue

You will hear a series of sentences about Harry and Lucy. Some of the words will not be in English - these words describe where Harry and Lucy are, relative to a given place.

As you hear each sentence, the corresponding picture will be displayed on screen. Every two sentences, the screen background will turn grey and you will see each of the pictures again, in random order.

Press the space bar to continue

Your task will be to repeat the sentence associated with each of the pictures.

You will only have a limited amount of time to repeat each sentence.

Press the space bar to start the practice

After practice:

You are now ready to begin the first part of this experiment.

Pay attention to each character's position in the picture. Different words are used for different positions, and your task is to learn which words are associated with which positions. You will be tested on this later.

Press the space bar to begin

Appendix C

Items

All nouns and corresponding pictures were used in Experiments 1 to 4, with the exception of the following items which were introduced from Experiment 2: open spaces - *airport, boat, catwalk, dam, market, ship, train station, vineyard*; enclosed spaces - *airplane, canteen, hairdresser's, hotel, living room, pub, shopping centre, shower*.

Open spaces

1. Airport

2. Beach

3. Boat

4. Bridge

5. Bus stop

6. Campsite

7. Car park

8. Catwalk

9. Cemetery

10. Dam

11. Desert

12. Dock

13. Field

14. Garden

15. Golf course

16. Hill

17. Ice rink

18. Island

19. Market

20. Motorway

21. Mountain

22. Pavement

23. Petrol station

24. Playground

25. Red carpet

26. Roof

27. Roundabout

28. Savannah

29. Ship

30. Skate park

31. Ski slope

32. Square

33. Stadium

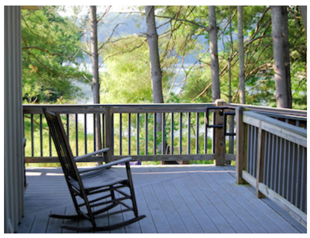
34. Stairs

35. Tennis court

36. Terrace

37. Train station

38. Veranda



39. Vineyard

40. Zebra crossing

Enclosed spaces

41. Airplane 

42. Aquarium 

43. Bakery 

44. Bank 

45. Bathroom 

46. Bedroom 

47. Canteen 

48. Casino 

49. Cave 

50. Cellar 

51. Changing room 

52. Church 

53. Classroom

54. Factory

55. Garage

56. Greenhouse

57. Gym

58. Hairdresser's

59. Hospital

60. Hotel

61. Kitchen

62. Laboratory

63. Library

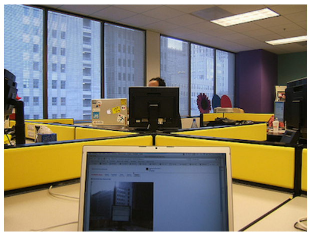
64. Lift

65. Living room

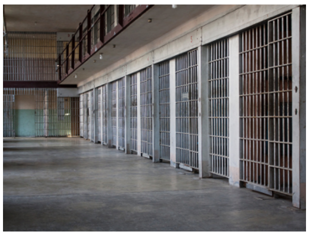
66. Museum

67. Nightclub

68. Office



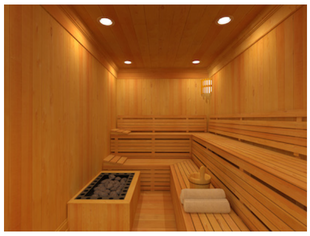
69. Prison



70. Pub

71. Restaurant

72. Sauna



73. Shopping centre

74. Shower

75. Stable

76. Supermarket

77. Swimming pool

78. Tent

79. Theatre

80. Tunnel

List of pseudonouns used

The following pseudowords were used to replace a subset of place names (n = 32) in Experiments 2 and 4:

groam	*breel*
fribe	*trape*
zule	*cray*
shay	*pline*
dreet	*kipe*
bleam	*frobe*
voom	*preem*
ploan	*blune*
prole	*hain*
brile	*frean*
snope	*neak*
glane	*soam*
koop	*shoop*
prait	*fay*
raim	*trule*
draid	*sloat*

References

Alikaniotis, D. (2018). *Understanding Semantic Implicit Learning through Distributional Linguistic Patterns*. PhD thesis, University of Cambridge. 175

Alonso, D., Fuentes, L. J., and Hommel, B. (2006). Unconscious symmetrical inferences: A role of consciousness in event integration. *Consciousness and Cognition*, 15(2):386–396. 31

Altmann, G., Dienes, Z., and Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4):899. 21

Amato, M. S. and MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, 116(1):143–148. 19

Ambridge, B., Pine, J. M., Rowland, C. F., Chang, F., and Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: word learning, morphology, and verb argument structure. *Wiley Interdiscip. Rev. Cogn. Sci.*, 4(1):47–62. 1

Anderson, J. (1993). *Rules of the Mind / [Edited by] John R. Anderson with the Collaboration of Francis S. Bellezza ... [et Al.]*. LErlbaum Associates, Hillsdale, N.J. 42

Anderson, J. (1996). ACT: A Simple Theory of Complex Cognition. *American Psychologist*, 51(4):355–365. 9

Anderson, J. R. (1983). *The Architecture of Cognition / John R. Anderson.* Number 5 in Cognitive Science Series. Harvard University Press, Cambridge, Mass. ; London. 9

Anderson, J. R. and Lebiere, C. (1998). *The Atomic Components of Thought.* Hillsdale, NJ: Lawrence Erlbaum Associates. 9

Andringa, S. and Curcic, M. (2015). HOW EXPLICIT KNOWLEDGE AFFECTS ONLINE L2 PROCESSING. *Studies in Second Language Acquisition*, 37(02):237–268. 78

Antoniou, M., Ettlinger, M., and Wong, P. C. M. (2016). Complexity, Training Paradigm Design, and the Contribution of Memory Subsystems to Grammar Learning. *PLOS ONE*, 11(7):e0158812. 145, 174

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science*, 9(4):321–324. 18

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412. 65

Baddeley, A. (1992). Working memory. *Science*, 255(5044):556–559. 129

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in cognitive sciences*, 4(11):417–423. 129

Baddeley, A. (2003). Working memory and language: An overview. *Journal of communication disorders*, 36(3):189–208. 129

Baddeley, A. (2015). 1. Working Memory in Second Language Learning. In Wen, Z. E., Borges Mota, M., and McNeill, A., editors, *Working Memory in Second Language Acquisition and Processing*, pages 17–28. Multilingual Matters, Bristol, Blue Ridge Summit. 129

Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, 105(1):158. 129

Baddeley, A. D. and Hitch, G. (1974). Working memory. In *Psychology of Learning and Motivation*, volume 8, pages 47–89. Elsevier. 129

Bardovi-Harlig, K. (2000). Tense and aspect in second language acquisition: Form, meaning, and use. *Language Learning: A Journal of Research in Language Studies*, 50:1. 25

Barlow, M. and Kemmer, S. (2000). *Usage-Based Models of Language*. Cambridge University Press. 40

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3). 65

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious Mixed Models. *arXiv:1506.04967 [stat]*. 65

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48. 65

Bates, E. and MacWhinney, B. (1989). Functionalism and the competition model. *The crosslinguistic study of sentence processing*, 3:73–112. 42

Batterink, L. J. and Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, 90:31–45. 15

Batterink, L. J., Paller, K. A., and Reber, P. J. (2019). Understanding the Neural Bases of Implicit and Statistical Learning. *Topics in Cognitive Science*, 11(3):482–503. 11

Batterink, L. J., Reber, P. J., Neville, H. J., and Paller, K. A. (2015a). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, 83:62–78. 55

Batterink, L. J., Reber, P. J., and Paller, K. A. (2015b). Functional differences between statistical learning with and without explicit training. *Learning & Memory*, 22(11):544–556. 11, 15, 34

Bertels, J., Franco, A., and Destrebecqz, A. (2012). How implicit is visual statistical learning? *J. Exp. Psychol. Learn. Mem. Cogn.*, 38(5):1425–1431. 11

Bigand, E., Perruchet, P., and Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*. 21

Bjork, R. A. (1994). Institutional impediments to effective training. In Druckman, D. and Bjork, R. A., editors, *Learning, Remembering, Believing: Enhancing Human Performance*, pages 295–306. National Academy Press, Washington, DC. 51

Bley-Vroman, R. and Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In Tarone, E. E., Gass, S. M., and Cohen, A. D., editors, *Research Methodology in Second-Language Acquisition*, pages 245–261. Erlbaum, Hillsdale, N.J. 17

Bock, K., Dell, G. S., Chang, F., and Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, 104(3):437–458. 42

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135. 66

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in second language acquisition*, 33(2):247–271. 18

Branigan, H. P., Pickering, M. J., and Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, 75(2):B13–25. 42

Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., and Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24(6):489–506. 42

Brashers-Krug, T., Shadmehr, R., and Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, 382(6588):252. 131

Burnham, W. H. (1903). Retroactive amnesia: Illustrative cases and a tentative explanation. *The American Journal of Psychology*. 131

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6):1563–1569. 52

Carpenter, S. K. and DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 19(5):619–636. 52

Cave, C. B. (1997). Very Long-Lasting Priming in Picture Naming. *Psychological Science*, 8(4):322–325. 40

Chambers, K. E., Onishi, K. H., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87(2):B69–B77. 2, 19

Chapman, G. B. and Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, 18(5):537–545. 31

Chen, W., Guo, X., Tang, J., Zhu, L., Yang, Z., and Dienes, Z. (2011). Unconscious structural knowledge of form–meaning connections. *Consciousness and Cognition*, 20(4):1751–1760. 23, 185

Cheng, A.-C. (1995). *Grammar Instruction and Input Processing: The Acquisition of Spanish" Ser" and" Estar"*. PhD Thesis, University of Illinois at Urbana-Champaign. 43, 44

Chetail, F. (2017). What do we do with what we learn? statistical learning of orthographic regularities impacts written word processing. *Cognition*, 163:103–120. 19

Chondrogianni, V., Vasić, N., Marinis, T., and Blom, E. (2015). Production and on-line comprehension of definiteness in English and Dutch by monolingual and sequential bilingual children. *Second Language Research*, 31(3):309–341. 42

Christensen, R. H. B. (2019). *Ordinal—Regression Models for Ordinal Data*. 90

Christiansen, M. H. (2019). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*, 11(3):468–481. 12

Clahsen, H. and Felser, C. (2006). Grammatical processing in language learners. *Applied psycholinguistics*, 27(1):3–42. 25, 41

Clahsen, H. and Felser, C. (2018). SOME NOTES ON THE SHALLOW STRUCTURE HYPOTHESIS. *Studies in Second Language Acquisition*, 40(3):693–706. 41

Clahsen, H., Felser, C., Neubauer, K., Sato, M., and Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning*, 60(1):21–43. 41

Cleeremans, A., Destrebecqz, A., and Boyer, M. (1998). Implicit learning: News from the front. *Trends in cognitive sciences*, 2(10):406–416. 5, 13

Cleeremans, A. and Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. *Implicit learning and consciousness*, pages 1–40. 6, 166, 181

Clifford, S. and Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, 1(2):120–131. 133, 178

Consortium, B. N. C. (2007). *British National Corpus, Version 3 BNC XML Edition*. 56

Conway, C. M. and Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1):24–39. 11, 21

Conway, C. M. and Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, 21(4):561–580. 11, 21

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114. 168

Cowan, N. (2015). 2. Second Language Use, Theories of Working Memory and the Vennian Mind. In Wen, Z. E., Borges Mota, M., and McNeill, A., editors, *Working Memory in Second Language Acquisition and Processing*, pages 29–40. Multilingual Matters, Bristol, Blue Ridge Summit. 168

Crump, M. J., McDonnell, J. V., and Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one*, 8(3):e57410. 133, 178

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 14(3):215–235. 52

Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *J. Cross. Cult. Psychol.*, 9(2):131–149. 183

Cunillera, T., Laine, M., Càmara, E., and Rodríguez-Fornells, A. (2010). Bridging the gap between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, 63(3):295–305. 19

Custers, R. and Aarts, H. (2011). Learning of predictive relations between events depends on attention, not on awareness. *Consciousness and cognition*, 20(2):368–378. 32

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4):450–466. 173

Davis, M. H. and Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536):3773–3800. 132

De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Appl Linguist*, 13(1):1–24. 37

De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language learning*, 46(3):529–555. 38, 40

De Jong, N. (2005). CAN SECOND LANGUAGE GRAMMAR BE LEARNED THROUGH LISTENING?: An Experimental Study. *Studies in Second Language Acquisition*, 27(2):205–234. 40

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, 47(1):1–12. 134

De Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., and Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: Evidence from sequence learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598):2065–2076. 19

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, 10(5):204–211. 181

DeKeyser, R. (2003). Implicit and Explicit Learning. In Doughty, C. J. and Long, M. H., editors, *The Handbook of Second Language Acquisition*, pages 312–348. Blackwell Publishing Ltd, Oxford, UK. 2, 8, 9, 129

DeKeyser, R. (2007). Skill acquisition theory. In VanPatten, B. and Williams, J., editors, *Theories in Second Language Acquisition: An Introduction*, volume 97113, pages 97–114. Lawrence Erlbaum Associates, Mahwah, N.J. 9, 45

DeKeyser, R. (2017). Knowledge and skill in ISLA. In Loewen, S. and Sato, M., editors, *The Routledge Handbook of Instructed Second Language Acquisition*, pages 15–32. Routledge, 1 edition. 42

DeKeyser, R. M. (1995). Learning Second Language Grammar Rules. *Studies in Second Language Acquisition*, 17(03):379–410. 26, 29, 30, 40, 41

DeKeyser, R. M. (1997). BEYOND EXPLICIT RULE LEARNING. *Studies in Second Language Acquisition*, null(02):195–221. 8, 9, 42, 45, 46, 49, 82, 187

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, 22(4):499–533. 7

DeKeyser, R. M. and Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46(4):613–642. 3, 26, 27, 29, 42, 45, 46, 76, 82, 163, 169, 170, 187

Dell, G. S. and Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Phil. Trans. R. Soc. B*, 369(1634):20120394. 46

Dell, G. S., Reed, K. D., Adams, D. R., and Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6):1355–1367. 19

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word preactivation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, 8(8):1117. 78

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*. 58

Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Macmillan International Higher Education. 126

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5. 125, 126

Dienes, Z., Altmann, G., Kwan, L., and Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5):1322. 14

Dienes, Z., Broadbent, D., and Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5):875. 14

Dienes, Z. and Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and brain sciences*, 22(5):735–808. 5, 6

Dienes, Z. and Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, 69(5):338–351. 13, 54, 55

Dörnyei, Z. (2005). *The Psychology of the Language Learner : Individual Differences in Second Language Acquisition*. Taylor & Francis Group, Mahwah, UNITED STATES. 183

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in second language acquisition*, 13(4):431–469. 40

Doughty, C. and Williams, J. (1998). *Focus on Form in Classroom Second Language Acquisition*. Cambridge University Press. 8

Doughty, C. J. (2003). Instructed SLA: Constraints, Compensation, and Enhancement. In Doughty, C. J. and Long, M. H., editors, *The Handbook of Second Language Acquisition*, pages 256–310. Blackwell, New York. 16, 32

Dumay, N. and Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18(1):35–39. 132

Ellis, N. C. (1994). Implicit and explicit learning of languages. 6, 9, 24

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language learning*, 48(4):631–664. 40

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, 24(2):143–188. 25

Ellis, N. C. (2005). AT THE INTERFACE: DYNAMIC INTERACTIONS OF EXPLICIT AND IMPLICIT LANGUAGE KNOWLEDGE. *Studies in Second Language Acquisition*, 27(02):305–352. 2, 8, 16, 18, 24

Ellis, N. C. (2006a). Language Acquisition as Rational Contingency Learning. *Applied Linguistics*, 27(1):1–24. 12, 13, 31, 32

Ellis, N. C. (2006b). Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning. *Applied Linguistics*, 27(2):164–194. 12, 31

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. *Implicit and explicit language learning*, pages 35–47. 2

Ellis, N. C. (2015). Implicit AND explicit language learning: Their dynamic interface and complexity. In *Implicit and Explicit Learning of Languages*, pages 1–24. John Benjamins. 8, 9

Ellis, N. C. and Cadierno, T. (2009). Constructing a second language: Introduction to the special section. *Annual Review of Cognitive Linguistics*, 7(1):111–139. 10

Ellis, N. C. and Sagarra, N. (2010). THE BOUNDS OF ADULT LANGUAGE ACQUISITION. *Studies in Second Language Acquisition*, 32(04):553–580. 31, 32, 36, 40, 177

Ellis, R. (1988). The role of practice in classroom learning. *AILA review*, 5:20–39. 40, 43

Ellis, R. (2006c). Modelling Learning Difficulty and Second Language Proficiency: The Differential Contributions of Implicit and Explicit Knowledge. *Applied Linguistics*, 27(3):431–463. 2, 16

Ellis, R. (2009a). Implicit and explicit leaning, knowledge and instruction. In Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., and Reinders, H., editors, *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, volume 42. Multilingual Matters. 11

Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., and Reinders, H., editors, *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, volume 42. Multilingual Matters. 16

Ellis, R. (2012). *Language Teaching Research and Language Pedagogy.* Language Teaching Research and Language Pedagogy. John Wiley and Sons. 2

Erlam, R. (2006). Elicited Imitation as a Measure of L2 Implicit Knowledge: An Empirical Validation Study. *Applied Linguistics*, 27(3):464–491. 16, 17, 55, 75

Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, pages 65–93. 16, 17, 18, 53, 55, 75

Erlam, R., Loewen, S., and Philp, J. (2009). The roles of output-based and input-based instruction in the acquisition of L2 implicit and explicit knowledge. *Implicit and explicit knowledge in second language learning, testing and teaching*, pages 241–261. 44

Estes, K. G., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological science*, 18(3):254–260. 19

Ettlinger, M., Bradlow, A. R., and Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics*, 35(4):807–831. 145, 174

Faretta-Stutenberg, M., Morgan-Short, K., Granena, G., Koeth, J., Lee-Ellis, S., and Lukyanchenko, A. (2011). Learning without awareness reconsidered: A replication of williams (2005). In *Selected proceedings of the 2010 Second*

Language Research Forum: Reconsidering SLA research, dimensions, and directions, pages 18–28. academia.edu. 23

Feld, G. B., Weis, P. P., and Born, J. (2016). The Limited Capacity of Sleep-Dependent Memory Consolidation. *Frontiers in Psychology*, 7. 175

Fenn, K. M., Nusbaum, H. C., and Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, 425(6958):614–616. 132

Ferry, A. L., Fló, A., Brusini, P., Cattarossi, L., Macagno, F., Nespor, M., and Mehler, J. (2016). On the edge of language acquisition: inherent constraints on encoding multisyllabic sequences in the neonate brain. *Dev. Sci.*, 19(3):488–503. 19

Ficca, G., Lombardo, P., Rossi, L., and Salzarulo, P. (2000). Morning recall of verbal material depends on prior sleep organization. *Behavioural brain research*, 112(1-2):159–163. 131

Fischer, S. and Born, J. (2009). Anticipated reward enhances offline learning during sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6):1586. 176

Fitch, W. T. and Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656):377–380. 19

Fletcher, P. C., Zafiris, O., Frith, C. D., Honey, R. a. E., Corlett, P. R., Zilles, K., and Fink, G. R. (2005). On the benefits of not trying: Brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. *Cerebral Cortex (New York, N.Y.: 1991)*, 15(7):1002–1015. 26, 35

Forkstam, C., Jansson, A., Ingvar, M., and Petersson, K. M. (2009). Modality transfer of acquired structural regularities: A preference for an acoustic route. In *31th Annual Conference of the Cognitive Science Society (CogSci 2009)*. Cognitive Science Society. 21

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition. 66

Francis, A. P., Schmidt, G. L., Carr, T. H., and Clegg, B. A. (2009). Incidental learning of abstract rules for non-dominant word orders. *Psychological research*, 73(1):60–74. 2, 20

Franco, A., Cleeremans, A., and Destrebecqz, A. (2016). Objective and subjective measures of cross-situational learning. *Acta Psychol.*, 165:16–23. 11

French, L. M. (2006). *Phonological Working Memory and Second Language Acquisition: A Developmental Study of Francophone Children Learning English in Quebec*. Edwin Mellen Press. 129

French, L. M. and O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29(3):463–487. 129

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4):128–135. 131

Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, 99(1):529–534. 15, 46

Frost, R. L. A. and Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147:70–74. 21

Fyfe, E. R., DeCaro, M. S., and Rittle-Johnson, B. (2015). When feedback is cognitively-demanding: The importance of working memory capacity. *Instructional Science*, 43(1):73–91. 130, 133

Gabrieli, J. D., Cohen, N. J., and Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition*, 7(2):157–177. 24

Gais, S. and Born, J. (2004). Declarative memory consolidation: Mechanisms acting during human sleep. *Learning & Memory*, 11(6):679–685. 131

Galambos, S. J. and Goldin-Meadow, S. (1990). The effects of learning two languages on levels of metalinguistic awareness. *Cognition*, 34(1):1–56. 183

Gaskell, M. G. and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89(2):105–132. 132, 174

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., and Stackhouse, A. (2014). Sleep Underpins the Plasticity of Language Production. *Psychological Science*, 25(7):1457–1465. 132, 174

Gass, S. M. and Mackey, A. (2007). Input, interaction, and output in second language acquisition. *Theories in second language acquisition: An introduction*, 175199. 40

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., and Ellis, R. (2015). TIMED AND UNTIMED GRAMMATICALITY JUDGMENTS MEASURE DISTINCT TYPES OF KNOWLEDGE: Evidence from Eye-Movement Patterns. *Studies in Second Language Acquisition*, 37(2):269–297. 15, 16

Godfroid, A. and Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In Rebuschat, P., editor, *Implicit and explicit learning of languages*, volume 48, page 325. John Benjamins Amsterdam. 15

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science*, 13(5):431–436. 19

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34(3):445–474. 130, 133

Goo, J., Granena, G., Yilmaz, Y., and Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In Rebuschat, P., editor, *Implicit and Explicit Learning of Languages*. John Benjamins Publishing Company, Philadelphia, UNITED STATES. 26, 28, 29, 30, 40

Grey, S., Williams, J. N., and Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences*, 38:44–53. 130, 174

Griffin, Z. M. and Ferreira, V. S. (2006). Properties of spoken language production. In *Handbook of Psycholinguistics*, pages 21–59. Elsevier. 40

Guerra-Carrillo, B., Katovich, K., and Bunge, S. A. (2017). Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample. *PLOS ONE*, 12(8):e0182276. 177

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in second language acquisition*, 35(3):423–449. 16

Hahne, A. and Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4(2):123–141. 7

Hama, M. and Leow, R. P. (2010). LEARNING WITHOUT AWARENESS REVISITED. *Studies in Second Language Acquisition*, 32(03):465–491. 23, 40

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learn. Individ. Differ.*, 44:9–15. 7

Hawkins, R. and Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, 118(4):595–612. 13

Hay, J. F., Pelucchi, B., Estes, K. G., and Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive psychology*, 63(2):93–106. 18, 19

Hendricks, M. A., Conway, C. M., and Kellogg, R. T. (2013). Using dual-task methodology to dissociate automatic from nonautomatic processes involved in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5):1491–1500. 160

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83. 176

Hrdlička, M. (2000). Předložky ve v̀yuce češtiny jako cizího jazyka: Acta Universitatis Carolinae philologica: Monographia. 50

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18(3):193–223. 2, 8, 9

Hulstijn, J. H. (2003). Incidental and intentional learning. In Doughty, C. J. and Long, M. H., editors, *The Handbook of Second Language Acquisition*, pages 349–381. Blackwell, New York. 10, 11

Hulstijn, J. H. (2005). THEORETICAL AND EMPIRICAL ISSUES IN THE STUDY OF IMPLICIT AND EXPLICIT SECOND-LANGUAGE LEARNING: Introduction. *Studies in Second Language Acquisition*, 27(2):129–140. 2

Hulstijn, J. H. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning. In Rebuschat, P., editor, *Implicit and Explicit Learning of Languages*. John Benjamins Publishing Company, Philadelphia, UNITED STATES. 8

Hulstijn, J. H., De Graaff, R., et al. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *Aila Review*, 11(97-112). 26

Hyltenstam, K. and Abrahamsson, N. (2003). Maturational constraints in SLA. In Doughty, C. J. and Long, M. H., editors, *The Handbook of Second Language Acquisition*, pages 538–588. Blackwell, New York. 31

Izumi, S. (2002). OUTPUT, INPUT ENHANCEMENT, AND THE NOTICING HYPOTHESIS: An Experimental Study on ESL Relativization. *Studies in Second Language Acquisition*, 24(4):541–577. 38

Izumi, S., Bigelow, M., Fujiwara, M., and Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in second language acquisition*, 21(3):421–452. 38

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, USA. 1

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, 30(5):513–541. 15

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, 25(4):603–634. 25

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language learning*, 57(1):1–33. 25

Jiménez, L. and A Vázquez, G. (2005). Sequence learning under dual-task conditions: Alternatives to a resource-based account. *Psychological research*, 69:352–68. 160

Jiménez, L. and Méndez, C. (1999). Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1):236–259. 32, 33

Johnson, E. K. and Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, 44(4):548–567. 19

Karpicke, J. and Roediger, H. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2):151–162. 52

Karpicke, J. D. and Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, 319(5865):966–968. 52

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., and Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain Lang.*, 127(1):46–54. 19

Kerz, E., Wiechmann, D., and Riedel, F. B. (2017). IMPLICIT LEARNING IN THE CROWD: INVESTIGATING THE ROLE OF AWARENESS IN THE ACQUISITION OF L2 KNOWLEDGE. *Studies in Second Language Acquisition*, 39(4):711–734. 133, 178

Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, 22(2):154–169. 129

Kinder, A., Shanks, D. R., Cock, J., and Tunney, R. J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General*, 132(4):551. 104, 172

Klein, W. (1986). *Second Language Acquisition*. Cambridge University Press. 25

Knowlton, B. J. and Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(1):169–181. 10, 21

Kormos, J. (2009). *Speech Production and Second Language Acquisition*. New York : Routledge, 2009, c c2006., New York. 37, 38

Kowal, M. and Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language awareness*, 3(2):73–93. 38

Kowal, M. and Swain, M. (1997). 14 From semantic to syntactic processing How can we promote it in the immersion classroom? *Immersion education: International perspectives*, page 284. 38

Krashen, S. (1981a). *Second Language Acquisition and Second Language Learning*, volume 3. Pergamon, London. 5, 8

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The modern language journal*, 73(4):440–464. 43

Krashen, S. (1994). The input hypothesis and its rivals. In Ellis, N. C., editor, *Implicit and Explicit Learning of Languages*, pages 45–77. John Benjamins, Amsterdam. 6, 26, 43

Krashen, S. D. (1981b). The "fundamental pedagogical principle" in second language teaching. *Studia Linguistica*, 35(1-2):50–70. 2, 5, 26, 43

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition.* Language Teaching Methodology Series. Pergamon, Oxford ; New York, 1st ed edition. 2, 5, 26, 43

Krashen, S. D. (1985). *Second Language Acquisition and Second Language Learning.* Language Teaching Methodology Series. Pergamon Pr, Oxford, reprinted edition. OCLC: 256792515. 5, 8, 42

Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Kováříková, D., Jelínek, T., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondřička, P., and Zasina, A. (2015). SYN2015: representative corpus of written Czech. *https://wiki.korpus.cz/doku.php/cnk:syn2015.* 56

Kulinich, E., Royle, P., and Valois, D. (2019). On the inefficiency of negative feedback in russian morphology L1 acquisition. *First Lang.*, 39(5):547–570. 1

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(1):1–26. 66

Lany, J. and Gómez, R. L. (2008). Twelve-Month-Old Infants Benefit From Prior Experience in Statistical Learning. *Psychological Science*, 19(12):1247–1252. 19

Lany, J., Gómez, R. L., and Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, 31(3):481–507. 19

LaPierre, D. (1995). Language output in a cooperative learning setting: Determining its effects on second language learning. 38

Lee, J. and VanPatten, B. (2003). *MAKING COMMUNICATIVE LANGUAGE TEACHING HAPPEN.* McGraw-Hill Education. 40, 43

Lee, M.-W. and Williams, J. N. (1997). Why is short-term sentence recall verbatim? An evaluation of the role of lexical priming. *Memory & Cognition*, 25(2):156–172. 17

Leow, R. P. (2015). Toward a model of the L2 learning process in instructed SLA. *Explicit learning in the L2 classroom: A student-centered approach*, pages 236–250. 30

Leung, J. and Williams, J. N. (2006). Implicit Learning of Form-Meaning Connections. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 28(28). 22, 24, 51, 54, 55, 76

Leung, J. H. C. and Williams, J. N. (2011). THE IMPLICIT LEARNING OF MAPPINGS BETWEEN FORMS AND CONTEXTUALLY DERIVED MEANINGS. *Studies in Second Language Acquisition*, 33:33–55. 2, 22, 25, 51, 54, 76, 176

Leung, J. H. C. and Williams, J. N. (2012). Constraints on Implicit Learning of Grammatical Form-Meaning Connections. *Language Learning*, 62(2):634–662. 22, 24

Leung, J. H. C. and Williams, J. N. (2014). CROSSLINGUISTIC DIFFERENCES IN IMPLICIT LANGUAGE LEARNING. *Studies in Second Language Acquisition*, 36(04):733–755. 32, 98, 185

Levelt, W. J. (1993). Language use in normal speakers and its disorders. In Blanken, G., Dittmann, J., Grimm, H., Marshall, J. C., and Wallesch, C.-W., editors, *Linguistic Disorders and Pathologies, An International Handbook*, pages 1–15. De Gruyter Mouton, Berlin, Boston. 37

Levelt, W. J. (1995). The ability to speak: From intentions to spoken words. *European Review*, 3(1):13–23. 37

Levelt, W. J. (1999). Language production: A blueprint of the speaker. In Brown, C. and Hagoort, P., editors, *Neurocognition of Language*, pages 83–122. Oxford University Press, Oxford, UK. 37

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* MIT Press, Cambridge, Mass. 37

Lew-Williams, C. and Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4):447–464. 78, 171

Li, M. and DeKeyser, R. (2017). PERCEPTION PRACTICE, PRODUCTION PRACTICE, AND MUSICAL ABILITY IN L2 MANDARIN TONE-WORD LEARNING. *Studies in Second Language Acquisition*, 39(4):593–620. 2, 3, 42, 169

Lombardi, L. and Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, 31(6):713–733. 17

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4):1494–1502. 66

Lukšija, M. (2010). *Korpus Jako Zdroj Dat Při Prezentaci Předložek Do/Na s Místním Směrovỳm Vỳznamem ve Vỳuce Češtiny pro Cizince*. PhD Thesis, Masarykova univerzita, Filozofická fakulta. 50

Lupker, S. J. (1988). Picture naming: An investigation of the nature of categorical priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):444. 40

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. *Cognition and second language instruction*, pages 69–90. 32

MacWhinney, B., Bates, E., and Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23(2):127–150. 32

MacWhinney, B., Pléh, C., and Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive psychology*, 17(2):178–209. 12

Marsden, E. and Chen, H.-Y. (2011). The Roles of Structured Input Activities in Processing Instruction and the Kinds of Knowledge They Promote. *Language Learning*, 61(4):1058–1098. 34, 45

Marsden, E., Williams, J., and Liu, X. (2013). LEARNING NOVEL MORPHOLOGY: The Role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, 35(4):619–654. 22

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., and Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69(4):574–588. 78, 171

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., and Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6):1083–1100. 21, 36, 51, 182, 187

Mattys, S. L. and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78(2):91–121. 19

Mattys, S. L., Jusczyk, P. W., Luce, P. A., and Morgan, J. L. (1999). Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology*, 38(4):465–494. 19

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I Error and Power in Linear Mixed Models. *Journal of Memory and Language*, 94:305–315. 65

McClelland, J. L., McNaughton, B. L., and O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419. 131

McLaughlin, B. (1978). The monitor model: Some methodological considerations. *Language learning*, 28(2):309–332. 6

McLaughlin, B. (1987). *Theories of Second-Language Learning*, volume 81. Edward Arnold London. 9

McLaughlin, J., Osterhout, L., and Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature neuroscience*, 7(7):703. 15

Mirković, J. and Gaskell, M. G. (2016). Does Sleep Improve Your Grammar? Preferential Consolidation of Arbitrary Components of New Linguistic Knowledge. *PLOS ONE*, 11(4):e0152489. 132

Mitchell, D. B. and Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 14(2):213–222. 40

Miyake, A. and Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. *Foreign language learning: Psycholinguistic studies on training and retention*, pages 339–364. 129

Monaghan, P., Schoetensack, C., and Rebuschat, P. (2019). A single paradigm for implicit and statistical learning. *Top. Cogn. Sci.*, 11(3):536–554. 12, 35

Morgan, J. L., Meier, R. P., and Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive psychology*, 19(4):498–550. 19

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., and Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, 17(01):56–72. 7, 145, 174

Morgan-Short, K., Finger, I., Grey, S., and Ullman, M. T. (2012a). Second Language Processing Shows Increased Native-Like Neural Responses after Months of No Exposure. *PLOS ONE*, 7(3):e32974. 34, 46, 132, 174

Morgan-Short, K., Steinhauer, K., Sanz, C., and Ullman, M. T. (2012b). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of cognitive neuroscience*, 24(4):933–947. 15, 34, 35, 36, 46, 132

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2):127–162. 19

Norris, J. and Ortega, L. (2003). Defining and measuring SLA. *The handbook of second language acquisition*, pages 716–761. 16

Norris, J. M. and Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning*, 50(3):417–528. 26, 27, 28, 29, 30, 40

O'Reilly, R. C. and Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in cognitive sciences*, 6(12):505–510. 131

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., and Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, 21(6):509–521. 15

Oxford, R., Nyikos, M., and Ehrman, M. (1988). Vive la difference? reflections on sex differences in use of language learning strategies. *Foreign Language Annals*, 21(4):321–329. 183

Paciorek, A. (2012). *Implicit Learning of Semantic Preferences*. PhD thesis. 22, 87, 124

Paciorek, A. and Williams, J. N. (2015). Semantic generalization in implicit language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(4):989–1002. 2, 22, 24, 83, 176

Packard, M. G. (1999). Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. *Proceedings of the National Academy of Sciences*, 96(22):12881–12886. 35

Pacton, S. and Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1):80. 19

Pacton, S., Perruchet, P., Fayol, M., and Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of experimental psychology: General*, 130(3):401. 2, 19

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and second language acquisition. In Ellis, N. C., editor, *Implicit and Explicit Learning of Languages*. Academic Press, Academic, London ; San Diego, London. 6, 7, 8

Paradis, M. (2004). Neurolinguistics of bilingualism and the teaching of languages. In *Multimodality of Human Communication: Theories, Problems and Applications*, University of Toronto. 6, 7, 8, 10

Payne, J. D., Stickgold, R., Swanberg, K., and Kensinger, E. A. (2008). Sleep Preferentially Enhances Memory for Emotional Components of Scenes. *Psychological Science*, 19(8):781–788. 175

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203. 58

Perdue, C. (1993). *Adult Language Acquisition: Volume 2, the Results: Cross-Linguistic Perspectives*, volume 2. Cambridge University Press. 31

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? transitional probabilities versus chunks in language learning. *Top. Cogn. Sci.*, 11(3):520–535. 12, 20

Perruchet, P. and Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5):233–238. 20

Pickering, M. J., Branigan, H. P., Cleland, A. A., and Stewart, A. J. (2000). Activation of syntactic information during language production. *Journal of psycholinguistic research*, 29(2):205–216. 42

Pickering, M. J. and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4):329–347. 46

Poldrack, R. A. and Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, 41(3):245–251. 35

Potter, M. C. and Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29(6):633–654. 17, 53, 173

Potter, M. C. and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38(3):265–282. 17, 42

Radulescu, S. (2014). *Patterns Bit by Bit. An Entropy Model for Linguistic Generalizations*. Master's Thesis. 53, 128, 173, 174

Radulescu, S., Wijnen, F., and Avrutin, S. (2019). Patterns bit by bit. An Entropy Model for Rule Induction. 53, 128, 160, 173, 174

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6):855–863. 2, 10, 13, 14, 20, 21, 176, 180

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 2(1):88–94. 26, 27, 51

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of experimental psychology: General*, 118(3):219. 26, 51

Reber, A. S. (2015). Foreword. In Rebuschat, P., editor, *Implicit and Explicit Learning of Languages*. John Benjamins Publishing Company, Philadelphia, UNITED STATES. 12, 20

Rebuschat, P. (2008). Implicit learning of natural language syntax (Unpublished doctoral dissertation). *University of Cambridge*. 13, 20, 41

Rebuschat, P. (2013). Measuring Implicit and Explicit Knowledge in Second Language Research: Measuring Implicit and Explicit Knowledge. *Language Learning*, 63(3):595–626. 14, 179

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., and Ziegler, N. (2015). TRIANGULATING MEASURES OF AWARENESS: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, 37(2):299–334. 23

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., and Ziegler, N. (2013). Implicit and explicit knowledge of form-meaning connections: Evidence from subjective measures of awareness. *Noticing and second language acquisition: Studies in honor of Richard Schmidt*, 255:275. 23

Rebuschat, P. and Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33(4):829–856. 2, 13, 20, 41, 51

Ritchie, S. J., Bates, T. C., and Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology*, 51(5):573–582. 177

Robinson, P. (1996). Learning Simple and Complex Second Language Rules Under Implicit, Incidental, Rule-Search, and Instructed Conditions. *Studies in Second Language Acquisition*, 18(1):27–67. 26, 29

Robinson, P. (2003). Attention and memory during SLA. *The handbook of second language acquisition*, pages 631–678. 30

Robinson, P. (2005). Cognitive Abilities, Chunk-Strength, and Frequency Effects in Implicit Artificial Grammar and Incidental L2 Learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and Their Relevance for Sla. *Studies in Second Language Acquisition; New York*, 27(2):235–268. 21, 130, 133, 145, 173

Rodd, J., Gaskell, G., and Marslen-Wilson, W. (2002). Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language*, 46(2):245–266. 167

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., and Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, 40(7):1095–1108. 167

Rogers, J., Révész, A., and Rebuschat, P. (2016). Implicit and explicit knowledge of inflectional morphology. *Applied Psycholinguistics*, 37(4):781–812. 22

Ruiz, S., Tagarelli, K. M., and Rebuschat, P. (2018). Simultaneous acquisition of words and syntax: Effects of exposure condition and declarative memory. *Front. Psychol.*, 9:1168. 7, 40, 41

Rummer, R., Schweppe, J., and Martin, R. C. (2013). Two Modality Effects in Verbal Short-Term Memory: Evidence from Sentence Recall. *Journal of Cognitive Psychology (Hove, England)*, 25(3):231–247. 17

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81(2):149–169. 11, 19

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, 274(5294):1926–1928. 2, 11, 18, 20

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of memory and language*, 35(4):606–621. 19

Saksida, A., Langus, A., and Nespor, M. (2017). Co-occurrence statistics as a language-dependent cue for speech segmentation. *Dev. Sci.*, 20(3). 19

Saults, J. S. and Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, 136(4):663. 168

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., and Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 372(1711). 131

Schmidt, E., Pérez, A., Cilibrasi, L., and Tsimpli, I. (2019). PROSODY FACILITATES MEMORY RECALL IN L1 BUT NOT IN L2 IN HIGHLY PROFICIENT LISTENERS. *Studies in Second Language Acquisition*, pages 1–16. 173

Schmidt, R. (1992a). Awareness and second language acquisition. *Annual review of applied linguistics*, 13:206–226. 30

Schmidt, R. (1992b). Psychological mechanisms underlying second language fluency. *Studies in second language acquisition*, 14(4):357–385. 30, 40, 43

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in second language learning*, 11:237–326. 10

Schmidt, R. (2001). Attention. In Robinson, P., editor, *Cognition and Second Language Instruction*, pages 3–32. Cambridge University Press, Cambridge. 31

Schmidt, R. and Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. *Talking to learn: Conversation in second language acquisition*, pages 237–326. 38

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied linguistics*, 11(2):129–158. 30

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in second language acquisition*, 15(2):147–163. 42

Schweppe, J., Barth, S., Ketzer-Nöltge, A., and Rummer, R. (2015). Does verbatim sentence recall underestimate the language competence of near-native speakers? *Frontiers in Psychology*, 6. 17

Schweppe, J., Rummer, R., and Fürstenberg, A. (2009). Beyond sentence boundaries: Grammatical gender information in short-term recall of texts. *Memory & Cognition*, 37(1):73–80. 17

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., and Hagoort, P. (2012). Shared syntax in language production and language comprehension–an FMRI study. *Cereb. Cortex*, 22(7):1662–1670. 42

Segalowitz, N. (2003). Automaticity and Second Languages. In Doughty, C. J. and Long, M. H., editors, *The Handbook of Second Language Acquisition*, pages 382–408. 6

Shanks, D. R. (1995). *The Psychology of Associative Learning*, volume 13. Cambridge University Press. 12

Shanks, D. R. and St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17(3):367–395. 14, 179

Skehan, P. (2015). 11. Working Memory and Second Language Performance: A Commentary. In Wen, Z. E., Borges Mota, M., and McNeill, A., editors, *Working Memory in Second Language Acquisition and Processing*, pages 189–202. Multilingual Matters, Bristol, Blue Ridge Summit. 130

Slobin, D. I. and Welsh, C. A. (1973). Elicited Imitation as a Research Tool in Developmental Psycholinguistics. In Ferguson, C. and Slobin, D. I., editors, *Readings in Child Language Acquisition*. New York: Holt, Rinehart and Winston. 16, 17, 173

Spada, N., Shiu, J. L.-J., and Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, 65(3):723–751. 17, 18

Spada, N. and Tomita, Y. (2010). Interactions Between Type of Instruction and Type of Language Feature: A Meta-Analysis. *Language Learning*, 60(2):263–308. 26, 28, 29, 30, 40

Speciale, G., Ellis, N. C., and Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied psycholinguistics*, 25(2):293–321. 129

Suzuki, W. L. and Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *Journal of comparative neurology*, 350(4):497–533. 6

Suzuki, Y. and DeKeyser, R. (2015). Comparing Elicited Imitation and Word Monitoring as Measures of Implicit Knowledge. *Language Learning*, 65(4):860–895. 9, 18

Suzuki, Y. and DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Lang. Learn.*, 67(4):747–790. 8, 9

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in second language acquisition*, 15:165–179. 38, 40

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian modern language review*, 50(1):158–164. 38, 40

Swain, M. (2005). The output hypothesis: Theory and research. In *Handbook of Research in Second Language Teaching and Learning*, pages 495–508. Routledge. 38

Swain, M. and Lapkin, S. (1995). Problems in Output and the Cognitive Processes They Generate: A Step Towards Second Language Learning. *Applied Linguistics*, 16(3):371–391. 38

Tagarelli, K. M., Mota, M. B., and Rebuschat, P. (2011). The Role of Working Memory in Implicit and Explicit Language Learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33(33). 130

Tamminen, J., Davis, M. H., Merkx, M., and Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, 125(1):107–112. 129, 132, 174, 175, 180, 184

Tamminen, J., Davis, M. H., and Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, 79:1–39. 53, 83, 96, 129, 132, 174, 175, 180, 184

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217):1632–1634. 78

Thompson, S. P. and Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language learning and development*, 3(1):1–42. 19

Tokowicz, N. and MacWhinney, B. (2005). IMPLICIT AND EXPLICIT MEASURES OF SENSITIVITY TO VIOLATIONS IN SECOND LANGUAGE GRAMMAR: An Event-Related Potential Investigation. *Studies in Second Language Acquisition*, 27(2):173–204. 15, 16

Tokowicz, N. and Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, 22(7):1092–1106. 16

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4(02). 6, 7, 23

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In Sanz, C., editor, *Mind and Context in Adult Second Language Acquisition*, pages 141–178. Georgetown University Press, Washington DC. 6, 7, 23, 31

Van den Noort, M. W., Bosch, P., and Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, 11(4):289–296. 129

van Schagen, J. (2017). *The Effect of Overloading the Memory Capacity on Rule Learning*. Master's Thesis. 160, 174

VanPatten, B. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition*. Greenwood Publishing Group. 33

VanPatten, B. (2002). Processing Instruction: An Update. *Language Learning*, 52(4):755–803. 33

VanPatten, B. and Cadierno, T. (1993). Explicit instruction and input processing. *Studies in second language acquisition*, 15(2):225–243. 33, 42, 43, 44, 186, 187

VanPatten, B., Collopy, E., Price, J. E., Borst, S., and Qualin, A. (2013). Explicit Information, Grammatical Sensitivity, and the First-Noun Principle: A Cross-Linguistic Study in Processing Instruction. *The Modern Language Journal*, 97(2):506–527. 29, 30, 34

VanPatten, B. and Oikkenon, S. (1996). Explanation versus Structured Input in Processing Instruction. *Studies in Second Language Acquisition*, 18(4):495–510. 33, 34, 46, 48

VanPatten, B. and Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. *Second language acquisition theory and pedagogy*, pages 169–185. 33, 34, 43, 44, 46, 187

VanPatten, B. and Uludag, O. (2011). Transfer of training and processing instruction: From input to output. *System*, 39(1):44–53. 44, 46

VanPatten, B. and Wong, W. (2004). Processing instruction and the French causative: Another replication. *Processing instruction: Theory, research, and commentary*, pages 97–118. 43

Vinther, T. (2002). Elicited imitation: A brief overview. *International journal of applied linguistics*, 12(1):54–73. 17

Wajnryb, R. and Maley, A. (1990). *Grammar Dictation*, volume 3. Oxford University Press. 38

Walk, A. M. and Conway, C. M. (2015). Implicit statistical learning and language acquisition. In Rebuschat, P., editor, *Implicit and Explicit Learning* *of Languages.* John Benjamins Publishing Company, Philadelphia, UNITED STATES. 12

Warker, J. A. and Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2):387. 19

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., and Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37(1):159–170. 7

Weber-Fox, C. M. and Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of cognitive neuroscience*, 8(3):231–256. 7

White, L. (2003). *Second Language Acquisition and Universal Grammar.* Cambridge University Press. 13

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., and Born, J. (2011). Sleep Selectively Enhances Memory Expected to Be of Future Relevance. *Journal of Neuroscience*, 31(5):1563–1569. 176

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, 27:269–304. 2, 14, 15, 22, 23, 24, 175, 176, 177, 178

Williams, J. N. (2009). Implicit learning in second language acquisition. In *The New Handbook of Second Language Acquisition*, pages 319–353. 5, 7, 9, 10, 11, 12

Williams, J. N. (2011). *Working Memory and SLA*. Routledge Handbooks Online. 129

Williams, J. N. and Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, 118(4):522–553. 20

Williams, J. N. and Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, 53(1):67–121. 129

Williams, J. N. and Rebuschat, P. (2012). Statistical learning and syntax: What can be learned and what difference does meaning make. *Statistical learning and language acquisition*, pages 237–264. 20

Williams, J. N., VanPatten, B., Williams, J., Rott, S., and Overstreet, M. (2004). Implicit learning of form-meaning connections. *Form-meaning connections in second language acquisition*, pages 203–218. 22

Winocur, G. and Moscovitch, M. (2011). Memory Transformation and Systems Consolidation. *Journal of the International Neuropsychological Society*, 17(5):766–780. 131

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, 65(1):1–14. 40

Wonnacott, E., Boyd, J. K., Thomson, J., and Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5year olds. *J. Mem. Lang.*, 66(3):458–478. 1

Wonnacott, E., Brown, H., and Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, 95(Supplement C):36–48. 40, 186

Wonnacott, E., Newport, E. L., and Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, 56(3):165–209. 186

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3485–3492. IEEE. 58

Zaromb, F. M., Karpicke, J. D., and Roediger, H. L. (2010). Comprehension as a basis for metacognitive judgments: Effects of effort after meaning on recall and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2):552–557. 52

Zaromb, F. M. and Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8):995–1008. 52

Zhang, R. (2015). MEASURING UNIVERSITY-LEVEL L2 LEARNERS'IMPLICIT AND EXPLICIT LINGUISTIC KNOWLEDGE. *Studies in Second Language Acquisition*, 37(3):457–486. 18