Statistical and explicit learning of graphotactic patterns with no phonological counterpart: Evidence from artificial lexicon studies with 6– to 7-year-olds and adults

Felicia Daniela Singh

A thesis submitted for the degree of

Doctor of Philosophy

Department of Language & Cognition

Division of Psychology and Language Sciences

University College London

2021

Declaration

I, Felicia Daniela Singh, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude towards my supervisor, Dr. Elizabeth Wonnacott. Your endless enthusiasm for the science of language learning and commitment to research integrity have truly inspired me. I am also very grateful for the many comments and restructuring suggestions to early drafts of this thesis—they have made it so much better and clearer—I take sole responsibility for all the inconsistencies and mistakes that managed to sneak through these pages. My first thanks also go to my supervisor and collaborator, Dr. Anna Samara, whose name is peppered throughout this thesis. Thank you for letting me take your work and turn it into my own. I cannot express how much it meant for me all your advice, pep talks, and detailed explanations of concepts around statistical learning and spelling. I would also like to thank my secondary supervisor, Professor Courtenay Norbury, for taking me under her wing: Thank you for always being supportive, helping us, early career researchers, navigate the academic world and often cheering us up with a slice of cake, a picnic or a visit to the pub.

I am eternally grateful to all the headteachers and teachers who allowed me to disrupt their busy schedule and take the children out of their classes, breaks or assemblies to play computer games and teach them alien languages. Thank you to all the children who participated and for their funny stories, and to all adult participants who were willing to humor me and play the alien language games. I owe my thanks to project students Tarisa Tan and Aslam Norhan for helping with some of the data collection.

Special thanks to Dr. Merle Mahon for always checking how I was doing and, along with Richard Jardine, making sure I had everything I needed for a smooth PhD experience. I was very fortunate to be part of the Language Learning Lab with Maša Vujović, Gwen Brekelmans, Hanyu Dong, Cat Silvey, Anna Samara and Liz Wonnacott: Thank you all for listening to my many experiment ideas and helping me shape them before presenting them at conferences and ultimately here, in this thesis. The PhD experience would have been frightening without my fellow 2016 starters: Jo Saul, Claire Murphy and Maša Vujović: Our regular catch-ups helped me feel I wasn't the only one struggling, at times, with how to structure my days. Thank you to Gwen Brekelmans for listening to me when I swiveled around to ask all sorts of questions, many about R script and the annoying "invisible" mistakes; and thank you for sharing your script with me, a starting point for my own. Thank you, Jo Saul, for lending your voice and Southern British accent to some of my stimuli. I really enjoyed my time in Room 202 and am fortunate to have known every single one of you who resided with me there, especially Vanessa Meitanis, Lena Blott, Claudia Bruns, Gwen Brekelmans, Maša Vujović, Anna Volkmer, Claire Murphy, and Jo Saul: The small breaks made my days so much easier and fun. I now reflect on those times, in light of the lockdown we all find ourselves in, and appreciate how important those connections were, whether work related chats in the hallways when moving from one seminar to another, or informal catch-ups during coffee breaks. They have all contributed to this work.

As the African proverb says, "It takes a village to raise a child", and I would like to extend my sincerest gratitude to everyone in my “village” who made it possible for me to complete this work and raise two beautiful children, particularly our “extended family” and childcare bubble, Karla, Ben, Zachary and Autumn, who have provided a home to Aryan and Arjun when I was rushing to schools to test children, running late from testing, and during the last few months of writing this thesis. A huge thank you to Mikaela for taking the children to the park and giving me a few extra hours to write and for always asking how I was doing. Thank you to all my friends who kept asking when and *if* I will ever be done with this PhD: They pushed me to stay focused and I am now hopefully closer to giving them an answer.

And finally, to my family: I am genuinely overwhelmed by the unconditional love and support I received from each and every one of you. My husband, Bhavneet Singh: you have pushed me to be my best—and then more, and are a true example to me and our children of what it means to be passionate about the work you do. Through my hard work, I seek not only to follow your example, but to be an example to our boys, who will see their mother juggle family and work life, and who is successful thanks to your support. Aryan and Arjun: thank you for all the cuddles and understanding when I wasn’t able to watch a movie with you or play a game. You have tried all my experimemt versions so thank you for being my pilot participants. My adopted parents, JP and Bubble Singh: thank you for your passionate support and encouragement. And thank you to my parents, Emil and Lia Comes and my brother, Ovidiu Comes: your unconditional love keeps me going. My mother, an educator herself, has inspired my fascination with how children learn and what a difference a dedicated educator can make in their pupil’s life.

Mama şi Tata: *Ştiu că nu e nevoie să vă mulţumesc, pentru că iubirea voastra şi suportul sunt necondiţionate. Totuşi, vă multumesc din suflet pentru ajutorul pe care mi l-aţi oferit în orice situaţie, pentru exemplul care sunteţi pentru mine, şi pentru că ne-aţi învăţat să fim autentici şi să apreciem munca.*

For Aryan and Arjun

“Înţelepciunea nu se împrumută cu carul ci se câstigă cu bobul”

Nicolae Iorga

My English translation: *Wisdom is not borrowed by the cartload but must be earned with each grain.*

Abstract

Children are powerful statistical spellers: They can learn novel written patterns with phonological counterparts under experimental conditions, via implicit learning processes, akin to “statistical learning” processes established for spoken language acquisition. Can these mechanisms fully account for children’s knowledge of written patterns? How does this ability relate to literacy measures? How does it compare to explicit learning?

This thesis addresses these questions in a series of artificial lexicon experiments, inducing graphotactic learning under incidental and explicit conditions, and comparing it with measures of literacy. The first experiment adapted an existing design (Samara & Caravolas, 2014), with the goal of searching for stronger effects. Subsequent experiments address a further limitation: Previous studies assessed learning of spelling rules which have counterparts in spoken language; however, while this is also the case for some naturalistic spelling rules (e.g., English phonotactics prohibit word initial /ŋ/ and accordingly, written words cannot begin with *ng*), there are also purely visual constraints (graphotactics) (e.g., *gz* is an illegal spelling of a frequent word-final sound combination in English: **bagz*). Can children learn patterns unconfounded from correlated phonotactics? In further experiments, developing and skilled spellers were exposed to patterns replete of phonotactic cues. In post-tests, participants generalized over both positional constraints embedded in semiartificial strings, and contextual constraints created using homophonic non-word stimuli. This was demonstrated following passive exposure and even under meaningful (word learning) conditions, and success in learning graphotactics was not hindered by learning word meanings. However, the effect sizes across this thesis remained small, and the hypothesized positive associations between learning performance under incidental conditions and literacy measures were never observed. This relationship was only found under explicit conditions, when pattern generalization benefited. Investigation of age effects revealed that adults and children show similar patterns of learning but adults learn faster from matched text.

Impact Statement

The role of statistical learning in language acquisition has been investigated from as early as a century ago; therefore, a large body of evidence suggests that humans are able to detect and extract the distributional probabilities from spoken language input. Spelling is a vital, yet understudied part of literacy development, and sensitivity to patterns in the written language has been shown to emerge early in development. These patterns are often untaught; however, current theories of spelling development do not account for how they are learned. Recent studies have demonstrated that sensitivity to spelling rules can emerge following brief incidental exposure to novel pattern-embedding stimuli, and this knowledge can generalize to novel instances. However, the visual features of written words used in these studies have been confounded by correlated phonological cues. Is learning possible when these do not have phonotactic counterparts? While early theories of spelling development postulated that phonology is an integral part of learning to spell, this thesis reviews extensive literature that shows that written language has evolved distinct features from the spoken language it represents, including spelling rules that can only be learned on a visual basis. Therefore, it is important to account for learning of purely visual linguistic patterns.

This thesis draws from two distinct research areas: visual statistical learning and spelling development; and extends a language-learning paradigm devised by Samara and colleagues (2014, 2019). The results from various design adaptations, together with innovative methods of statistical inference, provide substantial evidence that incidental visual graphotactic learning occurs independent of phonology. For the first time, these processes are shown to support relatively beginning spellers' learning. This thesis also provides a first direct demonstration that, although implicit learning is possible, explicit instruction benefits learning generalization of spelling rules in children, and it was only under explicit conditions that correlations with measures of literacy attainment were found.

The impact of this thesis concerns academic research and there are also direct implications for spelling instruction. This thesis has shown how Bayesian statistics can be used to explore learning effects and correlations between measures, and to differentiate between where there is evidence for null effects, versus ambiguous evidence. In addition, a novel paradigm was developed to probe implicit and explicit learning in different populations. Ongoing research could use this paradigm to look at how typically developing and dyslexic populations differ in their statistical learning abilities. This thesis goes beyond current research in statistical learning, to consider the relevance of explicit instruction—widely considered to be the primary form of access to literacy. It also contributes to the theories of literacy by demonstrating developing spellers' sensitivity to purely visual patterns. While current results show that early incidental exposure to print should be encouraged, they also show that explicit teaching is altogether more effective. Future work could look at how best to combine implicit learning and explicit instruction. By building from this work, it is possible to develop principles that can inform educational programs that aim to teach actual rules, to supplement classroom teaching.

Contents

List of Figures

List of Tables

1 Literature Review

Language and literacy skills form the basis of learning in an educational system, along with mathematical skills. However, talking, reading, and writing begin much earlier than the start of schooling. Learning to write, the topic of the current thesis, can be a long process and, given the complexity of the writing systems that evolved, it may require lifelong learning. By following the conventions of a writing system, learners can produce correct spellings and this is important: It enables the writer to record and communicate ideas that can be shared with others. An ability to spell words correctly is not guaranteed to emerge from exposure to written text (Ehri & Roberts, 1979; Bosman & Van Orden, 1997) and is a skill more difficult to learn (Treiman, 1998) than reading. Yet spelling development is relatively understudied in comparison to reading. While reading and spelling are correlated, spelling requires skills that are independent of reading (Shanahan, 1984) and difficulties tend to be more enduring (Caravolas, Hulme, & Snowling, 2001). A key difference lies in the fact that spelling requires the production (or encoding) of strings from memory (such as letters in the alphabetic system, or logograms in Chinese) while reading requires decoding print (Ehri, 1997). Therefore, it is important to understand how spelling-specific learning processes emerge and develop. The process of learning to spell does not stop when children have mastered the spellings of words that exist in their spoken language vocabulary (also referred to as the conventional stage; Frith, 1985). With increased exposure to their language, children and adults encounter new words and, in order to produce correct and conventional spellings for novel vocabulary, they must develop sensitivity to how their writing system is structured (Kessler, 2009). There is increasing evidence that spelling patterns (both deterministic and probabilistic) are picked up implicitly, that is, without direct instruction or feedback, through statistical learning mechanisms (Treiman, 2018). The goal of the current thesis is to further explore and establish the extent to which implicit statistical learning processes, which have been widely explored in the context of learning of spoken language, may also be relevant to learning to spell.

With this goal in mind, the first part of this chapter (1.1) considers the evolution of written systems, and the relationship with spoken language. The evidence presented reveals that various pressures led to the evolution of writing systems which are *not* necessarily a direct reflection of spoken language, but contain their own variation and complexity. The second section (1.2) considers the relationship between these complex systems and the cognitive processes that support them. The third section (1.3) presents the key aspects of writing systems which are of interest in the current thesis: graphotactic (spelling) rules. Evidence for readers' knowledge of these untaught, probabilistic constraints on possible spellings comes from both experimental and developmental studies, and the latter is presented in Section 1.4. Such evidence provides the basis for the aim of the current thesis: to explore whether these constraints can be learned via implicit statistical learning processes. To that end, the fifth section (1.5) reviews the literature on statistical learning in language and vision more broadly. A sixth section (1.6) reviews literature that explored statistical learning specifically in the context of learning graphotactic constraints. The experiments presented in this thesis follow up on this type of work, further probing the ability of adults and children to learn different types of graphotactic patterns. Another question addressed in the current thesis is the extent to which the types of experimental measures relate to literacy measures. The seventh section (1.7), therefore, reviews relevant literature and highlights the methodological difficulties that make this link elusive. Finally, although a central claim in this thesis is that implicit statistical learning may underpin learning of spelling, it is nevertheless possible that explicit instruction is beneficial. For this reason, in several of the studies reported in this thesis, an explicit condition was included as a comparison to the implicit learning condition. Therefore, Section 1.8 provides a brief review of relevant literature on the role of explicit teaching in spelling.

1.1 Evolution of written language

Written language is an expression of spoken language but this relationship is far from straightforward and has been studied extensively (Ludwig, 1983; Olson, 1977, 1994; for a review, see Perfetti & Harris, 2013). Unlike spoken language, which is acquired at an impressive rate and seemingly effortlessly by the infant brain (for a review, see Kuhl, 2011), written language is a skill that is learned at a slower pace. As Pinker (1994) explains in his book, *The Language Instinct*, human beings have “the instinct to learn, speak, and understand language” (Pinker, 1994, p. 15). Human beings not only have an innate and spontaneous tendency to speak but also an ability to subsequently apply the knowledge acquired instinctively to other areas of cognition and adapt behavior (i.e., developmental flexibility). In contrast, written language is a recent cultural innovation that lags historically behind spoken language. While all human beings speak or sign, not all read and write, indicating that there is no dedicated instinct, but instead learners must employ cognitive mechanisms that developed for other purposes (DeHaene, 2009). Hence, the distinction between the terms *acquisition* when talking about spoken language and *learning* when talking about written language. This section provides a brief account of the evolution of written language, to demonstrate how different writing systems evolved to provide a solution between the trade-off of representing different aspects of spoken language, and how they may also be shaped by chance events. The result is that written languages may show complex patterning and variation which is systematically distinct from the spoken language counterpart (Coulmas, Ehlich, & Winter, 1983). This provides motivation for the current work in that it explains the existence of purely visual patterns that humans must pick up when learning to spell, without relying on their knowledge of spoken language.

1.1.1 The birth of written language.

At first, writing took the form of a functional system consisting of pictures that represented meaning (pictographs). This system emerged from a need for trade and mass production (cuneiform writing), to represent royal iconography (hieroglyphic writing), and mark calendars (glyphic writing), among many practices. With the emergence of Sumerian Cuneiforms, the meaning-based graphs developed to become more abstract and, with the Egyptian hieroglyphic, which was the first writing system to have phonetic values, writing became “visible speech”, that is, spoken language in visual form (DeFrancis, 1989).

The first written texts in writing (in alphabetic scripts) replicated the spoken word in a transparent way through *scriptura continua*, typically read aloud by expert orators, where there was no punctuation or marks to indicate spaces between words or sentences, and each distinct unit of writing (grapheme) represented a spoken language unit (phoneme), such as Classical Greek and Latin. With the emergence of silent reading around the 7th century (Saenger, 1997; see also Rastle, 2019), graphic conventions were introduced (e.g., word spacing, punctuation, and layout) and written text diverged from the close relationship with fluent speech. It is hard to imagine that mass literacy, that is, teaching people to read and write, would have been possible without a transition to an accessible form of encoding language. From being restricted to elites and clergy, propelled by the introduction of paper in the 11th century, literacy has become a human right and an essential skill needed for lifelong learning.

A first indication that written language evolved distinct properties to represent spoken language comes from the distinction between writing and written language. Ludwig (1983) pointed out that, in human evolution, writing, the instrument of written language, precedes written language, but had an indirect impact on its development (Ludwig, 1983). Factors such as the process of learning, the impersonal environment in which it is used, as well as the structure of- and between- the written units played a more important role. Due to their abstract property, graphemes were believed to lack the properties that can be associated with phonemes in spoken language, such as intonation, accents, tone or prosody. However, certain elements of prosody and intonation can be reflected when larger written utterances are combined, such as variation in word order and punctuation. Nevertheless, many of the prosodic and intonational cues that help provide word and sentence meaning in spoken language are absent in written language (e.g., Grosjean & Gee, 1987; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001), as well as the audiovisual context such as lip-movement (e.g., McGurk & MacDonald, 1976) and gestures (e.g., Alibali & Kita, 2010). On the other hand, just as some features of spoken language were not overtly recorded in writing, the visual medium of writing evolved to provide discrete, separable and segmented information about language that did not translate directly into spoken language (such as word and sentence boundaries; Ludwig, 1983). With this divergence, written speech became written language and earned its own autonomous place in the overall language system.

Most research exploring the divergence between spoken and written language focused on the consistency of the mappings between individual—or group of graphemes—and spoken units or phonemes (i.e., orthographic depth; e.g., Frost, 2008; Katz & Feldman, 1983; Katz & Frost, 1992; Ziegler & Goswami, 2005). In *alphabetic* systems, for example, used by most European languages and some Asian (e.g., Korean)—where letters (i.e., visually simple symbols) represent phonemes—the extent to which a letter represents multiple sounds and, conversely, multiple letter combinations represent a sound (or speech unit, such as syllable), determines the consistency of a language—also referred to as transparency or depth. As a result, alphabetic languages are placed on a continuum, from consistent, transparent or shallow orthographies such as Finnish, Turkish, Serbo-Croatian, and most Romance languages: Spanish, Italian, Portuguese and Romanian, to inconsistent, opaque or deep orthographies, such as English. A language like English, for example, uses more than one letter to represent a sound (e.g., the diagraph *sh* represents the sound /ʃ/) and a sound can be represented by more than one letter (e.g., the sound /ə/ can be represented by *o* as in *gallop*, *u* as in *minus*, or *e* as in *planet*). Share (2018; see also Daniels & Share, 2018) points out that the phoneme-based approach explains inconsistencies found in alphabetic orthographies, but does not account for the variation in writing systems that do not have such orthography. For example, in the *abjad* systems, graphemes only represent the consonants (e.g., Arabic); in the *abugidas* system the consonants have a dominant status over vowels (e.g., Devanagari), and in *morphosyllabaries*, graphemes correspond to syllabic morphemes (meaningful internal word units; e.g., Chinese) or syllables (e.g., Japanese) in the spoken language. Therefore, in those systems, the degree of divergence between spoken and written language must be explained by factors other than phonology, such as morphological and semantic information (discussed in Section 1.1.2). Share (2018) explains that, while each written language is the expression of its corresponding spoken language, there are systematic differences between certain aspects of "written text and spoken vernacular" and "all scripts are, to some degree diglossic" (Share, 2018, p. 441). This means that, regardless of the degree of divergence, even native speakers need to learn new forms of their language through reading. In the more extreme instances of diglossia, the spoken language is used in informal, conversational context and does not have a written form, but instead, a distinct, formal and complex form is used in writing (e.g., vernacular Arabic/Standard Arabic). In the mildest instances, the formal written language utilizes forms less frequently used in the spoken counterpart, such as the use of more formal words (e.g., *discontinue* vs. *stop*), or the use of more complex syntactic structures. While the resulting inconsistencies weaken the link between spoken and written language, they provide additional support and enable rapid analysis of text (see also Rastle, 2019).

It has been argued that the enormous diversity among writing systems can be explained by a matched diversity in spoken languages (Halliday, 1977). However, as demonstrated above, the written and spoken languages are not perfectly matched and their divergence can explain much of this diversity among scripts. Furthermore, entire writing systems have been imposed on many cultures where there was no opportunity for a matched system to evolve. For example, although only around 500 African languages have their own written form, many of these use a Roman-based script brought into the culture by European missionaries (Share, 2012). As a third of the 7,000 or so (Ethnologue, 2020) living languages in the world are in Africa, many more remain without a written form (Bendor-Samuel, 1996).

Given that written language is not merely secondary to- and derived from- spoken language, the next subsections describe how the writing systems evolved to provide a trade-off of representing different aspects of language, and how chance events may have shaped them.

1.1.2 The "Grapholinguistic Equilibrium"[1]

As concluded above, written languages are not necessarily optimally matched with the spoken language they represent. Hence, when learning to read and write, it is not enough to learn the visual arrangement of graphemes and how they map on to sounds (orthographic characteristics; Goswami, 2012). Instead, the full linguistic environment is needed to form representations of the writing system. The visual objects (written units) also map on to meaning and are structured in interrelated units with grammatical and lexical function (morphology). Frost (2012) suggested that written language adjusted to provide unlimited information about meaning (semantics) and sounds (phonological information) with limited (even minimal) visual structure. This whole-language approach was a departure from an earlier belief that written language is an expression of spoken language and, as such, it is centered solely around phonology and not meaning (Perfetti, 2011; Perfetti, Zhang, & Berent, 1992). Seidenberg (2011) suggested that a trade-off between the complexity of orthography (i.e., whether an orthography is deep or shallow) and spoken

[1] Coined by Seidenberg (2011)

language complexity (i.e., whether a spoken unit contains complex inflections through morphemes) ensures that a “grapholinguistic equilibrium” is maintained so that comprehension is supported by any writing system (Harm & Seidenberg, 2004; see also Perfetti & Harris, 2013). Such approach explains that writing systems made use of this trade-off to achieve an efficient writing-language fit. For example, consistent orthographies (see 1.1.1 for examples) can support the use of complex inflections through morphemes (meaningful internal word units) through agglutinative processes. Morphemes are joined together in word strings with minimal changes to phonology giving the language its morphological transparency. In Finnish, for example, long words such as *Juoksentelisinkohan* (translated: “wondered if you should run around aimlessly?”) are not unusual. Morphemes are strung together around the word stem, such as *sydän* (heart) to form *sydämeni* (my heart), without changing the phonology. On the other hand, languages with inconsistent spelling-sound mappings do not accommodate complex morphological inflections but instead have irregularities packed in shorter and frequent units. In English, for example, inflectional and derivational morphology is preserved at the cost of phonological transparency. This means that a single morpheme can be expressed with different phonemes, such as the past tense *ed*, expressed with /t/ in *jumped*; /d/ in *robbed*; and /Id/ in *added*. Similarly, the morphological information from the word *heal* and *debit*, remains visible, while the pronunciation varies in words derived from them, such as *health* and *debt* (where the letter *b* is retained but not pronounced). Hebrew and Arabic go even further in conveying morphological information, with large part of phonological information missing (Frost, 1994, 1995; Perfetti & Harris, 2013). Here, morphemes take the form of consonant skeletons, such as the root morpheme KTB expressing the idea of writing, and vowels are not expressed in writing but instead, the context is used to indicate where and what vowel sounds are needed to form words (e.g., in Arabic ***k**a**t**a**b**a* means ‘he wrote’, *ya**kt**u**b**u* means he writes, and *ma**kt**a**b**a* means library). Finally, Chinese writing does not even encourage activating phonology but instead maps directly to meaning.

The inconsistencies found in the diverse writing systems described above tell us that the clue to how written languages become effective does not lie in the appearance or size of the spoken language chunk it represents but in how they map structurally to the spoken language. Hence, despite the intimate relationship between orthography and phonology, especially in the alphabetic systems, it is widely agreed that written language does not merely transcribe the entire speech.

1.1.3 Arbitrary influences

As evidenced in the previous sections (1.1.1 and 1.1.2), writing systems have not evolved solely to align with the phonological structure of spoken language, but instead, their development includes other linguistic factors such as morphology and semantics. However, the optimization process described above cannot explain all inconsistencies. This is partly because writing has a more conservative character than speech and thus, a process of optimization cannot support the rapid adaptation required to match the changes in spoken language. That is, sound changes happen at a quicker pace than written language and, while a writing system may start with a grapheme for each phoneme, if pronunciations change, written language will not catch up, resulting in inconsistencies. In Spanish, for example, there used to be two distinct phonemes /b/ and /v/, spelled regularly as *b* and *v* and, while the sounds merged into one phoneme, the spellings remained the same, as in *base* (meaning base), pronounced as /ba-se/ and *vaso* (meaning glass), pronounced as /ba-so/ (see also Treiman & Kessler, 2013). In addition, chance events outside the linguistic environment may play a role in the evolution of particular writing systems (Seidenberg, 2012). Such events—political, religious, and ideological—documented throughout the history and in all corners of the world, have interacted with the structure of language to shape the writing systems we use today. For example, Turkish and Romanian scripts changed from Arabic and Cyrillic, respectively, to Latin in the 19^{th} century, as a result of national ideology (Ghetie, 1978). Other scripts went through reform, such as the Hangul, created by Korea's King Sejong the Great to replace the idu system (based on Chinese characters) in 1443; and the Chinese script was reformed in 1949 by simplifying the Traditional Chinese characters in order to boost literacy. Such long, progressive and complex influences are arbitrary and reflected in the many inconsistencies we find in orthographies. Seidenberg offered an apt analogy: "Accidents of geography and history are to writing systems as mutation, migration, and genetic drift are to evolution" (Seidenberg, 2012, p. 306).

In sum, written language, far from transcribing spoken language (Frost, 1998), follows an adaptation process to spoken language and incorporates chance historical and cultural influences to represent the full linguistic environment, with an overarching goal: to convey meaning. The evolution of written language was not planned and linear but it has rather diverged on multiple paths, resulting in a variety of systems available for speakers of many human languages. Importantly, however, these systems need to be such that they can be learned and processed by the human brain. Thus, they are also shaped by the human cognition. Conversely, due to its objectified and permanent character, written language has the ability to shape learning by providing insights into aspects of language that are not available from the spoken language, such as elements of morphology that were preserved in the written form that would have otherwise been lost (see examples in Section 1.1.2). The next section, therefore, considers the relationship between literacy and cognition and aims to explain the cognitive processes that make it possible for humans to adapt to the variation.

1.2 The relationship between literacy and human cognition

1.2.1 Honoring the human cognitive system

There have been recent notable efforts to define a universal model of reading (Frost, 2012; Share, 2018), motivated by the universal attachment of written language to the same linguistic system used by spoken language (Coulmas et al., 1983; Perfetti & Harris, 2013). This view posits that writing and speech engage the same linguistic system, with differences in the modality of perception and production. Such conclusion has historically placed written language on a subservient level (Bolinger, 1946), with Mattingly (1972) famously proposing that reading is a “secondary language-based skill” (p. 142), “parasitical on spoken language” (p. 145), requiring linguistic awareness.

However, as noted in Section 1.1.1, there is huge variation in written language and how it relates to spoken language and this constrains how written language will be learned and processed, and the cognitive system must be attuned to this variation (e.g., Samuelsson et al., 2005). The field of cognitive sciences aims to identify the abilities needed for the learners to master their orthography in order to get to the linguistic meaning, as well as the functional constraints written languages satisfy. A two-way relationship is apparent: The writing systems must be tuned in to the human cognitive abilities and in turn, the cognitive system is tuned to pick up the information available in the writing system (Perfetti & Harris, 2013).

As noted in Section 1.1.1, unlike spoken/signed language, written language is not universal and is a human “invention”, thus the human brain is unlikely to have evolved specific functions dedicated for literacy (Gough & Hillinger, 1980; Pereaa & Carreiras, 2012). Instead, the human brain must utilize mechanisms dedicated for other tasks. For example, at a perceptual processing level, the brain areas that are engaged in identifying graphemes are the same as the ones activated for recognizing visual objects and faces (DeHaene & Cohen, 2011) and humans are already experts at the object recognition before they start to read. Unlike the three-dimensional world of faces, however, graphemes are two-dimensional objects and the visual processes needed to adapt to this feature. As a result, the units of writing such as words are perceived through a narrow “moving-window” of vision, so that the eye movements saccade from left to right to take in the words as objects when processing them (Rayner, 1998).

Research also focusses on how other aspects of human cognition constraint the cognitive processes involved in reading and writing. For example, the anatomical properties of the brain as well as the functions available for processing place unique constraints on the temporally sequenced information found in written language (Christiansen & Chater, 2015; Perfetti, 1985). Such biological constraints that evolved to process a world without writing became attuned to the specific challenges of processing written words through “neuronal recycling” (DeHaene, 2009). Cognitive science attempts to explain how learners make the transition from no predisposition to literacy to becoming expert readers and writers. Factors that have been established as important include working memory, attention, and introspective awareness. Working memory, for example, stores the sequential information characteristic of both spoken and written language to facilitate fast integration of information packed in the orthography (Christiansen & Chater, 2015; DeHaene, 2009; Share, 2018). The information ranges from the simple appearance of graphemes to the more complex units that contain meaning (morphemes and words). The relationship between parts (such as the sequence of graphemes) needs to be analyzed fast to then discover how they represent phonology and meaning (Whitney, 2001). All this is achieved while monitoring comprehension through introspective awareness (Oakhill, Cain, & Elbro, 2014). Limitations in attention, another cognitive function that influences the processing of written language, allow multitasking (i.e., a number of activities can be performed simultaneously) only if all but one activity is automatized. In other words, the human brain can only perform one cognitive activity at a time, hence the need to adapt quickly to a complex set of tasks (Share, 2018).

All of these processes interact with the type of orthography a learner is exposed to. A large body of research has explored the way in which the orthographic inconsistencies presented in Section 1.1.1 are processed by the human brain and, more pertinent to the current thesis, how children's rate of acquisition is affected while they resolve them. Ambiguities have been shown to affect the way certain language specific regularities are learned, as well as the pace of acquisition. While learning the complex characters in Chinese requires initial time demands, the process speeds up and becomes efficient following grapheme mastery (see McBride, 2016). Alphabetic writing, on the other hand, is easier to learn and children learning a consistent orthography such as Czech demonstrates an accelerated growth spurt at the start of formal education (Caravolas, Lervåg, Defior, Seidlová Málková, & Hulme, 2013). In contrast, children learning an inconsistent orthography such as English follow a slow and steady rate of acquisition (Seymour, Aro, & Erskine, 2003; Share, 2008; Ziegler & Goswami, 2005), since English requires its learners to first master its variable mappings in order to reach efficiency for skilled reading. With increasing exposure to a particular writing system, the brain tunes its processes and adapts behavior to pick up the specific features that encode the language. At the same time, writing systems take into account and select those cognitive procedures able to capture the features needed for efficient and fast processing. Examples of such general-purpose procedures utilized by writing systems are order (e.g., the invariance in the phonetic expression of syllable structure in English), prominence of certain constituent units (e.g., the prominence of consonants over vowels in abugida orthographies such as Devanagari), or simultaneous processing of parallel systems of codes (e.g., Hebrew letters and diacritics) (Frost, 2012; Levy, 2012).

Since understanding the skills involved in learning a written language requires a thorough consideration of the features of one particular system, the current thesis focuses on one particular script: the Latin alphabet. When learning about a sound-based writing system such as the alphabet, children have multiple cues they can incorporate: That graphemes (or group of graphemes) map on to units of sound (phonological processing) and that they are frequently combined in certain sequences (orthographic processing) and that they are classified by their functional role, such as root words and inflections (morphological processing). All of the experiments presented in this thesis test learners who are already familiar with a specific language which makes use of the Latin alphabet—that is, English—although the aim is to uncover general mechanisms at play when learning any written language that uses the alphabetic script. In the general discussion (Section 7.5.1), I return to discuss the limitations of this approach in relation to broader criticisms regarding anglogcentricies in literacy research (Share, 2008).

Many of the cognitive systems mentioned above are implicit, that is, they do not require awareness (Reber, 1967) and, as noted in the limitations of attention, most of the activities involved in processing written language need to be automatized for skilled processing. Importantly, although becoming literate relies on explicit instruction and years of practice, just like other learned skills such as playing the piano, children are not taught how to process the written language but are rather provided with the logic of their writing system. Moreover, skilled readers and spellers process written language in a rapid and automatized way (Share, 2011). Thus, a goal of this thesis is to consider the role of *implicit statistical learning processes* in human literacy, processes that were shown to be relevant in other aspects of human language and cognition more broadly (see Section 1.5 for a detailed review). All written languages are rich in statistical properties (Venezky, 1970), such as the distribution of- and correlation between- graphemes and phonemes (or syllables) but also graphic conventions that do not have a phonological explanation (referred to as graphotactics henceforth and discussed in detail in Section 1.3), and the cognitive system, in turn, is a correlation-seeking device. This thesis builds on the assumption that the ability to pick up the statistical information in the written language is particularly important in learning to spell. Just as spoken and written language do not carry the same information, and as different writing systems put distinct constraints on the cognitive system, so does spelling presents unique challenges for the learner, compared with reading. Thus, the next section addresses the less widely explored unique demands placed on the cognitive processes when learning to spell.

1.2.2 Spelling: not just the reverse of reading

Two decades since Treiman (1998) wrote a book chapter entitled: *Why spelling?: The benefits of incorporating spelling into beginning reading instruction*, spelling development remains significantly under-researched. Reading and spelling ability are, unsurprisingly, correlated (Ehri, 2008; Malmquist, 1958; Shanahan, 1984), albeit moderately. Some researchers point towards this tight relationship to claim that the processes involved in reading and spelling are the same (Ehri, 1980; Gough, Juel, & Griffith, 1992). However, this view has been challenged by evidence that some good readers are poor spellers (Bruck & Waters, 1990; Holmes & Castles, 2001) but not vice-versa (Frith, 1980). Thus, some variance must be explained by a balance of abilities required for spelling, independent of reading. It has been shown that reading alone is not sufficient to ensure children master the harder task of spelling (Bosman & de Groot, 1991, 1992), and is limited as a way of improving it (Treiman, 2018). When reading, words can often be identified without processing every letter (Byrne, 1992; Treiman, 2018), while each word needs to be read many times before its spelling becomes consolidated in memory (Van Leerdam, Bosman, & Van Orden, 1998). Thus, learning to spell is more difficult, requires additional experience and develops more slowly than word recognition (Bosman & Van Orden, 1997). This led some researchers to argue that the two do not follow the same path and are, at times, separate abilities (Bryant & Bradley, 1980; Burns & Richgels, 1989). As a lexical ability, spelling requires the production of correct and conventional spellings, an encoding rather than decoding skill (Ravid, 2012). One dominant view in early theories of spelling development emphasized the dominant role of phonology in spelling: For example, Frith (1979) suggested that spelling occurs by ‘ear’ as opposed to ‘eye’. This emphasizes the dominant role of phonology in spelling and this view has had great support in early theories of spelling acquisition (Frith, 1980; Read, 1986; this literature is reviewed in detail in Section 1.4). However, in their proposed dynamic (or cognitive) systems framework Van Orden and Goldinger (1994) used a recurrent network model that included interdependent “nodes” (phonologic, orthographic and semantic) and found a powerful bidirectional connection between orthographic and phonologic nodes (see also Bosman & Van Orden, 1997). This was taken to suggest that phonologic mediation is essential in both spelling and reading. Inconsistencies between the orthographic and phonological dimensions, that is, letter-phoneme dynamic, are thought to be resolved by different constraints: Reading is achieved through decoding, that is, engages in letter-to-phoneme associations; and spelling is achieved through encoding, that is, engages in phoneme-to-letter associations (Daniels & Share, 2018). Spelling, therefore, is more difficult because, in most orthographies, sound-to-spelling associations are more variable than spelling-to-sound associations (Treiman, 2018). This means that even shallow orthographies that have a consistent letter-sound correspondence, such as Spanish, can be deep in the sound-letter correspondence, that is, have varying options for spelling phonemes (Valle-Arroyo, 1990). For example, when reading words such as *baya* (berry), *valla* (fence), and *vaya* (may he go), all pronounced [ˈβaja], it is simple to find the correct pronunciation since almost all Spanish letters can only be pronounced one way. However, when spelling, it is impossible, in fact, to know which letters to use based on the sounds alone. In another example of sound-to-spelling variability, the letters *c* in *cerrar* (to close) and *s* in *serrar* (to saw), are both pronounced as /s/. This asymmetry between reading and spelling is evidenced by the fact that skilled readers whose performance is above average, can display below average spelling performance, but not vice-versa, and spelling problems of dyslexics are more persistent than reading difficulties (Bosman & Van Orden, 1997; Critchley, 1975; Daniels & Share, 2018). When dealing with phonologically complex orthographies in both sound-to-spelling and spelling-to-sound direction, such as English, the impact of deficits such as dyslexia will be greater and both reading and spelling will be affected equally (Daniels & Share, 2018). A phonological deficit, therefore, has been shown to be a reliable predictor in dyslexia due to the obstacle posed by the phonological ambiguity, at least in phoneme-based writing systems such as alphabets (DeFrancis, 1989; Perfetti, 2003; Share, 2008; Ziegler & Goswami, 2005).

Despite the close relationship between reading and spelling discussed above, given the asymmetry between these two skills, it is not surprising that the two have different developmental paths. Interestingly, despite the greater difficulty of spelling development in later stages of literacy, there is some evidence that children may be able to make use of the alphabetic principles (that sounds correspond to letters/letter groups) in spelling before they can read (Byrne, 1992; Chomsky, 1979; Read, 1975; Seymour & Elder, 1986). For example, children's early invented spellings (Treiman, 1993) reveal their attempts to produce plausible spellings by analyzing the phonetic features of spoken words, without representing the orthographic conventions. This effort, however, is not directly linked to reading and often children are not able to read back what they have just written down (Bradley & Bryant, 1979; Bryant & Bradley, 1980; Burns & Richgels, 1989). Chomsky (1979) suggested that children sometimes operate in a 'writing mode' without engaging the 'reading' mode and hence, are not necessarily concerned with reading what they have written. Due to the focus on the orthographic detail in words through spelling, that is, the smaller units of sound that form them and their link to letters/letter clusters (Mommers, 1987; Share, 2018), spelling has a beneficial effect on reading (Cataldo & Ellis, 1988). Interestingly, the reliance on phonology early in literacy development when producing spellings is less demanding than reading, since it does not require holding phonemes in memory to then blend them together in a spoken word (Stahl & Murray, 1994). However, as noted earlier, becoming a proficient speller requires more time.

The phonological ambiguity cannot be the only challenge encountered by readers and writers, especially in writing systems other than the phoneme-based systems. Although there is extensive literature focused on phonological ambiguity, in fact, spellings are haphazard but often governed at least probabilistically, by "rules" or "constraints" on the possible use of letters in particular contexts. These patterns concern the visual arrangement of graphemes, and are referred to as graphotactics. They are the focus of the current thesis and I turn to discuss them in the next section.

1.3 Spelling and graphotactics

As noted in Section 1.2.1, written language is a highly patterned domain, and as such, visually presented words comply with regularities and constraints that are well explained in statistical terms. For example, English words never begin with *ck*: Simple, deterministic patterns of this sort are easy to verbalize and overtly taught in school. Other patterns (e.g., "consonants often double after single-vowel rather than double-vowel spellings") receive less attention in formal literacy instruction at least in part due to their complex, probabilistic nature (Kessler, 2009). Such constraints may be more or less dependent on phonology: Patterns may directly parallel restrictions in spoken language— that is, phonotactics (e.g., in English, words do not begin with /ŋ/ and accordingly, written words cannot have *ng* in beginnings) (restrictions discussed in detail in Section

1.5.2 and 1.5.4); they can be orthographic but reflect the influence of the phonetic environment (e.g., /ɛ/ is particularly likely to be spelled as *ea* when the word ends in /d/; e.g., *dead*); or can be purely visual (graphotactic) in nature with no phonological counterpart (e.g., *dd* does not begin written words, /d/ does; e.g., **ddoll*). The statistical properties of graphotactics, the patterns of interest in this thesis, are explained by frequency of graphemes and their probability of occurrence and co-occurrence in a certain context. Doublets[2] are particularly relevant examples of purely graphotactic constraints whose learning does not necessarily require phonological input. Another term used to refer to double letters, predominantly in the phonological literature, is 'geminate'; however this term generally refers to the use of double letters to represent an acoustic feature (i.e., duration): In Finnish, for example, doublets stand for long consonants. However the doubling of letters does not always result in a change in how phonology is represented, and this is the case of interest in the current thesis, and I therefore use the more neutral term 'doublets'. In English and French, for example, many consonants (but not vowels[3]), preserve the same pronunciation in both their single and doublet form (e.g., in English, *l* in *old* and *ll* in *roll* represents the single sound /l/; in French, *ul* in *formule* and *ull* in *bulle* are both pronounced /yl/). The usage of doublets versus single consonants is not random, but instead constrained. In English, for example, some consonants never double (e.g., *k, y*), others rarely double (e.g., v) while many double frequently (e.g., *l*, *t*, *s*), and they only appear within word middles (e.g., *bunny)*, or word endings (e.g., *bell*). Similar doubling "rules" can be found in French, where some consonants double frequently (e.g., *m*, *l*), while others do not (e.g., *k*, *x*) and they can only appear in word medial (e.g., *pomme*) but never in initial or final positions. In addition to frequency based and positional conditioning, the use of letters may also be conditioned on aspects of contexts both phonological (e.g., short vowels are more likely to be followed by doublets as in *supper*, pronounced /ˈsʌpə(r)/,

---

[2] Note that this is a different use of the word "doublet" than "linguistic doublets" which refer to two or more words that share the same etymology but differ in the phonological form, such as *host* and *guest*. The doublets in spelling research are two identical letters that appear together in a word and represent one phoneme.

[3] Vowels do not double in French

vs. *super*, pronounced /ˈsuːpə(r)/), or purely visual (e.g., doublets occur more often after a single- than double-letter spelling, as in *bedding* vs. *heading*), and the latter is an example of “pure” graphotactic constraint. Frequency, positional and context based influences of this sort, which regulate consonant doubling, have been seen in naturalistic spellings of preliterate children (Treiman, 1993) and were shown to emerge in carefully designed experiments. I review this evidence in the next sub-sections.

1.3.1 Sensitivity to position of doublets

The work of Treiman and colleagues (Cassar & Treiman, 1997; Treiman, 1993) unequivocally demonstrates recognition-based knowledge of frequency and positionally based constraints on doublet usage in early childhood. In a seminal study, Treiman (1993) showed that Grade 1 (6–7-years-old) children’s invented spellings reflect sensitivity to allowable doublets in English. That is, the doublets used by children appeared more frequently in word medial or final positions—compared to word initial positions, which is not allowed in English—and these were the ones that appeared most frequently in their language (e.g., *ee*, *bb* vs. *hh*, *kk*). This was also shown in controlled experimental conditions using the orthographic constraints task (also referred to as word-likeness task): Children were presented visually with pairs of nonwords where one conformed to- and one violated orthographic rules of double letters in English, and heard one oral pronunciation for each pair, and were subsequently asked to choose which item was the most word-like. It was found that children chose conforming items more often than expected by chance. These results were replicated by Cassar and Treiman (1997), using a similar task: 6-year-old English-speaking children chose more spellings with final doublets over initial ones when asked to choose the stimulus that looks more like a real word (e.g., *baff* vs. *bbaf*). Since the identity of the consonants and their legality as a doublet did not influence the results (i.e., children chose *juss* over *jjus* even though *j* never doubles in English), this knowledge was interpreted to be at a generalized level. However, Pacton, Perruchet, Fayol, and Cleeremans (2001) argued that the choice of *juss* could have been in fact because *s* doubles frequently in English and not because *jj* could not occur at the word beginning, leaving unresolved the question whether knowledge on legal position of doublets is general. To control for this confound, they contrasted consonants that are equally frequent in French[4] in both single and double form (e.g., *m*, *l*) with consonants that are frequent only in the single but not double form (e.g., *c*, *d*). In the word-likeness task, French speaking children as young as 6 years old, showed sensitivity to the frequency of double consonants, that is, their selection of nonwords with doublets (e.g., *ommera* vs. *ovvera*) and with single consonants (e.g., i*dose* vs. *imose*) was consistent with the frequency in their language. Pacton et al. (2001) further investigated children's sensitivity to the legal position of double consonants (only in word medial position in French), and replicated Cassar and Treiman's (1997) results. Importantly, they also showed that this sensitivity holds not only for consonants that often double (e.g., children chose *bummor* over *bumorr*) but it extends also to consonants that never doubled in French (e.g., they chose *bukkox* over *bukoxx*). Children also performed similarly in a word completion task, where they were asked to choose between two consonants, one as a singlet and one as doublet, to complete trisyllabic nonwords (e.g., *tuba_ir* or *u_otir*), providing experimental evidence in line with Treiman's (1993) observations that children rarely make errors such as *bbal* for *ball*.

Wright and Ehri (2007) addressed the ecological validity of this knowledge by investigating how doublet legality influences 6−7-year-old English speaker's word learning and memory: They taught participants modified spellings of real English words (similar to invented spellings produced by beginner spellers, e.g., *rrag* for *rag*) and subsequently asked them to retrieve them in an immediate production and a delayed recognition post-test. Critically, spellings were learned faster to criterion, were misspelled less frequently, and were better recalled, when they were possible in English (doubled letters at the end of word as opposed to the beginning) relative to when they were not (doubled letters at the beginning of word). For example, children miss-recalled *RREK* as *REK* or *REKK*, converting the illegal letter string to a legal one.

To further probe children's ability to use graphotactic knowledge acquired from reading exposure, Pacton, Sobaco, Fayol, and Treiman (2013) presented 9 year old French-speakers with stories to be read silently for meaning: These

---

[4] Note that, as described earlier in this section, doublet patterns in English and French are comparable and hence, this extension of results is relevant.

embedded legally and illegally spelled French nonwords. Similar to Wright and Ehri's (2007) findings with younger participants, words that violated double patterns in word beginning and medial positions (Experiment 3) were remembered poorly and illegally spelled words were “regularized”, mainly via omission errors (i.e., “dropping” one consonant of a doublet). Qualitatively similar findings were obtained in a follow-up study with adults who were incidentally exposed to the nonwords within texts or in isolation (Sobaco, Treiman, Peereman, Borchardt, & Pacton, 2015). Interestingly, this sensitivity to the position of doublets has been reported even in languages where single consonants and doublets have different pronunciations. In Finnish, for example, long consonant phonemes can occur at the beginning and middle of a word but the consonant letters that represent them can only double in word medial positions. Learning such constraint, therefore, requires knowledge about the phonological principle that doublets represent long consonants but also about the “formal rule” that they can occur only in word medial positions. Lehtonen and Bryant (2005) provide evidence for 7 years old Finnish children’s sensitivity to the legal position of doublets (graphotactic constraints) before any knowledge of their phonological function.

The studies discussed above each explored spelling regularities that are conditioned by the position or frequency of letters. These may help learners resolve some spelling irregularities, such as where consonants double, but extensive analyses of the English orthography (e.g., Venezky, 1970) show that they are less powerful cues relative to patterns that condition spellings based on surrounding context (Kessler & Treiman, 2001; Treiman & Boland, 2017; see also Treiman & Kessler, 2019, for context effects in the reading direction). Contextual constraints that influence when (not just where) consonants double can be phonological or purely visual and I discuss both in the next sub-sections. The literature exploring the effect of the phonological context on spelling patterns has been particularly concerned with vowels, and this research is presented here, before returning to consonants and how their doubling is conditioned by the phonological and graphotactic (purely visual) context.

1.3.2 Contextual (phonological) constraints on vowel spelling

In English, vowel spellings are more variable than consonant spellings (Kessler & Treiman, 2001) and the surrounding context helps predict some of the inconsistencies (e.g., Treiman, Kessler, & Bick, 2002). For example, /ɑ/ is spelled as *a* when preceded by /w/ (e.g., *wand*) but as *o* when preceded by other consonants (e.g., *pond*) (Hayes, Treiman, & Kessler, 2006; Treiman & Kessler, 2006). While context effects on vowel spellings are most often operationalized as the influence of following letters, Varnhagen, Boechler, and Steffler (1999) and Treiman and Kessler (2006) investigated American English children's appreciation of how the onset—that is,the segment that *precedes* the vowel—influenced spelling of the sound /ɑ/: Its spelling depends on the sound of the previous letter, such that, while /ɑ/ is commonly spelled as *o* (e.g., *slop*), monosyllabic words beginning with *sw* always take an *a* spelling (e.g., *swan*, *swat*). In a fill-in-the-blanks task, children heard the pronunciation of a nonword and were asked to fill in the missing letters. Both studies found that an effect of onset-to-vowel associations (e.g., /ɑ/ is spelled as *a* when preceded by /w/ but not by /b/) emerged from 8−9 years and increased with age. Treiman and Kessler (2006) found that coda-to-vowel associations (e.g., /aɪ/ is spelled *igh* before /t/, as in *light*, but as *i* when followed by /e/ final, as in *time*) emerged only around 12−13 years old, contrary to the belief that rimes play an essential role in associating phonology and orthography (Treiman et al., 2002, see Section 1.3.3). The sensitivity to context was seen to be affected when the vowel had a single dominant spelling: 6–7-year-olds in Varnhagen et al. (1999) study spelled the words using the more common letter for /ɑ/, regardless of the context predicting the less common spelling; and Treiman and Kessler (2006) found that children were slow to learn with those vowels that had a single dominant spelling, compared to those that did not. This was taken to suggest that context was not used as early and as efficiently as expected, given the rich contextual information available in the natural language.

1.3.3 Contextual (phonological) constraints on consonant doubling

While consonants in their singlet form have more predictable spellings than vowels in English, their doubling is conditioned by their position (as discussed in Section 1.3.1) but more effectively by the surrounding context. One possible influence is the phonemic properties of the surrounding vowels. Medial vowels, for example, influence following consonants: Short vowels (for American English, /æ/, /e/, /ɪ/, /ɑ/, /ʌ/, /ʊ/) are more likely to be followed by consonantal doublets over singlets (e.g., *supper*) and long vowels tend to be followed by singlets (e.g., *super*).

Phonological context influences on consonant doubling were assessed in one condition of Cassar and Treiman (1997) by presenting (visually and auditorily) participants with nonwords featuring medial single/double consonants preceded by a short vowel (e.g., /ˈtɛbɪf/; "is it **tebif* or *tebbif*"?) or a long vowel (e.g., /ˈsobæp/; is it *sobap* or **sobbap*). Adults and 11–12-year-olds (but not younger children, 6–10-year-olds) were above chance at choosing items embedding correct long (i.e., *sobap* for /ˈsobæp/) and short (i.e., *tebbif* for /ˈtɛbɪf/) vowel transcriptions and this was taken to suggest that phonological context affects children's preferences later in development relative to constraints on positions and letter identity. Broadly, this finding holds across other languages, e.g., Danjon and Pacton (2009), although only older French children were sensitive to contextual constraints such as doublets occurring before, but not after, a single consonant. The exact age at which context effects emerge are likely to depend on methodological task differences and the pattern being learned (see e.g., Deacon, Leblanc, & Sabourin (2011) for evidence of long vs. short vowel context sensitivity among 9.5-year-olds).

1.3.4 Contextual (graphotactic) constraints on consonant doubling

As mentioned earlier in this section, the phonological pattern on vowel pronunciation described above is not the only possible cue to consonantal doubling. Hayes et al. (2006) investigated sensitivity to a second, *graphotactic* in nature, cue that operates independently from phonology: Doublets occur more often after a single- than double-letter spelling (e.g., *Jeff* vs. *deaf*; *bedding* vs. *heading*). Similar to Cassar and Treiman (1997), they asked children to choose the most word-like item between two nonwords that either conformed to or violated this pattern (e.g., *vaff* vs. **vaf*; *vaif* vs. **vaiff*) and also used a nonword production task. Pattern-conforming performance was shown in both, but the nonword production results are particularly important in one additional way: The graphotactic influence can be hard to distinguish from the phonological pattern on vowel pronunciation because, in English, short vowels almost always take single-letter spellings (e.g., *tell*) while long vowels often take two-letter spellings (e.g., *tail*). In many instances, therefore, the phonological and graphotactic cues correlate with each other. In the 7–8-years-old's spelling attempts, though, doubling was less likely to occur after two vowel letters even when they represented (somewhat unconventionally) a short vowel. For example, if a short vowel /æ/ in /sæf/ was spelled with two vowels (e.g., *ae*), *f* was more likely to be produced over *ff*. Thus, the graphotactic pattern influenced children's performance above and beyond the phonology. While Hayes et al. (2006) study did not find this with adults (who tended to prefer one-letter spellings for short vowels), Treiman and Kessler's (2015) did demonstrate that adults' choice of consonant doubling was influenced by the number of letters used for the preceding vowel, regardless whether the vowel was short or long. Other recent work (Treiman & Boland, 2017; Treiman & Wolter, 2018) has further demonstrated adults' purely graphotactic sensitivity using disyllabic words rather than monosyllabic words, and this adhered to context on doubling as in Hayes et al. (2006), on preceding as well as following context.

1.3.5 Concluding remarks

The studies reviewed in this section show that skilled (adult) and developing readers are sensitive to a range of spelling patterns, including when and where consonants double. Clues as to how to resolve ambiguous spelling situations are provided by frequency and positional statistics, as well as contextual constraints, both phonological and purely visual. Sensitivity to graphotactic regularities was shown to emerge early (6−7 years old) and develop gradually, with frequency and positional spelling correspondences being learned first, and conditional correspondences being utilized later on (Cassar & Treiman, 1997; Pacton et al., 2001).

While these effects are extensively demonstrated among English-speaking children, those learning Finnish (an almost exceptionless orthography in terms of letter-sound correspondences) and German (where many, but not all letters have one-to-one correspondences with sounds) also prefer nonwords that adhere to doublet constraints and graphotactic probabilities in their language (Ise, Arnoldi, & Schulte-Körne, 2014; Lehtonen & Bryant, 2005). Similarly, Spanish-speaking 7−12-year-olds prefer nonwords with frequent than infrequent bigrams in their orthography (e.g., *b* was preferred to *v* before *u*) (Carrillo & Alegría, 2014).

If children and adults are sensitive to these types of constraints, then theories of spelling development must account for how they are learned. Importantly, the above described graphotactic patterns are mostly untaught or, importantly, not taught at the age at which sensitivities were shown to emerge. Section 1.8 and 4.1 will address the possible benefits of explicit teaching of such patterns. The evidence reviewed in this section, therefore, strongly suggests that children pick up statistical spelling patterns implicitly, via exposure to print. However, this evidence is indirect and therefore cannot answer questions such as how much exposure is needed for pattern sensitivity to emerge; or are these patterns learned at a fast rate or in a protracted development. In line with this, previous work has failed to find associations between children's orthographic knowledge, measured in the lab, and levels of print exposure, estimated using questionnaire-based measures of home literacy environment (Ise et al., 2014). In sum, questions of learnability are hard to address in studies that probe children's (and adults') sensitivity to patterns in their actual orthography. Learning experiments allow control over the input learners receive and enable researchers to investigate whether and how the patterns are picked up implicitly. This approach is known as statistical learning and is the approach taken in experiments in the current thesis. Before I review relevant statistical learning experiments in Section 1.5, in the next section, I review theories of development of spelling more broadly.

1.4 Spelling development

Becoming a skilled speller in an alphabetic system requires years of daily practice and effort. Early, traditional, models of reading and spelling acquisition (Frith, 1985; Gentry, 1982) prescribed that the development follows in stages, initially reliant on and driven by phonology, followed by use of orthographic and morphological information. More recent theories and studies show that learners use all these sources of information when choosing between different spelling options (Roman, Kirby, Parrila, Wade-Woolley, & Deacon, 2009; Treiman & Kessler, 2014) from the beginning of experience with print (Kessler, Pollo, Treiman, & Cardosso-Martins, 2013; Treiman, Kessler, & Bourassa, 2001). Recent research (Treiman & Kessler, 2014) also emphasizes the early emergence of knowledge about graphotactics, that is,"writing's outer form" (Treiman, 2017), in contrast with views from the phonological perspective. I present these theories here, with a view to highlight when and how sensitivity to graphotactics emerge.

1.4.1 Phonological Perspective

The phonological perspective of spelling, with Gentry (1982) and Frith (1985) as its principal proponents, concluded that children only acquire orthographic and morphological knowledge once they have established knowledge of letter-sound correspondences. The theories that emerged from this perspective proposed that processes involved in encoding written language follow in a stage-like manner and progress from one stage depends on the mastery of the previous stage. The processes involved in mastery and progress were identified as both internal (cognitive and linguistic capabilities) and external (instruction, experience with print), both playing an equal role in the spelling development. While the labelling, number and time course of these stages differ slightly (sometimes also referred to as phases), the theories agree on a broad course of learning. In an initial, *pre-communicative stage*, children's spellings were seen to be completely random, with no basis in phonology, such as producing *HS* for *quick* (Treiman, 1993). Once children acquire knowledge about letters, they start to incorporate this phonological information into their spellings. Next, they demonstrate an ability to connect the sound values of letters (phonetic cues) with constituent sounds of words, albeit only partially. In this *semi-phonetic stage*, children attempt to partially represent the phonetic structure of words by using the most salient letter sounds, such as *KE* for *cake*. The *phonetic stage* is reached only when children acquire encoding skills, enabling them to connect all phonemes and graphemes, and bond pronunciations with spellings. The process was believed to be facilitated by phonemic awareness, that is, the ability to (i) segment words in component sounds, (ii) blend sounds into words and (iii) substitute sounds (Stuart & Coltheart, 1988). Awareness of orthographic conventions was not believed to emerge at this stage, and errors such as the spelling of *truck* with the onset *ch*, following the shared sound with *chair* (Read, 1986) were taken to demonstrate early phonological- but a lack of orthographic knowledge. The latter was seen to be demonstrated by children only later, in the *transitional stage,* when they chose between different letters that map onto the same sound, on the basis of convention (e.g., *c* at the beginning of word and not *k*). By the final, *conventional stage* in spelling, the rules of the orthographic system were believed to be consolidated. Children have both (i) developed the cognitive abilities to process the complex system embodied in their orthography, and (ii) received appropriate instruction and experience to master their writing system.

A limitation of the theories summarized above, however, is that the time required to achieve skilled and automatized reading follows a long process that does not account for overlap, that is, skills that fit in more than one stage may emerge at the same time. In addition, individuals differ in the pace and style of learning. It has also been argued that the stages are specific to English language, which has inconsistent sound to letter mappings. Children learning a shallow orthography, however, might skip some stages and reach the alphabetic stage faster (Caravolas, 2004; Caravolas et al., 2013; Cardoso-Martins, 2005; Cardoso-Martins, Corrêa, Lemos, & Napoleão, 2006; Marinelli, Romani, Burani, & Zoccolotti, 2015; Wimmer & Hummer, 1990). Another shortcoming of the phonological perspective is that it highlights the phonology as driver of spelling development, with much less attention to orthographic and morphological processes. However, letter representations that seem random in the pre-alphabetic stage could well have developed as a result of sensitivities that are not available in the phonological input. For example, in Bissex (1980), a four-year-old's *SSDICA* spelling for *welcome home* could have reflected that *s* doubles frequently in English, his native language (see Sections 1.3.1 and 1.3.4 for examples of graphotactic constraints on consonant doubling). Furthermore, the partial representation of words using letter names in the semi-alphabetic stage has not been consistently demonstrated (Shankweiler, 1994; Varnhagen, Mccallum, & Burstow, 1997). Varnhagen et al. (1997), for example, found that children at this stage do not consistently use the letter-name strategy (e.g., *CR* for *car*) and some letter names are used more than others, depending on their phonological properties (e.g., they spell *r* for *are* vs. *te* for *tea*). Importantly, letter name information is assumed to be unavailable at this stage, hence, these invented spellings may represent knowledge other than letter sound, such as graphotactic. For example, in the misspelling of *cake* as *kak* and *cack*, the choice of *k* and *ck* for the sound /k/ could represent different elements of graphotactic knowledge in English, such as *k* and *c* can occur at word beginning and end but *ck* ends but does not start a word. In addition to early orthographic knowledge, morphological knowledge is shown in errors such as over-applying the –*ed* rule (to represent past tense, as in *clapped*) to words that are not verbs, such as *sofed* for *soft* (Nunes & Bryant, 2006). These errors and invented spellings (for a review, see Read & Treiman, 2013) are believed to be constructed spontaneously by children, using the information available to them. In sum, there is evidence that even children's earliest spellings reflect the regularities in the visual input they are exposed to, speaking against an account in which early spelling is purely driven by phonological *concerns.*

1.4.2 The constructivist perspective

At the same time as the phonological perspective, the constructivist theory put forward by Ferreiro & Teberosky (1982) challenged the stage theories and proposed that children pick up a great deal of information about their writing system before they start formal instruction and before they understand that letters represent phonemes. According to this theory, children build hypotheses about how writing works by identifying, assimilating and internalizing regularities and applying them to new input. In a *presyllabic stage*, children develop abstractions based on the minimum quantity and variation hypothesis, that is, that words need to have three or four letters, and these letters must be different from each other within a word. They also hypothesize that the visual form of the word (i.e., length) reflects properties of the object it names, that is, that large objects need larger spellings (e.g., *dog* or *elephant* will need more letters, while *puppy* or *mouse* will need less) (Zhang & Treiman, 2015). In later spelling development, children become aware of the link between letter and sound and write words using the number of syllables that form that word (*syllabic stage*) and then learn that letters stand for phonemes (*alphabetic stage*). Models in the constructivist perspective, however, do not move away from the suggestion that children move progressively from one hypothesis to another.

Some of the assumptions regarding some of the specific hypotheses children form and use have been challenged and further studies have failed to consistently support the view that children go through a stage when they represent syllables with the same number of letters (Cardoso-Martins et al., 2006; Pollo, Kessler, & Treiman, 2009; Treiman, Pollo, Cardoso-Martins, & Kessler, 2013). Furthermore, children have been shown to become sensitive early on to many characteristics of the printed material in their environment other than those described through the constructivist hypotheses. For example, children were shown to also pick up on the typical length of words and frequency or doubled letters, among other patterns (Pacton et al., 2001; Pollo et al., 2009; Treiman, 1993). Finally, the constructivist theory does not account for processes after the alphabetic stage, when learning is not yet complete, especially in complex orthographies such as English.

1.4.3 The dual route theories

Dual route theories (Barry, 1994; Kreiner & Gough, 1990) move away from the suggestion that spelling development plays out in a sequence of progressive processes. Instead, they propose that learners reach proficiency via two separate routes: The nonlexical route facilitates learning through a set of rules, offering a systematic link between single phonemes and graphemes (either individual letters or letter groups, e.g., *sh* represents one sound); and, at the same time, the lexical route enables whole words to be accessed from memory when rules in the nonlexical route do not produce correct spellings (e.g., exception words). During spelling development, children start with a reliance on the non-lexical route while they build a memory of whole word instances with experience, in the same way that they build sight vocabulary (Ehri, 2005). The reliance on one of the two routes depends also on the consistency of orthographies that are learned. Learners of deep orthographies such as English must rely particularly on the lexical route even beyond beginning of literacy acquisition, since such spelling system lacks a systematic letter-sound link required for the nonlexical route. This can explain the different pace of learning observed in different alphabetic writing systems (Marinelli et al., 2015), a distinction not accounted for by the stages proposed in the phonological perspective. However, such a simplified model is not sufficient to acknowledge and account for all systematic patterns available to learners, many of which have been described in Section 1.3. While the nonlexical route relies on rules for simple, context-free associations, many of the patterns in written language are conditioned by context. Furthermore, even learning of simple phoneme-grapheme associations is not equally easy across all possible rules, since many of them are probabilistic (see also Treiman, 2017). Storing in memory a large number of whole words cannot be the most efficient way for the cognitive system to process a written language that is dynamic. While phonology has a close (albeit variable) relationship with spelling (see Section 1.1) and plays an important role in spelling development (see Section 1.4.1), it is not the only dimension to be learned, as established in Section 1.1.2. The role of morphology and orthography (particularly graphotactics) are acknowledged by theories that take a whole-language approach.

1.4.4 The whole-language approach

Stage models of literacy development presented earlier (e.g., Frith, 1985; Gentry, 1982), do not explain sensitivities to non-phonological input and instead, they treat children's letter representations in the pre-alphabetic stage as random. While the constructivist models acknowledge the early nature of pattern sensitivity, similar to stage theories, they suggest that children move progressively from one hypothesis to another. Even a model that moves away from the belief that learning to spell proceeds in a sequential manner, such as the dual-route, fails to acknowledge the role of probabilistic or context dependent patterns. Furthermore, all theories discussed so far share an assumption that processing of patterns that occur frequently does not vary based on the function and meaning they convey (i.e., all graphemes are processed similarly and the spelling of /ayt/ as either *–ite* or *–ight* as in *site* and *sight* is retrieved from memory; Ehri, 2005).

Accumulating evidence is challenging the views brought forward by the early theories and highlights the importance of orthographic (e.g., Cunningham, Perry, & Stanovich, 2001) and morphological (e.g., Deacon & Kirby, 2004) processing and the cumulative effect these have together with phonological awareness on reading and spelling development (e.g., Deacon, 2012; Roman et al., 2009). According to Treiman and Kessler's (2014) Integration of Multiple Patterns framework (IPM), children use multiple sources of information when choosing between different spelling options and their learning is better when these sources support the same spelling (Kemp & Bryant, 2003, with English children; de Bree, van der Ven, & van der Maas, 2017, with Dutch children; Pacton, Fayol, & Perruchet, 2005, with French children). In English, for example, children may choose the letter *s* at the end of a word like *fibs*—even though the final sound is /z/—by using knowledge of plural inflection as well as graphotactic information (i.e., words cannot end with *z*). When spelling a word like *fees*, however, their choice between letters *s* and *z* will be more difficult since graphotactics do not prevent *freeze* spelling. Children's spelling errors also reveal that orthographic processes interact with phonology, morphology and meaning from the beginning of experience with printed words (Goswami, 2012) and underlie the entire development of spelling acquisition (Pollo et al., 2009; Pollo, Treiman, & Kessler, 2007). In addition to the role of words' inner form (i.e., phonology and morphology), theories such as the IMP emphasize the role of words' outer form (i.e., non-phonological information). They suggest that children learn about the visual form of written words and find systematic patterns in print (graphotactics) much earlier than stage or dual-route theories propose—around 2−3 years of age and before children understand that letters represent phonemes (Kessler et al., 2013; Treiman et al., 2001). As children learn to read and spell, they increasingly develop tacit knowledge of statistical properties of printed words (Pollo, Treiman, & Kessler, 2008) and this knowledge is believed to be picked up implicitly through statistical learning mechanisms (Treiman & Kessler, 2014). This view of spelling development is further explored in this thesis, which applies a statistical learning perspective and investigates whether the context-based graphotactic constraints can be learned by children (and adults) under experimental conditions. To that end, in the next section (Section 1.5), I review general literature on statistical learning. This is followed by a discussion of statistical learning experiments that have been carried out to date which have specifically investigated learning of graphotactic patterns (Section 1.6).

1.5 Overview of statistical learning literature

Interest in statistical learning (SL) can be traced back a century (for a rreview, see Christiansen, 2019) when it focused primarily on language change (Esper, 1925). Decades of obscurity followed, in the shadow of the "cognitive revolution"[5] boosted by Chomsky's famous rejection of Skinnerian behaviorism[6] (Chomsky, 1959). The work of Saffran, Aslin, & Newport (1996) renewed interest in input driven theories of (spoken) language acquisition. They demonstrated that

[5] The cognitivist movement defined its subject matter as "all processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used.[…] Such terms as sensation, perception, imagery, retention, recall, problem-solving, and thinking, among many others, refer to hypothetical stages or aspects of cognition" (Neisser, 1967, p. 4).

[6] With the publication of Science and human Behavior (Skinner, 1953), Skinner became the main spokesperson for behaviorism, defined as the study of the functional relations between behavior and environmental events (Catania, 1988).

infants were able to track statistical relationships between syllables to help them solve a key initial problem in acquiring a lexicon: Segmenting words from fluent speech. Unlike written language, speech does not consistently mark word boundaries through acoustic cues such as pauses and, instead, statistical regularities found within and between words (or utterances) differ such that sound sequences within words recur with higher probability (high transitional probabilities, of TPs), while sound sequences across word boundaries are more accidental (low TPs). In statistical terms, a high TP indicates that one syllable is likely, given the previous syllable, while low TP indicates that such probability is unlikely (Harris, 1955; Swingley, 2005). A popular example is the sequence *prettybaby*, where the likelihood of the syllable *pre* being followed by *ty* is high (i.e., when infants hear the syllable *pre*, this is followed by *ty* around 80% of the time), while, conversely, the probability of *ty* (in *pretty*) followed by *ba* (in *baby*) is low (0.03%) since the word *baby* may be surrounded by many other words (Saffran, 2003). Saffran et al. (1996) demonstrated that 8-month-old infants were sensitive to this type of information when listening to a 2-minute continuous artificial speech stream (e.g., *pabikutibudogolatudaropi*…), consisting of four tri-syllabic words (*pabiku*, *tibudo*, *golatu* and *daropi)*: After listening, infants discriminated between “words” with high TPs (e.g., *TP* = 1.0 for *pabi*) and strings of syllables which crossed word boundaries, and which had lower TPs (e.g., *TP* = 0.33 *for kuti*). The Head-Turn Preference Procedure, whereby children show interest (i.e., turn their head) when they hear unfamiliar combinations (lower TPs) allowed the researchers to determine that infants were more interested in part-words relative to the target words, that is, they reliably identified word boundaries using distributional properties in the input.

In the decades since Saffran’s seminal work, the prolific and dynamic field of SL research generated multiple definitions of SL (Arciuli, 2017; Armstrong, Frost, & Christiansen, 2017; Aslin & Newport, 2012; see also Frost, Armstrong, Siegelman, & Christiansen, 2015 for a review of SL as a domain-general mechanism; Perruchet & Pacton, 2006). While there has been agreement on the broad function of this computational mechanism, that is, to detect and extract the distributional probabilities from the input (e.g., spoken language), two separate strands of literature have emerged: Statistical learning (SL) and implicit learning (IL) (for a review, see Christiansen, 2019; Perruchet, 2019). In essence, both SL and IL literatures focus on the same phenomena and are built on the premise that humans—among other species—have an ability to extract patterned regularities, across space and time, from the rich stream of information perceived from the environment via sensory modalities (e.g., auditory, visual, and tactile) (Armstrong et al., 2017). In the IL literature, spearheaded by Reber (1967, 1989), this ability was characterized as being intuitive and unconscious, that is, it evolves without intention to learn and without explicit awareness of what is learned. While SL is indeed an implicit process and both SL and IL act in incidental (unsupervised) learning conditions (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997)—the terms having occasionally been used interchangeably—the divide between the two literatures remains significant (Perruchet, 2019; Perruchet & Pacton, 2006). While the SL literature is represented by studies that focus on the learnability of particular types of statistical structure, such as the TPs (Miller, 1956), the IL literature involves artificial grammar learning (AGL) and serial-reaction-time (SRT) methodology and chunk-based learning (Reber, 1967). According to IL theory, cognitive units (or chunks) rather than statistical probabilities such as TPs, are formed via implicit learning processes. For example, Arnon, McCauley, and Christiansen (2017) claim that, when segmenting a speech stream, infants initially extract units larger than words, generally short multi-word utterances. Perruchet (2019) describes the discovery of chunks in *onceuponatime*: It may be segmented as *on/ceu/ponat/ime*, *onceu/po/nati/me*, *onc/eupo/natim/e* or the correct segmentation, "*once/upon/a/time*. Learners will need to select the segmentation that best meets a certain criteria, after considering all or some of the possible segmentations.

Although the current thesis uses the term implicit learning when referring to learning in unsupervised, incidental experimental conditions, the term does not refer to its interpretation as seen in chunk-based models. Nevertheless, an understanding of the distinction is necessary, particularly considering recent support for chunk-based models in IL over TP-based models in SL (Perruchet, 2019), and an emergence of the unified term "implicit statistical learning" (Christiansen, 2019; Conway & Christiansen, 2006).

In the following sections, studies from the SL literature are reviewed, to demonstrate that infants, children and adults are indeed able to detect environmental probabilistic patterns and process them in order to "make accurate decisions about novel stimulus circumstances" (Reber, 1989, p 219). The focus of early SL research was to provide evidence for its role in language acquisition. In the decades since, a substantial body of research has found that SL operates across distinct cognitive domains other than language, such as music (e.g., Creel, Newport, & Aslin, 2004; Tillmann & McAdams, 2004), vision (e.g., Fiser & Aslin, 2001, 2002; Kirkham, Slemmer, & Johnson, 2002), and movement (e.g., (Ongchoco, Uddenberg, & Chun, 2016), among many (for a review, see Armstrong et al., 2017). I begin, in Section 1.5.1, by reviewing the large body of research exploring the role of SL in the auditory domain, particularly highlighting non-verbal infants' sensitivity to regularities in their language.

1.5.1 Auditory Statistical Learning

It has been shown that infants as early as seven months are able to acquire complex information about the properties of their language by simply listening to speech and detecting patterns in input (Jusczyk, Luce, & Charles-Luce, 1994). They do this by extracting statistical properties (frequency of occurrence and co-occurrence of certain sounds, stress, and intonation) through distributional and probabilistic information (Jusczyk et al., 1994; Mattys, Jusczyk, Luce, & Morgan, 1999; Saffran et al., 1996). As described in Section 1.5 above, Saffran's findings provided the bedrock for a large number of subsequent studies and were replicated and extended in various ways. An outstanding issue identified and addressed by Aslin, Saffran, and Newport (1998) was the precise nature of statistical computations measured by the study. In the input used in the original experiment (Saffran et al., 1996), infants could have preferred part-words based on the more simple statistics (i.e., frequency of co-occurrence) rather than the conditional probabilistic ones (TPs). This is because TPs shared the variance with the frequent co-occurrence of syllables, confounding the statistical cues used to segment the words. Thus, Aslin et al. (1998) created a new speech stream where frequencies of syllable sequences were equated for within and across words, while maintaining the differences in TPs. They achieved this by making two of the four words appear twice as often in the corpus so that syllable sequences across boundaries between these two common words had relative high frequency. Infants continued to prefer part-words even when the two statistical cues were equated.

TPs are not the only statistical constraints that provide clues about word boundaries. Mattys and colleagues (Mattys & Jusczyk, 2001; Mattys et al., 1999) demonstrated that infants between 7.5- and 9-month-old make use of their sensitivity to probabilistic phonotactics as well. Phonotactics are probabilistic constraints on what sound sequences are allowed in a language and where they can occur. They are particularly relevant in the current thesis since they provide the most direct counterpart in spoken language to the graphotactic constraints on possible sequences of graphemes. An example of a phonotactic constraint is that the sequence *br* (as in *bright*) is likely to appear in word initial position, and *nt* (as in *vent*) at the end of English words, but certain consonant clusters (e.g., *vl*, *dn*, *pt*) cannot represent a valid word-beginning syllable (Halle, 1978). Thus, both *br* and *nt* are good candidates for word onset and word end, respectively. Importantly, these patterns have been shown to be language specific and thus, they must be learned (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk et al., 1994; Mattys & Jusczyk, 2001; Mattys et al., 1999). For example, English does not allow *ng* (/ŋ/) at the beginning of words while English-Vietnamese does (Onishi, Chambers, & Fisher, 2002). Mattys et al. (1999) showed that 9-month-old infants use phonotactic cues to segment words from fluent speech. They exposed infants to CVCCVC (consonants and vowels) sequences with word internally frequent or infrequent CC clusters. The infants reliably segmented the sequences into two words in the infrequent CC cluster cases but not the frequent ones.

Phonotactic knowledge, however, plays a role beyond word segmentation. For example, many patterns within words are illegal word-beginning and -end syllables but do not mark word boundaries (e.g., *mb* in *embed*) (Gambell & Yang, 2005). However, phonotactic knowledge is essential in learning L1 (first language) phonology, enabling children to acquire phoneme categories and group speech sounds into syllables that are valid in their own language (Fisher & Gleitman, 2002; Pitt, 1998; Smith & Pitt, 1999; Treiman & Zukowski, 1990; for an analysis of syllabification see also McCarthy, 1979). For example, in the word *embed*, knowing that *mb* is phonotactically impermissible at word beginnings informs learners that the sounds *m* and *b* are placed in separate syllables (*em-bed*) when they occur word medially. With this knowledge and continuing experience with a language, speakers accumulate and update their phonotactic knowledge (Frisch, Pierrehumbert, & Broe, 2004), and thus, adapt to new constraints. In line with this view, Onishi and colleagues showed that infants as young as 16.5-month-old (Chambers, Onishi, & Fisher, 2003), as well as adults

(Onishi et al., 2002), are able to pick up novel phonotactic constraints from exposure to pattern-embedding syllables and to generalize these to novel instances. Learning these types of phonotactic generalizations is particularly relevant to the scope of the current thesis: Just as graphotactics represent a system of “rules” in written language (see Section 1.3), phonotactics represent a system in spoken language which can be considered “rule based” and which, as pointed above, generalizes to novel stimuli. The work introduced above provides good evidence that such knowledge is learned from novel instances through statistical learning mechanisms, enabling learners to recognize nonce words or syllables. This type of work is discussed further in Section 1.5.4.

The studies presented above demonstrate that infants have an impressive ability to pick up probabilistic information in the input. However, due to the artificial nature of this input, their real life validity has been challenged. For example, some studies failed to find the same effects as in Saffran et al. (1996) and Mattys et al. (1999) when they used words of varying (2- and 3-syllable) length (Johnson & Tyler, 2010). Others challenged the use of input that is stripped from real speech characteristics (e.g., Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). In order to measure how infants integrate the different word boundary information such as prosodic markers (e.g., typical stress pattern of words), Mattys et al. (1999) designed stimuli that included conflicting cues on phonotactics and prosody. They combined sequences embedding good prosodic but poor phonotactic cues with sequences embedding good phonotactic but poor prosodic cues at word boundaries. They showed that infants rely more on prosodic cues at 9 months old, while still sensitive to phonotactics. Studies that tested younger infants found that, while the computational abilities were not the only prerequisites in language acquisition, they might be deployed earlier than speech cues such as stress (Thiessen & Saffran, 2003). Another study supporting the ecological validity of SL when extracting lexical information is Pelucchi, Hay, and Saffran (2009) who showed that English-learning infants identify word boundaries from fluent unfamiliar but naturalistic speech (i.e., infant-directed Italian speech), using the same computations as in (Saffran et al., 1996). While experimental conditions in studies of SL may not be able to reflect the complex structure of naturalistic languages, they provide promising evidence that patterns that mirror those in natural languages are easier to learn. For example, Saffran and Thiessen (2003) showed that phonotactic constraints that are found in natural languages are easier to learn than those that are not. While patterns regulated by voicing are prevalent across languages (e.g., voicing assimilation of /s/ in /kats/ vs. /dogz/ in English), the restrictions on individual segments are not (e.g., legality of /pat/ and /dat/ but not /gat/ is unlikely) (see also Christianson & Devlin, 1997; Saffran, 2002; Saffran & Kirkham, 2018). In Saffran and Thiessen's (2003) study, infants (7-month-old) heard words that followed either a phonotactic regularity constrained by voicing (i.e., syllables begin with voiced but end with voiceless consonants; e.g., *bapgok* and *ditbap*) or a phonotactic regularity constrained by the position of individual phonemes (i.e., two voiced and one voiceless phonemes can begin syllables (e.g., /b/, /d/, or /k/) and two voiceless and one voiced phoneme can end syllables (e.g., /p/, /t/, or /g/)). Infants only learned the phonotactic patterns when voicing supported syllable position, taken to suggest that inconsistent sound patterns are harder to learn than consistent ones. Importantly, the authors suggested that the positional regularities presented to infants are unlikely to occur in natural languages, explaining the difficulty in learning them.

SL's involvement in linguistic processing goes beyond learning about the sounds of a language and vocabulary and is further supported by evidence that it plays an important role in semantic (e.g., Graf Estes, Evans, Alibali, & Saffran, 2007; Lany & Saffran, 2010) and syntactic (e.g., Saffran & Wilson, 2003) language acquisition. Graf Estes et al. (2007) presented infants with similar sound streams to those in (Saffran et al., 1996) and then asked them to link the newly learned words to objects, in an object-label association task. Infants were able to link familiar sound sequences with objects, but not the unfamiliar sounds, taken to suggest that the mere frequency of sound sequences was not sufficient for object-label learning, but this was dependent on higher-order, conditional probabilities (i.e., the "productiveness" of sound sequences given by their strong internal structure). In addition to the ability to incorporate statistical cues when learning word meanings, infants were shown to also use these cues to learn about the structure in which these words occur (syntactic structure). A discussion of such studies, that show SLs involvement in learning syntactic structure, is presented in Section 1.5.4.

The evidence presented above suggests that statistical learning mechanisms may play a role in learning of spoken language at many levels. However, further evidence suggests that this type of learning in the auditory domain is not language specific. Experiments that used non-linguistic stimuli and tested non-human primates demonstrate that SL is not just a linguistic mechanism. Saffran, Johnson, Aslin, and Newport (1999) showed that 8-month-old infants were able to detect TPs of continuous tone stream analogous to the speech stream used in Saffran et al. (1996). In Hauser, Newport, & Aslin (2001) study, adult cotton-top tamarins (a species of New World monkey) were able to extract the TPs defining the word boundaries in a similar manner as human infants. The monkeys showed interest in nonwords and part-words when exposed to the same set of auditory stimuli as in Saffran et al. (1996).

Statistical learning is not only a generally relevant mechanism within the auditory domain but extensive research has demonstrated that it is also part of visual learning, such as face identification (e.g., Althoff & Cohen, 1999). The next section further explores the role of statistical learning outside the language domain, reviewing several studies in the area of vision.

1.5.2 Visual statistical learning

Many of the studies that demonstrated that SL is implicated in learning non-linguistic structure in the visual modality (also coined as visual statistical learning or VSL) have utilized similar methods as in the auditory modality. In Kirkham et al. (2002), for example, a sequential visual (rather than auditory) stream of colored shapes was presented to 2-, 5-, and 8-month-old infants. The transitional probabilities embedded in the visual stream were the same as in Saffran et al. (1996), that is, each shape corresponded to a symbol in the previous study and the probability of certain shapes following one another (e.g., yellow circle after pink diamond) was high (1.00) while the probability of other, random shape pairs was low (0.33). Infants as young as 2 months old reliably showed preference for the low probability (random) pairs rather than high probability ones, which was interpreted as showing a novelty effect for unfamiliar items, indicating that they had learned the highly probable sequences. To ensure that infants indeed learned the conditional probabilities (TPs) and not the frequency of pairs (i.e., similar to Aslin et al., 1998), Marcovitch and Lewkowicz (2009) manipulated the two types of statistics independently. They found that 4.5-month-olds learned both the simpler statistics (frequencies) as well as the conditional probabilities. Thus, sensitivity to temporal visual statistical cues emerges very early on.

Further studies aimed to expand the scope of visual statistical learning research by assessing adults' sensitivity to visual statistical relationships. Adults followed a similar procedure as infants, that is, they were first familiarized (equivalent to habituation in infants) with pattern-embedding stimuli and then were asked to judge which one of the two simultaneously presented sets of stimuli was familiar. Fiser and Aslin (2002) exposed adults to a continuous stream of abstract shapes organized into four triplets—a procedure known as the Triplet Learning Paradigm—that embedded joint probabilities (i.e., the probability of co-occurrence, equivalent to transitional probabilities in the auditory modality) of shape pairs and triplets. They found that the participants were able to discriminate not only (i) between "base triplets"—where the probability of three successive shapes was high (i.e., .083), and "impossible triplets"—where the probability was zero; but also (ii) between "base triplets" and "part triplets", where the probability was lower (i.e., .027). To demonstrate that sensitivity to higher-order conditional probabilities is measured, rather than frequency of co-occurrence (similar to similar to Aslin et al., 1998; Marcovitch & Lewkowicz, 2009), Fiser and Aslin (2002) also presented the shapes with unequal frequency within a triplet, while maintaining the conditional probability (i.e., the probability of one shape being present, given another). They found that adults were able to extract the conditional probability statistics even when the frequency of their co-occurrence varied. This sensitivity was shown to play a crucial role in associative learning of novel information (e.g., Atick, 1992).

In the visual domain, humans extract features about objects, surfaces and scenes from temporal as well as spatial input. These can be "highly complex two-dimensional shapes, three- dimensional objects, and multiobject scenes" (Fiser & Aslin, 2005). Research in this area aims to uncover how such representations are formed and learned via statistical learning processes. Fiser and Aslin (2001), for example, created complex visual scenes by using 3 x 3 grids where they displayed the same shapes as in the study presented above (Fiser & Aslin, 2002). In this study, the shapes were organized into "base pairs" rather than triplets, with a high probability of co-occurrence (i.e., .50), and were arranged into horizontal, vertical and oblique orientations, resulting in 144 possible scenes presented to adults during a passive viewing exposure phase. The participants were able to learn not only the position-dependent co-occurrences but also position-independent co-occurrences: They reliably discriminated between "base pairs" and “non-base pairs” when each individual shape in the “non-base pairs” either (i) did not appear in the tested grid position, and thus providing reliable shape position and co-occurrence information, or (ii) it did, and thus eliminating the positional information. As in their 2002 study, Fiser and Aslin (2001) also manipulated the sequences by doubling the frequency of some of the “base pairs” at exposure, so that co-occurrence frequency and conditional probabilities were not correlated. As a result, they showed that participants were sensitive to the predictability between shapes, that is, the conditional probability, while maintaining sensitivity to other, first-order statistics, such as frequency of single shapes. Such tests of visual statistical learning, particularly the Triplet Learning Paradigm, have been replicated and extended in more recent studies such as Arciuli and Simpson (2011, 2012), which used aliens instead of real or abstract shapes.

A premise of SL research is that learning of patterns described above proceeds in an intuitive, unconscious, or incidental way, without awareness of what is learned, often described as “automatic” (Fiser & Aslin, 2002, p.458), “spontaneous” (Fiser & Aslin, 2001, p. 502) or “as a byproduct of mere exposure” (Saffran et al., 1999, p. 30). Turk-Browne, Jungé, and Scholl (2005) aimed to test this assumption by adding an attentional dimension to the Triplet Learning Paradigm. They used the same shapes as in Fiser and Aslin (2001, 2002) but the triplets created were either green or red. The manipulation of attention was achieved by interleaving the red and green streams into one long stream and asking the participants to monitor just one of the streams of shapes (either green or red). Adults reliably discriminated between “base triplets” and foils (i.e., impossible or random triplets) only when the shapes tested were part of the ones presented in the “attended” color, regardless of how testing was manipulated by presenting all items in (i) black, (ii) original colors, or (iii) swapped colors. The same results were found even when performance was tested with an online reaction time task, rather than the forced choice task, so that judgments about implicitly leaned patterns were not measured with an explicit task. Such results were taken to suggest that visual statistical learning is not necessarily data driven but is influenced by selective attention. Nevertheless, participants were not able to report awareness of the relationships learned, in line with implicit learning studies in other domains and using different paradigms (see Stadler & Frensch, 1998 for a review). Furthermore, design features of studies reviewed in this section, such as the speed of stimulus presentation, brief exposure (usually 6 minutes) and passive viewing, and sometimes under distractor tasks, ensure that VSL in an implicit process.

This section has shown that many of the methods used in the experiments described in this section are similar to those carried out with spoken language (see Section 1.5.2). Thus, the literatures focusing on the two modalities look to answer similar questions regarding the types of statistics computed over auditory and visual stimuli. While some of the studies explore visual learning over sequentially presented stimuli, others also look at statistical learning across the spatial elements of static array. This latter dimension is not relevant for auditory learning, or spoken language, but is potentially important in the area of learning visual pattern in written language—that is, graphotactics, the focus of this thesis. However, the studies reviewed so far (in Section 1.5.1 and 1.5.2), in both the auditory and visual domain, have only tested infants' and adults' ability to extract patterns from trained items. However, we know that human leaners are capable of generalizing over linguistic patterns, both in spoken language and—as shown in Section 1.3—for graphotatic patterns in written language. In the next section, I present statistical learning studies that test generalization by using novel items that embed trained relationships.

1.5.3 Statistical learning and generalization

Considering that infants, children and even adults come across novel information throughout their learning experiences, we know that recognizing specific items is not sufficient in natural language acquisition. Human language has a generative power (Hockett & Hockett, 1960) and the ability to perceive relationships among categories of words in a sentence, for instance, plays a critical role in this productivity (Gómez & Lakusta, 2004). A large body of research, therefore, aimed to answer the question: Can the same SL mechanisms that enable infants to extract lexical information described in Section 1.5.1 support the acquisition of abstract language structure (e.g., Gerken, Wilson, & Lewis, 2005; Gomez & Gerken, 1999; Gómez & Lakusta, 2004)? In other words, can infants generalize to novel, untrained items using distributional cues such as transitional probabilities? One way of exploring category-based abstractions is by means of an artificial grammar learning paradigm (AGL). During studies using this paradigm, participants are typically exposed to a finite-state grammar that incorporates complex strings, albeit limited in their generative power (Chomsky, 1957), where dependencies can be defined in syntactic phrases. Given that such grammars can generate a range of strings, experimenters are able to assess generalization by presenting a subset of these in the training or exposure phase and then asking participants to discriminate between new grammatical and ungrammatical strings at test. Gómez and Lakusta (2004), for example, created two languages by combining two sets of function words, similar to those found in English (e.g., *the* and *a* can precede nouns but not verbs, whereas *will* and *can* precede verbs but not nouns). The resulting languages comprised of two "a-elements" (e.g., *alt* and *ush*) and two "b-elements" (e.g., *ong* and *erd*) and two sets of "category" words (six disyllabic "X-elements", e.g., *coomo*; and six monosyllabic "Y-elements", e.g., *deech*). In Language 1, the "a-elements" were paired with "X-elements" while "b-elements" were paired with "Y-elements" and vice-versa in Language 2. Only a subset of all strings were presented auditorily to 1-year-olds at exposure, while at test, infants were presented with a- and b-elements identical to the ones heard at exposure but paired with the novel, untrained X- and Y-words. Infants not only categorized the X and Y-elements based on their syllable number, they also generalized to novel elements and successfully discriminated between legal and illegal "marker-feature pairings". That is, they identified word categories (e.g., nouns, verbs, function words) through phonological regularities.

The implications of findings such as the ones presented so far have been challenged on grounds that the acquisition of language's complex grammar requires mechanisms beyond statistical learning (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Peña, Bonatti, Marina, & Mehler, 2002), such as the ability to abstract "algebra-like" rules. Gomez and Gerken (1999) demonstrated that 1-year-olds can also learn the predictive relationships between these categories, that is, abstract beyond specific pairs of items when all test items are novel, not just one. In one of their experiments (Experiment 4), infants were exposed auditorily to grammars composed of lists of 3- to 6-syllable strings in grammatical "sentences", such as VOT-PEL-JIC or PEL-TAM-JIC-RUD-TAM-JIC. In this grammar, the legality of endpoints was set such that the sentences could only begin with one of two possible words (e.g., VOT and PEL) and end with one of three (e.g., JIC, RUD or TAM) possible words. A test grammar was created by replacing the entire vocabulary in the exposure grammar but maintaining the sentence structure (i.e., sentences could only begin with one of two possible words and end with one of three possible words). This was created to test generalizations, by mapping every occurrence of a syllable (in the first grammar) to a new item (e.g., VOT was mapped to JED, PEL was mapped to FIM, and so on). Hence, at test, infants heard JED-FIM-TUP or FIM-SOG-TUP-DAK-SOG-TUP (grammatical) and JED-DAK-TUP or JED-DAK-TUP-FIM-JED-DAK (ungrammatical). Infants accurately discriminated between grammatical and ungrammatical strings (i.e., showed interest to the illegal sentences). Since all test items were novel and the resulting transitional probabilities between word pairs were zero, the results show that infants extracted a more abstract form of information, an ability considered to be a milestone in cognitive and language development, with a crucial role in syntax acquisition (Chomsky, 1957).

Peña et al. (2002) further showed that generalizations arise only when subliminal cues to word segmentation are added to the stimulus. They first demonstrated that adults segment a continuous stream of "words" in a similar procedure as Saffran et al. (1996). They exposed French-speaking adults to a continuous stream of "words" embedding non-adjacent TPs in trisyllabic items. They called the new grammar "AXC language", where A predicts C and is combined with three different Xs to create a family of words (e.g., *pu-li-ki*, *pu-ra-ki*, *pu-fo-ki*). Next, they aimed to assess whether such learning could be explained by acquiring a structural regularity rather than sensitivity to distant TPs, that is, could adults have generalized the rule "if A occurs then C will follow after an intervening X"? Participants were presented with "rule words" and part-words at test, instead of "words", that is, the intervening X syllable would not have appeared between A and C at exposure. Adults' failure to choose the "rule words" over part-words was taken to suggest that the computational mechanism that enabled participants to segment the stream does not support the detection of a structural regularity. Participants were only able to discover the underlying rules when brief pauses were added at word boundaries, making the job of segmentation (attributed to statistical learning) redundant. These findings held even when the exposure was (i) increased when a continuous speech stream was presented and (ii) decreased when brief pauses signaled boundaries. Peña et al. (2002) concluded that, while both statistical processes (e.g., sensitivity to distributional probabilities) and grammatical processes (e.g., rule learning) are involved in language acquisition, only the latter are powerful enough to enable the acquisition of the complexities of grammar, beyond word segmentation and building a lexicon.

However, Seidenberg, MacDonald, and Saffran (2002) provide an alternative interpretation. They suggest that the line between rule and statistical generalization can be blurred, and many of the constraints described in the studies reviewed in this section could be labelled as either rules or statistics. For example, in Peña et al. (2002), participants' responses were taken to have resulted from sensitivity to non-adjacent TPs (e.g., *pu* predicts *ki* after an intervening x such as *li*). However, as Seidenberg et al. (2002) pointed, the input offered an array of statistical cues that could have resulted in potential generalizations that correlated with the manipulated cues (e.g., "*pu* is not followed by *be*", "*ki* predicts *be*" and so on). Furthermore, the above studies do not provide a theory regarding the mechanisms required to arrive at the correct generalizations. Aslin and Newport (2012) offer an alternative hypothesis and suggest that, both specific item recognition and generalization over untrained "rule" items are the outcomes of a single, statistical learning mechanism. They suggest that the dimensions of a stimulus that are salient are first extracted, reducing the ambiguity of the input, and these are then generalized to all stimuli that share the patterns on these salient dimensions. While these dimensions can be perceptual, such as auditory pitch or visual features (e.g., falling objects), in language learning, these dimensions are functional, with no perceptual basis. When more elements (e.g., nouns) occur in the same context interchangeably, then a rule is acquired about all those elements, whereas, when the patterns apply to individual elements, specific instances are acquired. For example, Xu and Tenenbaum (2007) showed that, when preschool children heard one label for three objects (e.g., *fep* for pepper, carrot and onion) they were more accurate and confident about the correct hypothesis for generalization (e.g., they are all vegetables), compared to when they heard a label repeated three times for the same object (e.g., *fep* for red pepper). Such results were explained by sensitivity to statistics beyond label and object co-occurrence and provide a Bayesian account of learning, that is, children keep track of the number and variability of instances labelled to make statistical inferences when learning word meanings. In other words, children made inferences by evaluating the evidence in the sample: When they heard the label applied to the same object three times, they acquired a specific instance, but when they heard the label was applied to three more varied instances, there was evidence of broader learning (rule acquisition).

While the studies above looked at learning of purely form-based syntactic rules, in natural languages, syntactic rules operate over meaning. Research exploring the relationship between syntax and semantics showed that infants' experience with statistical cues in the auditory input influences their ability to subsequently acquire semantic properties of category members (Lany & Saffran, 2010). 22-month-old infants listened to grammars similar to those in Gómez and Lakusta (2004) and saw pictures of either vehicles or animals. When tested on novel phrases (i.e., same a- and b-elements as in training but different X- and Y-words) and novel pictures belonging to the two categories learned at exposure (i.e., vehicles and animals), they were able to generalize the meaning of individual words to novel category members. This was not true for a control group that received training on strings where category membership was not cued.

In addition to an ability to generalize to novel stimulus, language learners also need to contend with learning exceptions and arbitrary restrictions on the structure of natural languages, in order to avoid overgeneralization errors. Wonnacott and colleagues (Wonnacott, 2011; Wonnacott, Brown, & Nation, 2017; Wonnacott, Newport, & Tanenhaus, 2008) demonstrated that learners also generalize above the lexical level, that is, above distributional statistics (e.g., item co-occurrences), and this learning depends on the statistical properties of the input. Wonnacott (2011), for example, created two artificial languages where the relationship between items was manipulated. 6-year-olds were presented auditorily with sentences composed of particle-noun-particle, where the noun was an English label for an animal (e.g., *pig*, *giraffe*) while the particles were novel function words (e.g., every sentence started with *moop*, meaning "there are two" but ended with either *dow* or *tay*, labelled particle1 and particle2, with no semantic meaning). Children also saw pictures of two identical animals on the screen. The frequency of the end particles was manipulated such that one was more frequent in both languages. However, the relationship between nouns and end particles was different in the two languages: (i) in one language (Lexicalist), each noun occurred consistently with one particle (e.g., *dow* followed giraffe and rabbit and *tay* followed pig and cow); while in the other language (Generalist), both particles could occur with each noun. Children were asked to produce sentences while looking at pictures that were either familiar (i.e., from the input) or entirely novel (i.e., occurred only at test). When producing sentences for pictures containing familiar nouns, children's responses approximately matched the input, that is, for the Generalist language they used both particles with each noun, whereas in the Lexicalist language they restricted particles to particular nouns. However, children were also tested on nouns referred to as minimal-exposure nouns. These were nouns that occurred only during the test session and were presented four times, always with the same particle. Learners previously exposed to the Lexicalist and Generalist conditions treated these very differently: In the Lexicalist condition, children continued to use the noun only with the particle with which it was presented in the minimal input, whereas in the Generalist condition they were more likely to use both particles, using the overall dominant particle more frequently. This was taken to suggest that, when generalizing, children use frequency statistics but this generalization is balanced using item-specific co-occurrences when such information is reliable (e.g., in the Lexicalist language). Thus, such behavior is the result of a balance between generalization and item-specific distributional statistics, that is, prior learned expectations, consistent with the Bayesian inference account (e.g., Wonnacott et al., 2008). As mentioned in Section 1.5.2 when discussing learning of phonotactic patterns, it was suggested that learners build expectations about distributions (relationship between particles and nouns, in the case of Wonnacott, 2011) by tracking structures in their input to then make inferences about the reliability of these distributions for future usage.

1.5.4 Learning phonotactic constraints

While generalization in language is often discussed in terms of relevance to learning of syntactic and morphological patterns, it was also shown to be relevant to learning of rules over sound patterns, that is, phonotactics. These rules about phonemes, both deterministic and probabilistic, can occur within words of a language, and naïve speakers are aware of them, that is, they can discriminate between possible nonwords and those that do not follow the rules (Jusczyk et al., 1993), and their errors respect these rules (Stemberger, 1985). Jusczyk et al. (1993) showed that 9-month-old infants listened longer to words with sound patterns that were legal in their own language. They presented English and Dutch infants with low frequency bi- or tri-syllabic English and Dutch words and tested their sensitivity to the phonotactics in their language using the Head-Turn Preference Procedure. The items used in one language contained patterns that were not permissible in the other language, thus the phonotactics were not comparable. For example, in Dutch, words do not end with *d*, while in English they do. Similarly, Dutch syllables can begin with a sequence like *kn* (as in *knoest*) or *zw* (as in *zweten*) while English words don't. Both English and Dutch children listened longer to words with sound patterns that were legal in their native language. Onishi et al. (2002) argued that acquiring phonotactic regularities such as the ones described above (i.e., certain sound representations cannot begin or end syllables or words) are not absolute (i.e., deterministic) but must take context into consideration. For example, the sequence /ʌ/ occurs with /f/ (/ʌf/) as in *bluff,* more often than expected, and /æ/ with /l/ (/æl/) as in *pal,* less often than expected, given the frequency of the individual sounds (Kessler & Treiman, 1997). Thus, the restrictions on phoneme position and co-occurrences can be explained in statistical, probabilistic terms. Bayesian inference provides a neat account for learner's ability to process new input: By updating their phonotactic expectations with increasing experience with spoken words, learners accumulate perceptual information that, in turn, alters their phonotactic expectations (Church & Fisher, 1998). Learning such patterns requires abstractions that are flexible, that is, they retain the detailed contextual information, as seen in Wonnacott (2011) above. Dell, Reed, Adams, and Meyer (2000) tested such generalization in a novel word production study, where they embedded novel phonotactic regularities in four-syllable (CVC) sequences (e.g., /gɛŋ/, /kɛg/, /hɛm/, /nɛs/). In one experiment, particular consonant sounds (e.g., /f/ and /s/) were restricted to a particular position (either beginning or end of syllables, counterbalanced between participants), across all the words in the exposure set (what they refer to as a *language-wide* constraint), while for other sounds (e.g., /n/, /m/, /g/) their position was *item specific* (e.g., /fɛn/, /ges/). Other regularities respected English phonotactcics, that is, /h/ could only begin and /ŋ/ could only end syllables. Adults were asked to repeat the syllables at a fast pace (determined by a metronome) following visual presentation of words, and their productions were analyzed for errors. Previously, speech errors have been shown to respect the phonotactic constraints in the learner's language and therefore, the likelihood of producing illegal structures was considered to be low (Stemberger, 1983). For example, if they were to missorder the sounds in "left hemisphere" to produce "heft lemisfere", English speakers do not violate the phonotactics, but if they were to missorder /ækt/ to produce /ætk/, they would (Dell, 2007). Analyzing speech errors enables researchers to investigate how phonotactics are used during speaking. In line with this, Dell et al. (2000) found that adults' speech errors conformed to the phonotactics of the artificial lexicon used in the experiment following just one training session: They very rarely (2% of the time) produced (i) impossible sound sequences that violated the language-wide constraints (i.e., position of /f/ and /s/ was obeyed), and (ii) sequences that violated the item-specific (local) positional constraints, albeit to a larger extent (32% of the time). This provided additional support for the flexible nature of phonotactic abstractions. In another experiment in the same study, Dell et al. (2000) demonstrated that, in addition to positional constraints, context-based constraints are learned. They conditioned the position of the consonant sounds to the presence of a particular middle vowel (e.g., /f/ at the beginning of word if followed by /æ/). Such learning, however, was shown to become reliable only after two days of training (Warker & Dell, 2006), suggesting that representations of context-based constraints require more time to form. The evidence presented by speech error studies was taken to demonstrate that, when learning about the distribution of speech sounds, infants, children and adults detect implicitly the regularities in their language and use this knowledge to generalize. Dell et al. (2000) referred to this mechanism as implicit learning hypothesis.

A different technique for investigating phonotactic learning and generalization is provided by the work of Onishi and colleagues introduced in Section 1.5.1. In a series of learning studies they demonstrated that infants and adults learn novel phonotactic patterns from brief listening experience (Chambers, Onishi, & Fisher, 2010; Onishi et al., 2002; Seidl, Cristià, Bernard, & Onishi, 2009). For example, in a two-phase task, Onishi et al. (2002), first exposed adults to a stream of nonce CVC syllables where the position of consonants was restricted (e.g., /fɪp/, /f/ and /p/ were constrained to syllable-initial or syllable-end positions, respectively). In a second, test phase, adults were asked to listen and repeat test items "as quickly as possible". The test items consisted of both (i) studied syllables and (ii) unstudied (novel) syllables that either conformed to- or violated the positional constraints presented at exposure. Adults were faster at repeating the novel syllables that conformed to the newly learned constraints (e.g., /fɪp/, where /f/ was permissible onset) than those that violated them (e.g., /pɪf/, where /p/ is not permissible onset). As in Dell et al. (2000), Onishi et al. (2002) also tested learning of context-based constraints and found that adults were also faster at repeating novel syllables that conformed with the constraints that conditioned consonants' position to the adjacent vowels (e.g., /b/ can be an onset when the following vowel is /æ/ but not /ɪ/). Furthermore, the responses for the novel test items were similar to the studied items throughout the study, indicating that learning occurred rapidly and was not merely reflecting familiarity with sequences heard at exposure.

Evidence presented in Section 1.5.1 shows that not only adults, but also infants learn sequences of whole syllables from brief exposure to pattern-embedding auditory stimulus (e.g., Aslin et al., 1998). Some research has also pointed towards their ability to generalize across syllables (e.g., Hollich, Jusczyk, & Luce, 2001). Chambers et al. (2003) showed that 16.5-month-old infants learned the phonotactic regularities following familiarization with the same positional regularities presented to adults (Onishi et al., 2002). That is, they listened longer to novel syllables that violated the newly learned constraints, and their ability to generalize was observed within the same timeframe as adults. However, Chambers, Onishi, and Fisher (2011) failed to find the same effect with context-based constraints, concluding that more complex regularities are harder for infants (10.5-month-old) to learn in such short timeframe.

A question arising from the above described work is how abstract are the positional phonotactic representations? When presented with stimuli that embed positional constraints, did participants learn the "rule" that /b/ is a permissible onset, or did they learn that /b/ precedes /æ/? Both might be true from the input, so it is difficult to determine what type of generalization was formed. To test if redundant contextual information interfered with learning of positional constraints, Chambers et al. (2010) familiarized adults with CVC syllables containing consonants with restricted positions and one of two intervening training vowel (e.g., /fæs/, /bɛs/) but tested generalization to novel syllables that contained a different (transfer) vowel (e.g., /fɪs/, /bɪs/). Participants were faster at repeating the syllables that conformed to the trained constraints than those that violated them (i.e., position of consonant was illegal, e.g., /sæf/, /sɪb/), and this effect was similar regardless of the vowel used (studied or transfer) and acoustic similarity between the two types of vowels (e.g., similar: /æ/ vs. /ɪ/; dissimilar: /ɛ/ vs. /ʊ/).

Thus, the vowel context did not interfere with learning, taken as evidence that the phonotactic generalizations were abstracted on position.

Learning phonotactic generalizations is particularly relevant to the scope of the current thesis. Just as graphotactics represent a system of “rules” in written language (see Section 1.3), phonotactics represent a system in spoken language, which can be considered “rule based” and which, as pointed above, generalizes to novel stimuli. The next section presents studies that have directly assessed graphotactic learning and generalization.

1.6 Statistical learning applied to graphotactic rules

The work reviewed in the previous section demonstrates that adults and infants can use statistical learning to pick up patterns in spoken language, as well as visual patterns. Both of these are potentially relevant in learning to spell. Samara and Caravolas (2014) provided a first demonstration of graphotactic learning under incidental experimental conditions, extending the methodology used in studies that explored phonotactic learning (Section 1.5.4). They created a child-appropriate task whereby 7-year-olds and adults saw CVC nonwords and were told that they were going to play games with words from an alien language. There was no mention of any patterns or rules; however, all of the stimuli embedded novel patterns. In one experiment, the consonants were constrained to certain positions (e.g., in one counterbalanced list, *l* was restricted to onset position, and *p* to coda, e.g., *lep*) while in another condition the consonants' position was conditioned by the identity of the middle vowels, in both word beginnings and ends (e.g., participants saw stimuli like *tof*, which, *unbeknown* to them exemplified “*t* can be followed by *o*, never *e*” as well as “*o* can be followed by *f*, never *t*”). Generalization was tested immediately following exposure, using legality judgments: Participants were shown novel stimuli that either conformed to or violated the trained patterns and were asked to decide if these could be part of the language they were exposed to. Both children and adults were able to discriminate between legal and illegal stimuli in both conditions: positional (e.g., *les* vs. *sel*), and contextual (e.g., *tog* conforming to the above pattern vs. *teg* that did not). The positional constraints however, were learned easier than the contextual ones by both children and adults, taken to suggests that learning is modulated by pattern complexity, in line with evidence from studies that investigated the acquisition of spelling patterns from natural language (e.g., Caravolas, Kessler, Hulme, & Snowling, 2005; Cassar & Treiman, 1997; Pacton et al., 2001; see also Section 1.3). Nevertheless, even the 7-year-olds in Samara and Caravolas' (2014) study generalized over the more complex (first-order) contextual graphotactics, albeit a very small effect.

Later work has replicated the learning of both positional (Nigro, Jiménez-Fernández, Simpson, & Defior, 2015) and contextual constraints (Samara, Singh, & Wonnacott, 2019), in similar implicit learning tasks. Nigro et al. (2015) found that 8–9-year-old Spanish-speaking typically developing children were able to learn the positional constraints embedded in four-letter pronounceable strings (e.g., "stimuli only start with *m*, *l*, *t* — never with *f*, *n*, or *s*", as in *mifo*). Samara et al. (2019) replicated and extended Samara and Caravolas' (2014) findings on contextual constraints, showing that English- and Turkish-speaking children could learn patterns which were conditioned on *either* word-beginning or word-ending positions (rather than both), thus demonstrating learning in more naturalistic stimuli. This study also attempted to investigate whether learning was stronger in either word-initial (CV) or word-final (VC) positions in light of the debate on the relative importance of word-final units (referred to as rimes, i.e., the unit containing the vowel and word-final consonant/s) in literacy development: According to a well-regarded view of spelling development (Goswami & Bryant, 2016; Treiman & Kessler, 1995), learning is stronger when word-final consonants are conditioned on the preceding vowel (orthographic rime). In fact, Samara et al. (2019) did not find such evidence, that is, discrimination between novel pattern-conforming and pattern-violating stimuli in word-beginning was not substantially different from discrimination in word-ending units, as measured by Bayes factor statistics. However, what is critical for the current purpose is to note that substantial learning was found in each of onset (CV) and coda (VC) separately.

Evidence from studies using well-controlled experimental conditions, such as those described above, contributes to evidence from studies that used experience with natural languages (Kessler, 2009; Pollo et al., 2007; Steffler, 2001; Treiman & Kessler, 2013; see Section 1.3) and demonstrates that statistical learning processes can be applied to the learning of spelling patterns. This supports a view in which this learning plays a role in literacy development, and, at least to some extent, underpins the learning of spelling patterns.

However, there is an important limitation in this literature: In all of the experiments described above that used pronounceable alphabetic letters, although learning effects have been interpreted as graphotactic in nature, the contextual constraints had a phonological counterpart, that is, phonotactic constraints such as, "/t/ can only follow /o/, never /ɛ/", alongside graphotactics ("*t* can only follow *o*, never *e*"). Thus, although the participants did not hear the words and were not asked to pronounce them overtly, they may have accessed the pronunciation covertly and therefore learned the patterns as sound based—as phonotactic constraints. It is also possible that some participants could have read the stimuli aloud, since this was not prevented in any of the experiments by Samara and colleagues, nor in the study by Nigro et al. (2015). It is, thus, possible that *phonotactic sensitivity* of the type operating in spoken language contributed to, or even entirely drove, the learning effects. This possibility is especially supported by evidence from studies carried out by Onishi and colleagues (e.g., Chambers et al., 2003; Onishi et al., 2002) discussed in the previous section (1.5.4) that clearly demonstrate phonotactic learning in similar experimental conditions. However, as discussed in Section 1.3, not all orthographic patterns have a phonological counterpart. Although phonotactic constraints are reflected in natural language orthographies (e.g., in English, words do not begin with /ŋ/ and accordingly, written words cannot have *ng* in beginnings), there are also orthographic constraints that, while conditioned on phonology, do not reflect phonology in a direct way (e.g., /ɛ/ is particularly likely to be spelled as *ea* when the word ends in /d/; e.g., *dead*), and some orthographic patterns are purely visual (graphotactic) in nature (e.g., *dd* does not begin written words, /d/ does; e.g., *ddoll; doublets occur more often after a single– than double-letter spelling; e.g., *Jeff* vs. *deaf*; *bedding* vs. *heading*).

To date, only only a handful of studies have used fully artificial systems (Chetail, 2017; Lelonkiewicz, Ktori, & Crepaldi, 2020; Nigro, Jiménez-Fernández, Simpson, & Defior, 2016; Y. Vidal, Viviani, Zoccolan, & Crepaldi, 2021) but only two have used pseudoletter strings or child participants to test the learning of graphotactic patterns. Both of the latter moved to using unfamiliar—and thus unpronounceable—symbols in place of familiar graphemes. Chetail (2017) used stimuli comprised of characters unknown to the subjects (Phoenician Moabite alphabet letters) to investigate *adults*' learning over the distribution of co-occurring letters, in familiar or novel positions within the stimuli. As noted above, Samara and Caravolas (2014) found that such contextual (first-order) constraints are harder to learn for both children and adults, than positional (zero-order) ones. Nevertheless, Chetail (2017) found above-chance discrimination between patterned and random character sequences in a wordlikeness task (whereby participants were asked to judge which of two stimuli was more like the words seen in a previous exposure phase). She also found evidence that frequency-based learning is represented in the visual word recognition system as seen by participants' faster reaction times for high-frequency relative to low-frequency characters in a speeded detection task. However, as the study was only carried out with adult participants, the relevance for typical literacy development has not yet been established. There is no similar work with children, presumably due to the increased demands of using unfamiliar symbols that necessitate longer training. However, Nigro et al. (2016) tested 8–9 year old typically developing as well as dyslexic children on both linguistic and non-linguistic *positional* regularities. They designed non-linguistic stimuli of the same structure as the letter strings in Nigro et al. (2015) by replacing each letter with an unfamiliar shape and presented them in one study, while keeping the linguistic stimuli in another. They aimed to (i) determine whether linguistic properties of the material moderate learning effects and (ii) if dyslexic children differ from typically developing ones in their implicit learning ability. Since the constraints were the same as Nigro et al. (2015), they were positional, that is, certain symbols were restricted to certain positions (beginning or end) in the sequences. They found that both groups (typically developing and dyslexic children) learned the novel graphotactic constraints, that is, their performance did not differ when correctly identifying the training (i.e., studied) items. However, dyslexic children were unable to generalize over novel items, taken to suggest that transfer of such knowledge is more challenging for them. While this study did show that children can learn purely visual patterns, the shapes, albeit still symbols, were relatively more familiar than the Phoenician Moabite alphabet used in Chetail (2017). Most critically, however, the constraints were positional rather than contextual, and, as Samara and Caravolas (2014) demonstrated, they are substantially easier to learn.

With knowledge of this gap in the literature, one of the goals of this thesis was to establish that children can implicitly learn contextually conditioned constraints which have no phonological counterpart. Along with this goal, the thesis aimed to (i) look at the relationship between implicit learning in these experiments and measures of literacy, and (ii) to compare this type of implicit learning with learning under explicit conditions. In the next two sections, I review literature relevant to these final two goals.

1.7 Does performance in statistical learning tasks predict measures of language and literacy

1.7.1 SL and language

The evidence reviewed above demonstrates that children extract statistical regularities from carefully designed stimuli and, combined with infant and adult spoken language acquisition research, clearly demonstrates that statistical learning is a ubiquitous human ability. This raises the intriguing possibility that differences in this ability may underpin individual differences in language and literacy development. Evans, Saffran, and Robe-Torres (2009), for example, showed that children with specific language impairment (SLI) significantly underperformed typically developing ones in a task similar to Saffran et al. (1996), when they were presented with nonsense speech in continuous stream. Importantly, this performance was also seen in a non-linguistic task, where children were presented with tones following the same transitional probabilities as those in speech (see Section 1.5.1 for a review of methods used in investigating auditory statistical learning). This demonstrates that these processing difficulties are not restricted to speech.

This link has received much recent attention, particularly for oral language skills such as language comprehension (Conway, Bauernschmidt, Huang, & Pisoni, 2010), relative clause processing ability (Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010), and syntactic ability (Kidd, 2012).

Turning to literacy, Arciuli and Simpson (2012) were the first to report a positive correlation between reading ability (as measured using the standardized WRAT test; Wilkinson & Robertson, 2006) and performance on a triplet learning paradigm which embedded temporal relationships among visually presented “alien” figures in children (5–12-year-olds) and adults. A regression analysis partialling out age and attention during the learning phase of the experiment revealed that statistical learning accounted for a small but significant amount of unique variance in reading ability. Positive associations between a broad range of statistical learning tasks and three core early literacy predictors (oral language skill, vocabulary knowledge, and phonological processing ability) were also seen in a much larger study with over 500 4.5–10.5-year-olds (Spencer, Kaschak, Jones, & Lonigan, 2015). Consistently, Frost and colleagues (Frost, Siegelman, Narkiss, & Afek, 2013) found correlations between performance on a visual statistical learning task and English-speakers' developing reading performance in a second language (Hebrew). As highlighted in Section 1.2.2, spelling is not just the reverse of reading and Steffler's (2004) study is among the few to show a correlation between performance on the implicit task (e.g., the consonant preceding *t* doubles—*gosst*, but not when preceding *k*—*dafk*) and spelling ability. When found, however, such correlations are typically weak to medium ($r < .4$).

1.7.2 Issues with measures and correlations

Correlations between measures of implicit or statistical learning and language attainment are not consistently found (Nigro et al., 2015; Schmalz, Moll, Mulatti, & Schulte-Körne, 2019; West, Vadillo, Shanks, & Hulme, 2017). West et al. (2017) for example, examined aspects of declarative (explicit) and procedural (implicit) memory in 7–8-year-olds and compared this performance with language, literacy, and mathematical abilities. They found that only the declarative and not procedural learning measures correlated with literacy and language attainment and this was explained by low split-half test-retest reliabilities (ranging from 0 to .24) for three measures of implicit learning (as opposed to the good reliability in explicit versions of the same tasks). At least in part, null findings may be an artefact of poor statistical properties (e.g., low internal consistency and reliability) of the statistical learning measures (Krishnan & Watkins, 2019; Siegelman, Bogaerts, & Frost, 2017). Siegelman et al. (2017) argued that most lab-based learning experiments are designed to capture group-level effects, thus, are, by design, unlikely to capture patterns of association: They feature few trials, the level of difficulty is intentionally restricted across test items, and they often suffer from floor effects (i.e., many participants performing at chance levels).

Inconsistent patterns of correlations are not only found with domain-general tasks as those discussed above, but also with measures of children's naturalistic orthographic sensitivity. Treiman and Boland (2017), for example, found that good spellers (according to their WRAT spelling performance) were more likely to double medial consonants in a nonword spelling task measuring graphotactic sensitivity, which, by one interpretation, may be due to their preference for more frequent letter strings in their orthography. Steffler (2004) tested 10–11-year-olds' knowledge of an untaught spelling rule for consonant doubling when adding *ed* to verb stems, and compared this with performance on an artificial grammar learning task (AGL; see Section 1.5 for a description). They found that memory performance (i.e., correct identification of trained items) but not generalization (i.e., transfer of trained patterns to novel letter strings) correlated with *ed* word reading as well as spelling ability (r = .21). Ise et al. (2014), on the other hand, found no concurrent or longitudinal relationship between their measure of orthographic knowledge (nonword choice task) and reading/spelling ability in German-speaking children.

In sum, it is hard to establish whether statistical property confounds may have concealed a relationship between statistical learning and literacy development, or whether the null findings suggest that there is no fundamental link between variations in the ability to detect statistical structure and literacy acquisition. Findings in this literature may also be distorted by publication biases and false positive (type I) results, both of which are pertinent concerns for psychology and other social sciences (Camerer et al., 2019; Rosenthal, 1979). Finally, when null findings *are* reported, they should be treated with caution for an additional reason: In the frequentist analyses which are typically applied, a nonsignificant result *cannot* be taken as evidence for the null: There may be evidence for the null or no evidence for any conclusion at all (or indeed evidence against the null). Yet researchers routinely take a nonsignificant result to indicate that they should reduce their confidence in a theory that predicts a difference. Bayes factors, on the other hand (Dienes, 2008, 2015), can provide quantifying evidence of H_0. To date, only one study (Schmalz et al., 2019) has utilized this method of analysis to establish whether nonsignificant associations between statistical learning and literacy skill are substantial evidence against theoretical accounts that predict a relationship or an artefact of data insensitivity. The experiments reported in this thesis employ Bayes factors to evaluate evidence for, and, crucially, against hypotheses and this metric is used both for evaluating performance in our learning tasks and—for all of the experiments conducted with children and some of the experiments conducted with adults—relating learning to their performance on standardized measures of literacy. For some of the experiments, “explicit” versions of the implicit learning experiments are also included and, to anticipate the results, some interesting differences emerged concerning how performance related to measures of literacy. In the final section of this introduction, I present a brief review of the literature on explicit teaching of spelling rules and further discuss this in Chapter 5 (Section 5.1.2).

1.8 Explicit teaching of spelling rules

As argued in Sections 1.1 and 1.2, the relationship between written and spoken language is complex. From a cognitive perspective, unlike for spoken language, humans do not have a dedicated instinct for reading and writing and thus, they must utilize cognitive general mechanisms (DeHaene, 2009). When learning a spoken language, most relevant rules and structure are picked up implicitly, well before starting formal instruction in school, from infancy. It is well established that explicit instruction plays a very limited role in the development of first language, and various grammatical and sound-based regularities are learned without explicit instruction (see Section 1.5.1). For example, some phonotactic “rules” such as “/ŋ/ cannot begin an English word” are not explicitly taught but infants demonstrate learning of such patterns. In contrast, learning the structure of written language starts when children enter formal education and are exposed to extensive literacy instruction (e.g., "*i* before *e*, except after *c*"). Hence, the term *acquisition* is primarily associated with spoken language while the term *learning* is associated with written language.

Much of the early literacy curriculum in alphabetic languages focusses on phonics instruction which builds on the premise that each letter (or group of letters, e.g., *sh*, *tch*) has a corresponding sound (phonemes, e.g., /ʃ/, /ʧ/). As established in Section 1.1, many features of written language do not have a corresponding feature in spoken language, and almost every alphabetic system has multiple spellings for the same phoneme (see Section 1.1.1 for examples). While some of the irregularities are taught through “rules”, not all can be explained in this way and taught explicitly. For example, “the rabbit rule” teaches children that, in two-syllable words, they need to spell the middle sound with a double consonant if the preceding vowel is short, as in *rabbit*; and with a single consonant if the vowel is long, as in *tiger* (Treiman, 2018). However, some words deviate from the rabbit rule, such as *comic* and *valid*, where, even though the vowel of the first syllable is short, the consonant is not doubled (Treiman &

Boland, 2017). Furthermore, as shown in Section 1.3.4, doubling of consonants also occurs in a graphotactic context, where the preceding vowel phoneme is spelled with multiple letters (Hayes et al., 2006). The fact that rules are probabilistic and have exceptions suggests that they are unlikely to be mastered entirely through explicit instruction. While traditional education does offer ways of memorizing exceptions, it does recognize the existence of patterns that do not comply with the letter-sound correspondences. Some strategies offered for remembering spellings of individual words range from using a rhythm (e.g., "say the names of the letters in a rhythm", e.g., (p-e) (o-p) (l-e)), to "say it like it looks" (e.g., *knock*: say the silent *k*), and mnemonics, e.g., *ne**c**e**ss**ary*: one **c**ollar two **s**ocks). However, it is also notable that phonics instruction is generally limited to the early years of education and it does not allow for patterns that are more complex and acquired later in literacy to be explored, such as the context-based graphotactic patterns presented in Section 1.3 (e.g., doublets occur more often after a single– than double-letter spelling as in *Jeff* vs. *deaf*; *bedding* vs. *heading*).

The debate regarding what kind of instruction is best, has been controversial in English-speaking educational systems (see Treiman, 2018). As discussed in Section 1.2.2, learning to spell conventionally engages processes that differ from reading. To this end, some studies have proposed that, instead of focusing on teaching specific conventions, educators should aim to understand and explain how the writing system works. Such information, in turn, would provide developing spellers with tools for unlocking the written language (Gaskins, Ehri, Cress, O'Hara, & Donnelly, 1997; Graham & Santangelo, 2014; Invernizzi, Abouzeid, & Bloodgood, 1997) (I return to this discussion in a later Chapter 5, Section 5.1.2).

Although explicit instruction is not the whole story on how children learn to spell, it is important to investigate whether teaching the underlying rules or regularities is helpful over and above implicit learning from exposure to print. Evidence to date is mixed: While Bhattacharya and Ehri (2004) showed that adolescents (11–14-year-olds) with dyslexia benefit from explicit instruction, Kemper, Verhoeven, and Bosman (2012) showed that implicit and explicit instruction of orthographic patterns was equally effective for both typically developing Dutch children (7-year-old) and those with spelling disabilities (9-year-old). It is, therefore, unclear whether the effectiveness of these types of processes depends on aspects of the learner.

Further research which has directly addressed the benefits of explicit learning relative to implicit learning (Bosman, van Hell, Verhoeven, Hell, & Verhoeven, 2006; Butyniec-Thomas & Woloshyn, 1997; Kemp & Bryant, 2003; Nunes, Bryant, & Olsson, 2003; Rastle, Lally, Davis, & Taylor, 2021) is discussed in the in-depth review of literature of the implicit relative to explicit learning in Chapter 5, Section 5.1.2. One of the aims of the current thesis is to find evidence for the differential effectiveness of these two types of instructions.

1.9 Objective of the Thesis

This chapter reviewed different lines of research relevant to the main goal of the current thesis: To explore the extent to which implicit statistical learning processes, which have been shown to operate in spoken language, may be relevant to learning to spell. A brief account of the *evolution* of writing demonstrated how different writing systems evolved to provide a solution between the trade-off of representing different aspects of spoken language, and how they may also have been shaped by chance events and human cognitive abilities, resulting in complexity and variation of written languages that is distinct from spoken language. As a result, some of the written language patterns—graphotactics, represented by probabilistic “rules” regarding the legal position and combination of letters—do not have a spoken language counterpart, and have been largely overlooked. These patterns are the focus of this thesis. The evidence presented in this review strongly suggests that these rules are mostly untaught, thus must be picked up implicitly via exposure to print. Statistical learning research provides an account of the ability to learn patterned regularities from the input and this inspired research that aimed to directly test sensitivity to graphotactics (e.g., Samara & Caravolas, 2014).

The current thesis follows on from this work. An initial objective was to see if learning effects could be increased compared with those in Samara and colleagues’ (Samara & Caravolas, 2014; Samara et al., 2019) work, since these were small, and this may have limited the ability to see smaller differences between conditions and relationships with literacy measures. Although we were not wholly successful in meeting this objective, since learning effects—though larger—remained small, we did validate an additional test, which was retained throughout the thesis. The second, and key goal of the thesis, was to establish that children can implicitly learn conditioned spelling constraints which have no phonological counterpart, that is, removing confounds with phonotactics, using methods adapted from Samara and colleagues. To date, only two studies have controlled for phonological confounds, but the experiments either featured adults (Chetail, 2017) or very simple (positional) constraints (Nigro et al., 2016). Two other goals of this thesis were: (i) to investigate the relationship between implicit learning measured in the current experiments and measures of literacy; and (ii) to compare this type of implicit learning with learning under explicit conditions. The elusive link between implicit learning and literacy attainment was discussed in light of methodological issues. The current thesis aimed to address these by using Bayes factor analyses, which are able to provide evidence for, as well as against, hypotheses. Finally, while spoken language was shown to be learned implicitly, spelling is usually taught explicitly in schools. Many of the complexities that need to be resolved when learning to spell cannot be explained through sound-letter correspondences or explicit "rules". Hence, at least some spelling rules are learned implicitly, as also established in the previous sections. Nevertheless, it is important to investigate how explicit instruction contributes to learning over and above implicit learning.

1.10 Overview of the thesis and experiments

Throughout this thesis, the primary method of statistical inference was the Bayes factor. Because this type of inference is less known and used in the psychological literature, and since the method for computing Bayes factors in this thesis is also relatively novel, the next chapter (**Chapter 2**) introduces this statistical approach in detail, along with the assumptions used in each empirical chapter.

Chapters 3–6 report a series of empirical experiments which build on Samara and colleagues' work (Samara & Caravolas, 2014; Samara et al., 2019) and use an incidental graphotactic learning task that was adapted to test children's and adults' learning of positional and context-based graphotactic patterns.

The aim of **Chapter 3** was to address issues raised in regards to task and stimuli design in statistical learning studies with the goal of increasing the magnitude of learning effects seen previously. This was achieved by using

Samara and Caravolas' (2014) stimuli with an adapted experimental procedure. The most substantial modification was to (i) increase the salience of correlated phonotactic constraints by overtly providing the matched phonology during graphotactic learning; other adaptations were: (ii) the use of a new exposure task to ensure that attention was focused on the “form” of the stimuli, (ii) the exposure time was increased to two sessions (as in Samara et al., 2019) from the one (as in Samara & Caravolas, 2014); and (iv) a novel test task (fill-in-the-blanks) was devised alongside the existing legality judgment, to simulate more closely what learners do in naturalistic situations when producing spellings. For both children and adults, standardized tests of English word reading and spelling ability were also administered. The hope was that effect sizes and reliability would be increased and that there might therefore be correlations with measures of literacy, which had not been previously seen. However, while a boost in learning was observed, this was rather small and the Bayes factor analyses demonstrated that evidence for the correlations between graphotactic learning and literacy was generally ambiguous, despite the paradigm adaptations.

Chapter 4 moved to explore whether children and adults learn graphotactic constraints when their ability to use phonotactics is removed. This was achieved by embedding positional constraints in semiartificial “word” stimuli, where the pronunciation was not provided and in which the middle character was kept as a pronounceable letter while the edge characters were unfamiliar symbols. As learning constraints using symbols was expected to be relatively more difficult than when using familiar letter strings, only learning of positional constraints was tested (given that these are easier for children to learn (Samara & Caravolas, 2014). Here, for children only, standardized tests of reading and spelling were included. This study showed learning of the visual positional constraints by both 7-year-olds and adults. Once again, no relationship was found between children’s learning of positional patterns and literacy measures.

In **Chapter 5**, learning of *context-based*, rather than positional, purely visual graphotactic constraints was tested with both children and adults. Since this type of constraints is harder to learn, instead of using unfamiliar symbols to create an artificial language, we returned to using fully English orthography and created homophone stimuli using single versus double letters, thus, ensuring that learning when letters double was a purely graphotactic effect. The pattern complexity was also manipulated to produce one relatively simple “rule” and a second, more complex one. Both children and adults learned the purely visual homophonic context-based constraints, both simple and complex. For children, a second question explored within this experiment was whether children’s context-based graphotactic learning is different under implicit and explicit conditions. To that end, for one of the constraints (simple), the implicit learning condition was compared against an explicit learning condition where the relevant spelling rule was explained to the children before the exposure phase. Again, for children, standardized tests of reading and spelling were included. This investigation revealed that both participant groups learned both the more simple and complex constraints under implicit conditions, however, children’s learning was stronger under explicit conditions and interestingly, performance in the experiment related to measures of literacy only in the explicit learning condition.

In **Chapter 6**, graphotactic learning of context-based patterns was further probed under different types of implicit and explicit conditions. In Chapter 5, it was demonstrated that participants could learn the purely visual context-based graphotactic constraints under implicit condition. However, even in the implicit condition, the task was such that the participants’ focus was mainly on *form*. In contrast, in real life literacy, children’s incidental learning of spelling will most often occur when they learn novel written words in a meaningful context. Therefore, in Chapter 5, the statistical learning methodology was adapted and incorporated into a task where participants learned word meanings. This is a new paradigm and, since learning patterns in a more indirect way, with less attention to form, is expected to be more difficult, as a first step, this study was carried out only with adults. Testing adults also meant that learning of the more complex patterns could be tested. Half of participants were tested under implicit conditions (where they were not told about the spelling rule) and half were tested under explicit conditions (where they were told about the rule before the exposure phase). While participants learned the complex, context-based graphotactic patterns under both implicit and explicit conditions, their performance was significantly better when they were explicitly informed about the rules governing them. Interestingly, however, their performance on word learning was better in the implicit condition, that is, when explicit instruction about spelling did not precede the word learning exposure phase. Implications and limitations of all these results are discussed **Chapter 7**.

2 Data Analyses

The primary method of inference in this thesis are prior-informed Bayes factors (BFs) (see also Samara et al., 2019) which, unlike frequentist *p* values, provide evidence both for and against the null. This method allowed us to compare the strength of evidence for H_1 (e.g., participants use the patterns with above chance accuracy in their legality judgment/fill-in-the-blanks task performance) over H_0 (e.g., participants' performance in a given test is at chance), given the data. In frequentist analyses, a nonsignificant *p* value cannot be interpreted as evidence for the null but it may be evidence for the null or no evidence for any conclusion (or, in some cases, even evidence against the null). Yet, nonsignificant *p* values are routinely taken as evidence that weakens the confidence in H_1. A crucial advantage of BFs is that they can provide quantifying evidence of H_0 (Dienes, 2008, 2015). In addition, BFs help support studies that use robust methods to differentiate between ambiguous results in support of the null that are valuable to research. This, in turn, may help address and remediate the file-drawer problem and publication bias.

In calculating BFs, we used the prior-informed approach outlined in Dienes (2008, 2015) (implemented in R by Baguley & Kaye, 2009). For each analysis, computing the evidence for H_1 over H_0 requires (a) a summary of the data being analyzed and (b) a model of H_1. I describe below the way each of these requirements have been generated in the current thesis.

For the summary of the data, we model the data using the sample mean difference (e.g., the difference between mean and chance, or mean difference between conditions) and the standard error (*SE*) of that mean. Because the Bayes factor calculator requires that the data are normally distributed, we fit logistic mixed effect models and use the betas and *SE*'s for the relevant coefficients: For the difference from chance, the beta for the intercept gives the difference between the grand mean in log-odds space (i.e., averaged across legal and illegal items given that a centered coding is used for all fixed effects) and 0 that is, 50% chance in log-odds space; for comparisons between age groups or experiments, beta for the fixed effect for age-group/experiment gives the difference between the relevant groups in log odds space (averaged across legal and illegal items). Further details of the models are given in Section 2.2 below.

The values generated in the log-odds space allow us to meet normality assumptions.

For the model of H_1 (or the "theory"), we need a predicted probability distribution of the relevant difference—be it probability distribution of the difference between our manipulation and chance, or between two conditions. For the model of H_0, that is, "obtaining evidence that something does not exist" (Dienes, 2020, under review), we need to specify what would be the effect, were it to exist. When the effects are predicted in both directions (e.g., for correlations), we determine if the effect is smaller than the specified values by defining a null interval (Lakens, Scheel, & Isager, 2018). Throughout this thesis we follow recommendation from Dienes (2014) and, since our predictions are in a particular direction, we model H_1 using a half-normal distribution with a mean of 0 and an *SD* of *x*, where *x* is an estimate of the predicted difference. When testing the null in equivalence testing, we use the estimate against values at least as extreme as the lower equivalence bound, that is, *-x*. Note that this distribution favors smaller values close to 0, which is appropriate for most experimental work (see Dienes, 2014 for details on other ways of modelling this distribution). It is important, but challenging, to specifying the value of *x* objectively, so I return to modelling of H_1 in Section 2.2 below, where I detail the method and the priors used for each test in this thesis.

With the summary of the data and a model of H_1, the BF is calculated, providing us with the ratio between the likelihood of the data given the model of the H_1, and the likelihood of the data given the null (i.e., H_1 over H_0). We interpret values larger than 3 as substantial evidence for H_1, values less than .33 as evidence for H_0, and values between these .33 and 3 as inconclusive or ambiguous evidence (Jeffreys, 1961). The notation $BF_{H(0,x)}$ is used (following advice by Dienes, see https://osf.io/hzcv6/) to indicate that the Bayes Factor was computed as a ratio of likelihood of the data, where *x* is the standard deviation (*SD*) of the half normal used to model H_1, versus the null.

Throughout this thesis, I report Bayes factor analyses alongside frequentist statistics (logistic mixed effect models). *p* values are included due to their familiarity and following Dienes (2008) recommendation of reporting "a *p* for every *B*". However, only the BFs are interpreted throughout.

2.1 Logistic mixed effect models used to obtain data summary

As noted above, data summaries for BF calculations were obtained by running logistic mixed effect models. Two dependent variables were included in all but one experiment[7]: Accuracy in a fill-in-the-blanks task and accuracy in the legality judgment task, and these were both binary, that is, were correct (coded with 1) and incorrect (coded with 0) responses. Logistic mixed effect regression models used the lme4 package in R (Bates, Mächler, Bolker, and Walker, 2015; R, 2014) and were carried out separately for each of the two binary dependent variables.

Fixed effects were (i) legality (within subjects), in all analyses of legality discrimination test performance; and (ii) experiment (between subjects) for analyses comparing experiments (e.g., for analyses comparing implicit and explicit conditions), age-groups (for analyses comparing adults and children) and block (for analyses comparing performance over time; i.e., blocks and sessions). All fixed factors were centered to reduce collinearity between main effects and interactions, and so that the intercept reflected the grand mean, and fixed effects of experiment and age-group were averaged across legality where that factor was included. The legality factor is regarded as a control factor, that is, to control for whether participants were better at correctly accepting legal items, if biased to say “yes”, or better at correctly rejecting illegal items, if biased to say “no”. Therefore, this fixed effect is not reported in the thesis. Random effects were random intercepts for participants with a full random-slope structure (that is, by-participant slopes for all experimentally manipulated within-subject effects), as recommended by Barr, Levy, Scheepers, and Tily (2013). All reported models converged with BOBYQA (Bound Optimization BY Quadratic Approximation; Powell, 2009).

[7] In Experiment 2 (semiartificial positional), we were unable to implement the fill-in-the-blanks task due to methodological constraints. Here, only the legality judgment was included.

2.2 Modelling the H_1

Our approach was to model H_1 using a half-normal distribution with an *SD* of x, and a mean of 0. In this thesis, we determined *x* (i.e., expected effect sizes under H_1) in one of the following ways:

2.2.1 Method A

Where directly relevant independent data were available (e.g., from a methodologically similar study), we used it to infer rough estimates of the expected learning effect (the estimates for each key analyses are listed in the tables below; Section 2.4). For example, Samara et al. (2019) study with 7-year-olds was used to estimate children's predicted legality judgment performance across all experiments with children reported here, that is, we use the beta value for the intercept coefficient from the equivalent logistic regression run in that study ($\beta = 0.16$). This is because they used the same methods as in the current thesis and similar stimuli embedding complex patterns, and showed robust effects using BF analyses. For adults, the equivalent value from the relevant experiment in that age group in Samara and Caravolas (2014) study are used. In other cases, we use a value from another experiment within this thesis. In particular, the value from Experiment 1 was used as prior for Experiments 3, 4, and 5, for the fill-in-the-blanks task with children and adults and for the legality judgment task with adults as well. Similarly, if there was evidence of an effect in one between-subjects condition (e.g., between age groups in Samara & Caravolas, 2014), those values were used to model H_1 for the studies in the current thesis (e.g., age group comparison in Experiment 1). When our study, in turn, showed evidence of between-subjects effect, we used this for our following studies (e.g., age group comparison in Experiment 2).

2.2.2 Method B

In some occasions, independent data was used as an estimate of the *maximum* possible effect that might be expected. Here, we set *x* to be half of that maximum, given that *x* is the *SD* of the half-normal and a maximum is approximately 2 *SD*. An example of this approach were the *BF* analyses carried out over adults' performance in the fill-in-the-blanks task. Because the new, fill-in-the-blanks task was only validated in Experiment 1, it was the most appropriate estimate for the same task in the following experiments. However, because the graphotactic patterns in this first experiment were visual as well as auditory (with matching phonotactics), adults' performance in the following experiments, when the phonotactics were removed, was not expected to exceed the performance in Experiment 1. Therefore, we set *x* to be half of the maximum (i.e., 1.02/2 = 0.51).

2.2.3 Method C

Finally, when neither rough estimates of the expected, nor a maximum effect could be specified from independent data, we determined a plausible maximum using a value from within the data itself (see Dienes, 2014, 2015 for a similar approach). For a main effect with two levels, a_1 and a_2 (where a_1 and a_2 are the means for each condition or group), a plausible maximum difference between two conditions (main effect) corresponds to a situation in which one level performs at chance, and therefore the whole effect is carried by the other level. For example, when comparing children in Experiment 5 (explicit learning) and Experiment 3 (implicit learning), we predict that the difference in these groups would be, at most, equivalent to the scenario where children in the *implicit* condition would be guessing (i.e., at 50% accuracy, 0 log odds) while those in the *explicit* condition would have an average performance equal to twice the grand mean (in log odds space). Since we use centred coding in the mixed-effects models (see Section 2.1 above), twice the grand mean corresponds to twice the intercept estimate. Since this is our estimate of the *maximum* effect, we set *x* to be equal to half this value, that is, to the coefficient for the intercept (See Appendix A for the calculation of a main effect with two levels).

2.3 Modelling H_1 for correlations

In addition to the analyses looking at performance in the two tests (fill-in-the-blanks and legality judgment), we also include some correlational analyses relating the performance in the experiment to performance to measures of literacy. For these analyses, in contrast to previous work, we again compute Bayes factors and compare the null against a half normal distribution. Here, to get our model of the data, we compute a Pearson's *r* correlation and *z* transform *r* to make it normal, and for the *SE* we calculate 1/squareroot(df - 1). This data will provide the coefficient needed to model the theory. To estimate value *x*, in some cases we use Method A (Section 2.2.1 above), and use the directly relevant independent data from within this thesis (e.g., since we find correlations in Experiment 5 with children, we use this to inform the theory in Experiments 3 and

4). When independent data cannot be specified, we use a new method (Mathod D), described below.

2.3.1 Method D

We used an estimate of the expected coefficient under the H_1 based on a small-to-medium effect (*r* = .40) as this seemed most consistent with previous literature. That is, we predicted that, if H_1 was true, we would find a correlation approximately as strong as that found in previous studies reporting significant associations (Fisher's *z*-transformed coefficient =.42).

2.4 Tables with specific priors and method used to model H_1

In the tables below, for each study and each Bayes factor, I indicate which method was used to compute it and what studies are used to inform theory. I refer back to these tables in each results section throughout this thesis.

Table 2.1: Modelling the H_1 for Experiment 1, for Fill-in-the-blanks and Legality Judgment

EFFECT OF INTEREST	DIRECTION OF PREDICTION	SUMMARY DATA FROM:	TASK: fill-in-the-blanks: value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	fill-in-the-blanks: source of *x*	fill-in-the-blanks: H1 estimation type	legality judgment task: value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	legality judgment task: source of *x*	legality judgment task: H1 estimation Method
comaprison bewteen children's overall performance in Experiment 1 (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant child data	0.16	estimate of intercept from equivalent model of children's performance in legatlity judgment task* in Samara, Singh & Wonnacott (2019)	A	0.16	estimate of intercept from equivalent model of children's performance in legatlity judgment task in Samara, Singh & Wonnacott (2019), children	A
comaprison bewteen adult's overall performance in Experiment 1 (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant adult data	0.20	estimate of intercept from equivalent model of children's performance in adults legality judgment task* in Samara, Singh & Wonnacott (2019), children	A	0.20	estimate of intercept from equivalent model of adult's performance in legatlity judgment task in Samara, Singh & Wonnacott (2019), children	A
comparison bewteen adults and children in Experiment 1	adults outperform children	beta and *SE* of fixed effect of age-group in glmer model over combined data from adults and children	0.11	estimate of fixed effect of age group from equivalent model of adults and children's performance in legatlity judgment* task in Samara & Caravolas (2014)	A	0.11	estimate of fixed effect of age group from equivalent model of adults' and children's performance in legatlity judgment task in Samara & Caravolas	A
comparison between children's performance in Experiment 1 and in relevant study in Samara & Caravolas (2014)	higher performance in Experiment 1 than in Samara & Caravolas (2014)	beta and *SE* of fixed effect of experiment in glmer model over combined data from children in Experiment 1 and children in Samara & Caravolas (2014)	NA	NA	NA	0.11	estimate of grand mean (intercept) from same model as used to obtain data summary	C
comparison between adults' performance in Experiment 1 and in relevant study in Samara & Caravolas (2014)	higher performance in Experiment 1 than in Samara & Caravolas (2014)	beta and *SE* of fixed effect of experiment in glmer model over combined data from adults in Experiment 1 and adults in Samara & Caravolas (2014)	NA	NA	NA	0.29	estimate of grand mean (intercept) from same model as used to obtain data summary	C
correlations between children's performance in Experiment 1 tasks and performance in TOWRE and WRAT	postivie correlations	Fisher transformed *r* value for each correlation for children ; *SE* = 0.24	0.42	Fisher's *z* -transformed value of *r* = 40 - which is based on estimate a small-to-medium effect effect	D	0.42	Fisher's *z* -transformed value of *r* = 40 - which is based on estimate a small-to-medium effect effect	D
correlations between adults' performance in Experiment 1 tasks and performance in TOWRE and WRAT	postivie correlations	Fisher transformed *r* value for each correlation for adults; *SE* = 0.26	0.42	Fisher's *z* -transformed value of *r* = 40 - which is based on estimate a small-to-medium effect effect	D	0.42	Fisher's *z* -transformed value of *r* = 40 - which is based on estimate a small-to-medium effect effect	D

Table 2.2: Modelling the H_1 for Experiment 1, for Correlations Between Tests

EFFECT OF INTEREST	DIRECTION OF PREDICTION	SUMMARY DATA FROM:	value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	source of *x*	H1 estimation Method
correlations between children' performance in fill-in-the-blanks and legality judgment task in Experiment 1	positive correlations	Fisher transformed r value each correlation for adults; SE = 0.24	0.58	Fisher's *z* -transformed value of *r* = 52 - which is based on estimate from equivalent model for correlation of adults' performance in the two tasks in Experiment 6	A
correlations between adults' performance in fill-in-the-blanks and legality judgment task in Experiment 1	positive correlations	Fisher transformed r value each correlation for adults; SE = 0.17	0.58	Fisher's *z* -transformed value of *r* = 52 - which is based on estimate from equivalent model for correlation of adults' performance in the two tasks in Experiment 6	A

Table 2.3: Modelling the H_1 for Experiment 2, for Legality Judgment

EFFECT OF INTEREST	DIRECTION OF PREDICTION	SUMMARY DATA FROM:	TASK		
			legality judgment task		
			value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	source of *x*	H1 estimation Method
comaprison bewteen children's overall performance in Experiment 2 with trained and untrained consonants (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant child data	0.16	estimate of intercept from equivalent model of children's performance in legatlity judgment task in Samara, Singh & Wonnacott (2019), children	A
comaprison bewteen children's performance with trained and untrained consonants in Experiment 2	higher performance with trained than withy untrained consonants	beta and SE of fixed effect of consonant type in glmer model over relevant combined data from trained and untrained consonant	0.49	estimate of grand mean (intercept) from same model as used to obtain data summary	C
comaprison bewteen adult's overall performance in Experiment 2 with trained and untrained consonants (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant adult data	0.20	estimate of intercept from equivalent model of adult's performance in legatlity judgment task in Samara & Caravolas (2014), adults	A
comaprison bewteen adults's performance with trained and untrained consonants in Experiment 2	higher performance with trained than with untrained consonants	beta and SE of fixed effect of consonant type in glmer model over relevant combined data from trained and untrained consonant	0.92	estimate of grand mean (intercept) from same model as used to obtain data summary	C
comparison bewteen adults and children in Experiment 2, separate analyses for trained and untrained consonants	adults outperform children with both trained and untrained consonants	beta and *SE* of fixed effect of age-group in glmer model over combined data from adults and children, from separate analyses with trained and untrained consonants	0.42	estimate of fixed effect of age group from equivalent model of adults and children's performance in legatlity judgment task in Experiment 1	A
correlations between children's perforamnce in Experiment 2 tasks and performance in TOWRE and WRAT	postivie correlations	Fisher transformed *r* value for each correlation for children ; *SE* = 0.18	0.42	Fisher's *z* -transformed value of *r* = 40 - which is based on estimate a small-to-medium effect effect	D

Table 2.4: Modelling the H_1 for Experiment 3, 4, and 5, for Fill-in-the-blanks and Legality Judgment

			TASK					
			fill-in-the-blanks			legality judgment task		
EFFECT OF INTEREST	DIRECTION OF PREDICTION	SUMMARY DATA FROM:	value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	source of *x*	H1 estimation type	value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	source of *x*	H1 estimation Method
comaprison bewteen children's overall performance in Experiment 3, Experiment 4 and Experiment 5 (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant child data	0.39	estimate of intercept from equivalent model of children's performance in fill-in-the-blanks task in Experiment 1	A	0.16	estimate of intercept from equivalent model of children's performance in legatlity judgment task in Samara, Singh & Wonnacott (2019), children	A
comaprison bewteen adult's overall performance in Experiment 3 and Experiment 4 (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant adult data	0.51	half of the estimate (0.1.02/2) of intercept from equivalent model of adults' performance in fill-in-the-blanks task in Experiment 1	B	0.32	half of the estimate (0.65/2) of intercept from equivalent model of adults' performance in legality judgment task in Experiment 1	B
comparison bewteen children's performance in Experiment 3 and Experiment 4	better performance in Experiment 3	beta and SE of fixed effect of experiment in glmer model over combined data from Experiment 3 and Experiment 4	0.23	estimate of grand mean (intercept) from same model as used to obtain data summary	C	0.17	estimate of grand mean (intercept) from same model as used to obtain data summary	C
comparison bewteen adults' performance in Experiment 3 and Experiment 4	better performance in Experiment 3	beta and SE of fixed effect of experiment in glmer model over combined data from Experiment 3 and Experiment 4	0.35	estimate of grand mean (intercept) from same model as used to obtain data summary	C	0.21	estimate of grand mean (intercept) from same model as used to obtain data summary	C
comparison bewteen children's performance in Experiment 3 and Experiment 5	better performance in Experiment 5	beta and SE of fixed effect of experiment in glmer model over combined data from Experiment 3 and Experiment 4	0.83	estimate of grand mean (intercept) from same model as used to obtain data summary	C	0.38	estimate of grand mean (intercept) from same model as used to obtain data summary	C
correlations between children's perforamnce in Experiment 3 and Experiment 4 tasks and performance in TOWRE and WRAT	postivie correlations	Fisher transformed *r* value for each correlation for children ; *SE* = 0.18 (Experiment 3) and SE = 0.21 (Experiment 4)	0.52	Fisher's *z*-transformed value of *r* = .48 - which is based on estimate from Experiment 5	A	0.52	Fisher's *z*-transformed value of *r* = .48 - which is based on estimate from Experiment 5	A
correlations between children's perforamance in Experiment 5 tasks and performance in TOWRE and WRAT	postivie correlations	Fisher transformed *r* value for each correlation for children; *SE* = 0.21	0.42	Fisher's *z*-transformed value of *r* = 40 - which is based on estimate a small-to-medium effect effect	D	0.42	Fisher's *z*-transformed value of *r* = 40 - which is based on estimate a small-to-medium effect effect	D

Table 2.5: Modelling the H_1 for Experiment 6 and 7, for Fill-in-the-blanks and Legality Judgment

EFFECT OF INTEREST	DIRECTION OF PREDICTION	SUMMARY DATA FROM:	TASK: fill-in-the-blanks			TASK: legality judgment task		
			value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	source of *x*	H1 estimation type	value of *x* used as rough estimate of predited effect size (used to model H_1 as a half normal with mean of zero and *SD* of *x*)	source of *x*	H1 estimation Method
comaprison bewteen adults' overall performance in Experiment 6 and Experiment 7 (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from each glmer model over relevant adult data	0.20	estimate of intercept from equivalent model of children's performance in fill-in-the-blanks task in Experiment 4	A	0.16	estimate of intercept from equivalent model of children's performance in legality judgment task in Experiment 4	A
comparison between adults' performance in Experiment 4 and Experiment 6	higher performance in Experiment 4 than in Experiment 6	beta and *SE* of fixed effect of experiment in glmer model over combined data from adults in Experiment 4 and adults in Experiment 6	0.18	estimate of grand mean (intercept) from same model as used to obtain data summary	C	0.16	estimate of grand mean (intercept) from same model as used to obtain data summary	C
comparison between adults' performance in Experiment 6 and Experiment 7	higher performance in Experiment 7 than in Experiment 6	beta and *SE* of fixed effect of experiment in glmer model over combined data from adults in Experiment 6 and adults in Experiment 7	0.50	estimate of grand mean (intercept) from same model as used to obtain data summary	C	0.46	estimate of grand mean (intercept) from same model as used to obtain data summary	C

Table 2.6: Modelling the H_1 for Experiment 6 and 7, for the Word Learning Task

TASK					
word learning					
EFFECT OF INTEREST	DIRECTION OF PREDICTION	SUMMARY DATA FROM:	value of x used as rough estimate of predited effect size (used to model H1 as a half normal with mean of zero and SD of x)	source of x	H1 estimation Method
comaprison bewteen adults' overall performance in Experiment 6 (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant adult data	0.71	estimate of intercept from equivalent model of adults' performance in word learning task in Experiment 7	A
Comparison between blocks in Experiment 6	increasing performance from block 1 to block 6	beta and *SE* of fixed effect of block in glmer model over Experiment 6 data	0.20	estimate of fixed effect from equivalent model of adults' performance in word learning task in Experiment 7	A
comaprison bewteen adults' overall performance in Experiment 7 (their grandmean) and chance	above chance	beta and *SE* of intercept coefficient from glmer model over relevant adult data	1.46	estimate of intercept from equivalent model of adults' performance in word learning task in Experiment 6	A
Comparison between blocks in Experiment 7	increasing performance from block 1 to block 6	beta and *SE* of fixed effect of block in glmer model over Experiment 7 data	0.33	estimate of fixed effect from equivalent model of adults' performance in word learning task in Experiment 6	A
correlations between adults' performance in word learning and fill-in-the-blanks task in Experiment 6 and Experiment 7	postivie and negative correlations	Fisher transformed r value each correlation for adults; SE = 0.17	0.29	half of Fisher's z-transformed value of r = 52 - which is based on estimate from equivalent model for correlation of adults' performance in fill-in-the-blanks and legality judgment task, in Experiment 1	B
correlations between adults' performance in word learning and legality judgment task in Experiment 6	postivie and negative correlations	Fisher transformed r value each correlation for adults; SE = 0.16	0.29	half of Fisher's z-transformed value of r = 52 - which is based on estimate from equivalent model for correlation of adults' performance in fill-in-the-blanks and legality judgment task, in Experiment 1	B
correlations between adults' performance in fill-in-the-blanks and legality judgment task in Experiment 6 and Experiment 7	positive correlations	Fisher transformed r value each correlation for adults; SE = 0.17	0.58	Fisher's z-transformed value of r = 52 - which is based on estimate from equivalent model for xorrelation of adults' performance in the two tasks in Experiment 1	A
comparison between adults' performance in word learning in Experiment 6 and Experiment 7	higher performance in Experiment 7 than in Experiment 6	beta and *SE* of fixed effect of experiment in glmer model over combined data from adults in Experiment 6 and adults in Experiment 7	1.05	estimate of grand mean (intercept) from same model as used to obtain data summary	C

2.5 Robustness regions

The statistical approach used in this thesis is relatively novel and the choice of H_1 is subject to debate (Dienes, 2008). Therefore, the methods described above provide only a rough estimate of the likely effect size, and there is subjectivity in the choice of value. Given these considerations, we follow Dienes' (https://osf.io/hzcv6/) recommendation and estimate the likely size of the effect by calculating robustness regions (*RR*) for each BF, notated as: *RR* [x_1, x_2]. These show the range of estimates of H_1 (x_1 being the smallest and x_2 the largest *SD)* for which our data would support the same qualitative conclusion (i.e., if we report substantial evidence for H_1 with our chosen *x*, this is the range of values of *x* that would also have led to substantial evidence for H_1; if we report substantial evidence for H_0, this is the range of values of *x* that would also have led to substantial evidence for H_0; if we report ambiguous evidence, this is the range of values of *x* that would also have led to inconclusive evidence). The robustness regions should be interpreted bearing in mind that larger values of *x* bias the computation to find evidence for the null, whereas smaller values are bias in favor of H_1. They were calculated by testing values of *x* which are reasonable given our scale, taking increments of 0.001: Specifically, from 0 (corresponding to no difference from chance/no difference between the groups) to 4.595 log odds space (corresponding to almost perfect performance compared to 50% chance/one group has almost perfect performance at 99% and the other one has minimal performance and is at chance).

2.6 Required sample size estimation

Following the first study in this thesis (Experiment 1), we used the Bayesian approach to estimate the sample size that was likely to be necessary to see a substantial effect (i.e., *BF* = 3), similar to carrying our power analyses in a frequentist approach. Our H_1 assumption was that children perform above chance in our implicit learning tasks while H_0 would show at chance performance. We aimed to see at least the effect observed in Samara et al. (2019) (β=0.155) and at best as shown in our Experiment 1 (β=0.288). This is because the following studies in this thesis do not incorporate phonology and even if access to phonology was attempted covertly, it was not meant to be useful. Therefore, the Samara et al. (2019) study is expected to provide a closer alternative to the current data, as it did not incorporate phonology at all.

Bayes factor calculation showed a BF larger than three with Samara et al. (2019) sample of 60 participants (*BF*=5.3) so we used a Bayes calculation to find out if a BF larger than three can be obtained using fewer participants. The analysis suggested that, with 44 participants, a BF larger than three could be obtained. To round up and allow for increased power, we planned to test a maximum of 50 participants in each study and each condition. However, since this is no guarantee that we will find evidence with this number of participants, and since we also might find evidence with fewer participants, we planned to use optional stopping, given that Bayesian statistics are unaffected by the stopping rule, provided their priors are data informed (i.e., as in our work) (Dienes, 2016; Rouder, 2014). Specifically, we planned (and in some cases pre-registered, see Section 2.7) that we would look at the data after 25 participants and planned to stop testing if BF was larger than 3 (we have substantial evidence for H_1) or less than .33 (we have substantial evidence for H_0) (Jeffreys, 1961). If the data were insensitive, we increased the number by 10 participants until the null/alternative hypothesis were shown, or up to 50 participants. Note that we followed this procedure for comparisons against chance, our main investigation, and not when the results for comparisons between conditions or for correlations were ambiguous. Hence, when we stopped testing, some of our results remained ambiguous. In these cases, we ran further analyses to determine how many more participants would have been needed to find substantial evidence for or against the H_1 and discussed possible limitations and solutions.

2.7 Pre-registrations

As a solution to the replication crisis and publication bias, it was proposed that researchers report their study design and hypothesis prior to data collection (Nosek & Lakens, 2014). Any additional analyses or hypotheses tested outside the ones listed in the pre-registered report, therefore, are unambiguously exploratory. In this thesis, for the majority of the analyses in Chapter 5 (critically, all comparisons against chance), we prespecified our priors in a preregistered plan on OSF (Open Science Forum, https://osf.io/), a free public repository. I provide links to pre-registrations, as well as the collected data and R analyses scripts for each chapter (also listed in Appendix B) and indicate in the relevant results sections when the values that informed the alternative hypothesis, H_1, were not prespecified (e.g., exploratory correlations).

3 Incidental Phono-Graphotactic Learning of Novel Contextual Patterns

3.1 Introduction

The research reviewed in Section 1.3 and 1.6 strongly suggests that children and adults pick up statistical patterns from written language implicitly, via exposure to print (e.g., Pollo et al., 2009; Treiman & Wolter, 2018). Some patterns are purely visual, without a phonological counterpart (see Section 1.3.4 for examples). Graphotactics, the patterns of interest in the current thesis, are explained by frequency of graphemes and their probability of occurrence and co-occurrence in a certain context. Evidence of children's and adults' sensitivity to graphotactics comes from studies that looked at experience with natural languages (Kessler, 2009; Pollo et al., 2007; Steffler, 2001; Treiman & Kessler, 2013; see also Section 1.3), as well as studies that used well controlled experimental conditions (Chetail, 2017; Nigro et al., 2015, 2016; Samara & Caravolas, 2014; Samara et al., 2019; see Section 1.6). For example, the work of Treiman and colleagues (Cassar & Treiman, 1997; Treiman, 1993) unequivocally demonstrates that sensitivity to the frequency and legal position of doublets emerges in early childhood. The evidence comes from both children's naturalistic spellings (Treiman, 1993) and controlled experimental conditions, where a word-likeness task showed that children chose nonwords conforming to orthographic rules of double letters in English. Samara and Caravolas (2014) provided a first demonstration of graphotactic learning under incidental experimental conditions. Importantly, they demonstrated that, following brief exposure to novel patterns embedded in made-up pronounceable nonwords, English-speaking children and adults were able to generalize to novel stimuli that conformed to the newly learned graphotactic restrictions. This was the case when the patterns were both easier (i.e., embedded zero-order positional constraints; e.g., *l* can begin words and *p* can end words as in *les*) and more complex (i.e., embedded first-order contextual constraints; e.g., *t* can be followed by *o*, never *e* and *o* can be followed by *f*, never *t*, as in *tof*). However, the easier, positional patterns were learned more readily than the conditional, contextual ones. Because the constraints in Samara and Caravolas (2014) were embedded in both beginning (body) and end (rime) units in the stimuli, participants had redundant cues to rely on when judging the legality of each stimulus (see Section 3.2.3 for further discussion). Samara et al. (2019) addressed this limitation and extended the findings on contextual constraints by showing that English- and Turkish-speaking children could learn patterns which were conditioned *either* in word-beginnings *or* word-ending positions (rather than both). While overall learning has been consistently shown in both Samara and colleagues' studies above, the effects have been small—e.g., 2% (children) and 5% (adults) above chance (50%) level in Samara and Caravolas (2014), and 3% (children) in Samara et al. (2019)—and this may have been the reason for a lack of—or inconclusive—evidence for some exploratory questions. For example, while Bayes factor calculations in Samara and Caravolas (2014) study revealed no evidence that learning constraints embedded in body units was substantially different than in rime units (as noted in Section 1.6), this evidence was inconclusive. Furthermore, when correlations between pattern learning and literacy measures were explored, no associations were found. The current chapter builds on Samara and colleagues' work by making adaptations to their design, with the aim of increasing the learning effects seen previously, and exploring the relationship between learning measured in the lab and literacy. In the next three sections, I present studies that point to a need for such adaptation as well as how these changes are likely to boost learning.

3.1.1 The issue of reliability

Previous studies that looked for evidence of a relationship between statistical learning and language attainment either found weak to medium correlations (see Section 1.7.1) or no correlations at all (see Section 1.7.2). West et al. (2017) argued that null findings are explained by poor reliability of implicit learning tasks, compared to good reliability in explicit tasks (see also Krishnan & Watkins, 2019). Some of the reasons for such low reliability, as pointed by Siegelman et al. (2017), could be the methodological limitations of lab-based implicit learning experiments, such as short tasks with few trials and a restricted level of difficulty, as well as a limited selection of structures to be learned. Some of these limitations have been addressed recently by Kuppuraj, Duta, Thompson, and Bishop (2018) with adults, in an auditory-picture triplet learning task, using familiar words and varying types of dependencies (adjacent deterministic and probabilistic as well as nonadjacent probabilistic). They also devised an online task in addition to the offline recall task, to measure the online perceptual detection of regularities rather than the skill performance (e.g., the motor reactions typically required and measured in a serial reaction time task, SRT). They achieved a test-retest reliability of 0.67, a significant improvement to what was previously found in similar tasks (West et al., 2017). It is, therefore, possible to obtain higher reliability in lab-based studies, and that is the goal of the current study. This, of course, does not mean that we will find evidence of the relationships in question: Of note in Kuppuraj's study (Kuppuraj et al., 2018) is that, even with acceptable levels of reliability, they did not find a correlation between learning and a measure of verbal short-term memory. However, given our use of Bayes factors, it may allow us to find conclusive evidence in one direction—that is, either for H_1 or for the null.

3.1.2 Design adaptations

The goal of the current chapter is to address issues raised in regards to task and stimuli design, and improve reliability of our tasks, within the context of working with a developmental population, for whom longer and complex tasks prove challenging. With this in mind, we test a new version of Samara and Caravolas (2014) experiment where we modified the following aspects of the design:

3.1.2.1 Increased the exposure time

In Samara and Caravolas (2014), participants were tested in a single session; however, more exposure generally leads to higher learning increase, and spreading over multiple days allows for a period of sleep that could also be helpful (Brown, Weighall, Henderson, & Gaskell, 2012; Henderson, Weighall, Brown, & Gaskell, 2012). We therefore increased the exposure time to as much as in Samara et al. (2019), that is, two sessions on two consecutive days. Given the time constraints of working with children in schools, two days seemed reasonable.

3.1.2.2 Modified the exposure task

In Samara and colleagues' (Samara & Caravolas, 2014; Samara et al., 2019) studies, a color-detection task was used at exposure, whereby participants saw the words on the screen and one of the three letters (Samara & Caravolas, 2014) or all three letters (Samara et al., 2019) changed from black to either green, blue or red and participants were required to respond by pressing one on the colored buttons on the keypad. The goal of such cover task was to ensure that participants attend to the words and do not look away, without encouraging them to search for patterns. However, it is possible that this task encouraged them to attend to one letter in isolation, in the case of letter color detection, or just the color, a redundant cue, distracting from the form. Therefore, we modified the exposure task to encourage participants to attend to the form of the nonwords as a whole rather than their irrelevant properties. This was achieved by replacing the color-detection task with a one-back-task, where participants were asked to respond with one button if the word they saw was the same as the one before. Such a task has been used in other visual statistical learning studies (e.g., Arciuli & Simpson, 2011) to ensure that attention is paid during familiarization (or exposure) phase.

3.1.2.3 Added a fill-in-the-blanks test alongside the existing legality judgment

To test learning, we retained the legality judgment task used in both Samara and Caravolas (2014) and Samara et al. (2019) and added a novel task: fill-in-the-blanks. The purpose of introducing a second task was to show how robust the learning effects are. We chose as our new task a production task more directly related to the tasks of producing correct spellings in text. As noted in Section 1.2.2, although spelling is not just the reverse of reading, it does require spellers to encode language, a skill arguably harder than word recognition (Bosman & Van Orden, 1997). Following exposure, during which regularities are encoded and learned, a legality judgment task can measure the knowledge acquired. While this test has been used extensively to test sensitivity to novel patterns, the task involved in spelling, a production task, requires additional skills to those required in recognizing familiar items. Newly acquired knowledge about patterns needs to be retrieved from memory and reassembled in novel structures (particularly when generalization is tested).

3.1.2.4 Added correlated phonotactic cues to the graphotactics

The most substantial adaptation to Samara and Caravolas' (2014) study was the overt addition of phonology, thus adding phonotactic cues to the graphotactics. As pointed out in Section 1.6, while the experiments in Samara and colleagues (Samara & Caravolas, 2014; Samara et al., 2019) were designed to test learning of patterns in written language, in all cases the stimuli formed pronounceable nonwords of CVC letter strings that conform to phonotactic rules. Therefore, it cannot be ruled out that learners attempted to covertly pronounce the stimuli and thus learning benefitted from the fact that the graphotactic patterns all had a phonotactic counterpart, e.g.., the graphotactic rule "*t* can only follow *o*, never *e*" had a sound based phonotactic counterpart "/t/ can only follow /o/, never /ɛ/". This possibility is especially supported by evidence from research that clearly demonstrates that sensitivity to probabilistic phonotactics emerges in infants as young as 7-months-old (e.g., Aslin et al., 1998; Mattys et al., 1999; McCarthy, 1979; Onishi et al., 2002; Pitt, 1998; see Sections 1.5.1 and 1.5.4). Onishi and colleagues demonstrated that infants and adults learn novel phonotactic patterns from brief listening experience (Chambers et al., 2010; Onishi et al., 2002; Seidl et al., 2009) using a rather similar methodology to the studies by Samara and colleagues' work and thus it is possible that the same type of phonotactic learning underpinned apparent graphotactic learning in these studies. Of note, however, is that fact that in Samara and colleagues' studies, the verbalization was not encouraged and thus, it is possible that, despite the presence of phonotactic cues, participants might not have accessed or used them. In the current study, we used the same written stimuli as Samara and Caravolas (2014), but we adapted the exposure design such that the phonology was provided overtly—that is, participants hear the words as well as see them.

Some studies from the statistical learning literature have directly assessed the relationship between information perceived through the auditory and visual modalities (cross-modality facilitation). As discussed in Section 1.5, statistical learning has been shown to be a domain general ability that operates across domains beyond language (e.g., Saffran et al., 1996), such as vision (e.g., Fiser & Aslin, 2001, 2002; Kirkham et al., 2002), although there are constraints relevant to the cognitive functions specific to each modality (Frost et al., 2015). Humans perceive and process real-life information through multiple modalities at once and integrate this information into unified events (Stein & Meredith, 1993). During reading and spelling, information is processed through the visual modality (see Section 1.2 for a discussion of the cognitive system engaged in processing written language), as well as via phonological representations that engage the auditory modality. When the stimuli generate simultaneous (or near simultaneous) information in both modalities, as is the case in the stimuli used by Samara and Caravolas (2014), the effect is multimodal. When stimuli in one sensory modality influence the perception of another modality, the effect is crossmodal (Spence, Senkowski, & Roder, 2009). These effects can either facilitate or hamper processing, depending on the match or mismatch between the information perceived. To that end, Robinson and Sloutsky (2007) presented adults with colored abstract shapes alongside the auditory sequences used in Saffran et al. (1996), in a continuous stream. The shapes were arranged into triplets with a structure which matched with the auditory syllable sequences (e.g., for the word *pakibu* there is a shape representing “pa” one for “ki” and one for “bu”, etc.), thus sharing the same transitional probabilities (see Section 1.5.1 for more details regarding embedded statistics). The results showed cross-modality facilitation, that is, adults were better at correctly identifying the trained visual sequences when the statistics in the auditory stimuli were correlated with those in the visual stimuli, compared to when the auditory information was absent. Interestingly, this facilitation was not uniform: The auditory sequence learning did not benefit from correlated visual statistics.

Other studies have shown that, when there is a mismatch between auditory and visual cues, this can hinder learning in each modality (Mitchel & Weiss, 2011), although participants are capable of learning patterns in different perceptual modalities which are fully orthogonal (Conway & Christiansen, 2006).

The evidence from the study by Robinson and Sloutsky (2007) suggests that statistical learning may benefit from the availability of correlated multisensory input. Given the multimodal nature of the learning environment (Shams & Seitz, 2008) and perceptual systems (Stein & Stanford, 2008), such conclusion seems reasonable. It is expected, therefore, that the presence of overt phonology (auditory modality) will facilitate learning of graphotactic patterns (visual modality), if the underlying statistics are matched. In the current study, therefore, we expect to see more robust learning effects than in the study using the same stimuli but where the phonology was not overtly provided and might not have been fully accessed by all participants (Samara & Caravolas, 2014).

3.1.3 The current study

A key goal of our study is to assess whether, by adapting Samara and Caravolas' (2014) design using identical stimuli, we can observe an increased effect of learning the contextual graphotactic constraints. We call this study *phono-graphotactic*. We used the same written stimuli as in Samara and

Caravolas (2014), that is, context-based patterns at both beginning and end of made-up words; however written stimuli were accompanied by aural recordings of the words during exposure. Following Samara & Caravolas (2014), we included skilled adult readers with ample print experience and relatively beginner readers (6–7-year-olds) in line with previous work measuring pattern sensitivity in natural orthography (e.g., Cassar & Treiman's (1997) youngest group; Hayes et al., (2006) participants). Both were exposed, under identical task procedures, to pattern embedded visual and aural stimuli in two brief sessions (compared with one in Samara & Caravolas, 2014). To encourage participants to attend to the details of the written stimuli, in place of the color-detection task used in Samara and Caravolas (2014), we used the one-back task discussed above (Section 3.1.2.2). Following exposure, learning generalizations were tested in two tasks: First, the new *fill-in-the-blanks test* discussed above (Section 3.1.2.3), in which participants were asked to construct novel nonwords (i.e., not in the training set) by choosing between one of two possible vowels to “fill-in” a consonantal frame (C_C), where the target vowel conformed to the learned pattern and the foil violated the known pattern; and a second, legality judgment task (as used in Samara & Caravolas, 2014; Samara et al., 2019), in which participants were required to provide yes/no answers to unseen stimuli that either conformed to or violated the learned patterns. We sought to validate the novel task and expected to see above chance performance in both. We also tested the correlation between the two tasks with the expectation that, if they are, indeed, measuring outcomes of the same learning mechanism, they will be positively correlated. We also compared the results of the legality judgment task against those in Samara and Caravolas (2014) to see whether the various changes to the paradigm have indeed improved learning.

We sought to examine the relationship between incidental graphotactic learning and general reading or spelling ability and, in addition to the incidental graphotactic learning task, our experimental design included standardized measures of English word reading and spelling ability to all child and adult participants. We looked for correlations between these and the experimental task and employed a Bayes factor method of analysis to test whether nonsignificant associations between statistical learning and literacy skill are substantial evidence against theoretical accounts that predict a relationship or an artefact of data insensitivity.

As in all studies in this thesis, we also administered a post-experiment questionnaire to obtain subjective reports of participants' awareness of the experimenter-included patterns, as well as to tap into their intuition as to what was driving their performance in both post-tests. While we acknowledge the limitation of this method (Batterink, Reber, Neville, & Paller, 2015), we use the questionnaire, as previous work (e.g., Treiman & Boland, 2017), to shed light on previous inconsistent findings regarding children's and adults' ability to report on untaught spelling patterns. In this experiment, the questionnaire was verbal for both children and adults.

3.2 Experiment 1

3.2.1 Method

3.2.1.1 Participants

We were not able to follow the procedure outlined in Section 2.6 on estimating the required sample size and therefore, did not use optional stopping. This was because this experiment was designed and carried out before we established this procedure. Guided by a subset of data that was available at the time for an experiment carried out in our lab: Samara et al. (2019), we aimed to test 20 participants in each group.

Twenty typically developing Year 2 children (12 female, 8 male; mean age = 7.1 years, *SD* = 0.31) took part in the study. They were all recruited using an opt-in procedure from a primary school in London, had no known language, hearing or vision impairments and no history of learning difficulties. 3 children reported English as their second language while 17 were monolingual English speakers and all 20 had received the same amount of formal literacy tuition (2 years). Children were rewarded with stickers and a certificate. As in previous work (Samara et al., 2019), the mean reading and spelling performance in our sample was above average (mean reading = 121.25, *SD* = 8.13; mean spelling = 129.9, *SD* = 7.16), which is relatively typical in experimental studies with child participants.[8]

Eighteen adults (5 female, 13 male; mean age = 30.6 years, *SD* = 9.96) also participated in the study. The mean reading and spelling performance in our sample was average (mean reading = 89.45, *SD* = 5.42; mean spelling = 111, *SD* = 13.5). They were recruited via UCL's Sona system and were tested in the university's lab. They reported being monolingual native speakers of English with no language, hearing, vision impairments or any learning difficulties. They all provided informed consent and were paid for their participation.

3.2.1.2 Materials

Phono-Graphotactic learning task. Following the paradigm of Samara and Caravolas (2014), and using the same stimuli, the first-order context-based constraints were maintained, so that consonant position depended on the vowel type. Two sets of consonants (Set 1: *d*, *m*, *l*, and *f*; Set 2: *t*, *n*, *p*, and *s*) were combined into thirty-two C_C syllable frames. For half of the syllable frames, Set 1 consonants were used as onsets and Set 2 as codas (C_1_C_2, e.g., *d_t*); and for the other half, the reverse was true (C_2_C_1, e.g., *t_d*). Two vowels (*o* and *e*) were used to fill in the syllable frames, giving rise to a total of sixty-four CVC pronounceable non-words (see Figure 3.1). These were arranged in four lists, two of which conformed to- and two of which violated a novel phono-graphotactic constraints. Nonwords from the pattern-conforming lists served as exposure and legal unseen test items, and nonwords from one of the two pattern violating lists served as illegal test items. List assignment was counterbalanced between participants such that, for half of the participants C_1_C_2 frames were combined with the vowel *o* (e.g., *fos*) and C_2_C_1 frames were combined with *e* (e.g., *tem*), while for the other half of the participants, the reverse was true (e.g., *fes* and *tom*). The full set of items used in this experiment is shown in Appendix C.

[8] WRAT-IV standardization is drawn following normative data collected from the US, where formal literacy instruction begins one year later relative to the UK. Thus, the standard scores reported here may overestimate the reading level of our participants relative to their age group in England (see Marinus, Kohnen, & McArthur, 2013).

Exposure items (n = 16) and legal unseen items (n = 16) conformed to the following first-order statistics: One vowel predicted C_1 as onset and C_2 as coda, that is, in one counterbalanced list, there was an equal joint probability of vowel *o* occurring with any C_1 as an onset [e.g., P(d, o) = P(m, o) = .125] and any C_2 as a coda [e.g., P(o, t) = P(o, n) = .125]; While the reverse was true for the other vowel [e.g., P(t, e) = p(n, e) = .125; and P(e, d) = p(e, m) = .125]. Consequently, the joint probability of appearance of any C_1 letter and *e* [e.g., p(d, e)] or any C_2 letter and *o* [e.g., p(m, o)] was zero. As a result, participants could benefit from first-order contingencies in both CV and VC portions of the stimuli and these contingencies occurred systematically and with the same statistical probabilities throughout the task. No other statistics were predictive of legality: The probability of appearance of any C_1 or C_2 letter as an onset or coda was equated [e.g., P(d) = P(t) = .125].

Illegal items (n = 16) violated the graphotactic rule: The vowels were preceded and followed by consonants that were not permissible (i.e., had zero probability) during exposure.

The choice of onset, vowel and coda graphemes was made with the intention to conform to English graphotactic rules for permissible onset, vowel and coda spellings for monosyllabic words, to ensure that legality discrimination at test would not be influenced by violation of English orthography. The graphemes were selected and combined with an effort to avoid real English words but seven of the 32 C_1VC_2 sequences (*dot*, *lot*, *mop*, *men*, *den*, *met*, and *let*) and seven of the C_2VC_1 sequences (*ted*, *tel*, *tod*, *tom*, *nod*, *pod*, *sod*) were unavoidably real English words. Since Samara and Caravolas (2014) have found that the pattern of results was not influenced by the presence of real words, these were not excluded from the main analyses reported, however, separate additional analyses are carried out excluding these items when comparing children and adults' performance.

For each of the sixty-four syllable non-words, an audio file was recorded using the voice of a monolingual Southern English female speaker.

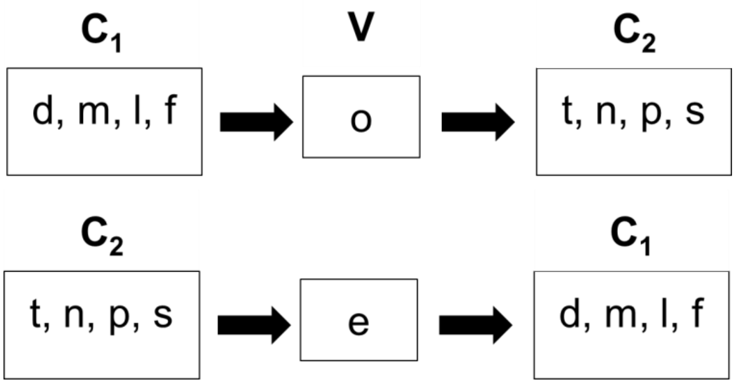


Figure 3.1: Schematic Representation of the Underlying Graphotactic Restrictions in List 1

Literacy Measures. The Word Reading and Spelling subtests of the Wide Range Achievement Test—Forth Edition (WRAT-IV; Wilkinson & Robertson, 2006; green forms) were used to measure children and adults' reading and spelling achievement.

3.2.1.3 Apparatus

The experiment was run on a Windows 7 Enterprise PC with a 13.3-inch CRT color monitor. Visual stimulus presentation and millisecond accurate response registration was achieved using PsychoPy 1.82.01. Participants responded on a standard QWERTY keyboard. The sound stimuli were recorded using Audacity Cross-Platform Sound Editor and participants heard them through Beats BP headphones.

3.2.1.4 Procedure

Child testing was carried out individually in school and adult testing was also carried out individually in the UCL's lab. The experiment ran in PsychoPy3 (Peirce et al., 2019). All participants were seen in two 30-minute sessions over two consecutive days, except for six child participants who completed the sessions with a one-day gap.

Phono-Graphotactic learning task. A child-appropriate learning task was modelled based on Samara and Caravolas (2014). At the beginning of the experiment, children were told that they will be seeing and hearing some words from an alien language, called Zorib, and they would have to play games with Zorib words. In session 1, the game (cover practice and exposure task) was to detect consecutive word repetitions. In session 2, further (covert) exposure was given, followed by disclosure of the patterned nature of Zorib words (i.e., informing participants of the existence of patterns in the words seen). Subsequently, two new games (administered in fixed order) were (i) a task where children were asked to produce Zorib words by filling in a missing letter ("fill-in-the-blanks" test), and (ii) a game where they classified new words as possible/not possible Zorib words ("legality judgment" test). Note that no auditory input was provided in the fill-in-the-blanks and legality judgment tests. Procedures for each task are detailed below.

Practice task. Participants were presented visually and auditorily with real simple English words (*cat*, *dog*, and *cow*) and were asked to detect the consecutive word repetitions by pressing a button when repetitions occurred consecutively (procedure detailed in the section below: Exposure task).

Exposure task. A total of 288 Zorib "words" (144 presented in 3 blocks in each session; 9 repetitions/string in each session) were shown in the context of a one-back cover task (illustrated in Figure 3.2). Participants were instructed to look at each word and (a) press the green button (green tape glued over the 'AltGr' button on the keyboard) when repetitions occurred consecutively (16 in each session) or (b) the red button (red tape was glued over the 'Alt' button). No other instructions or feedback was given. Stimuli were presented in black in the middle of a white background and remained there until a response was given. A response was allowed only after 350ms. A fixation point (black cross, presented for 500ms) followed the response, in the middle of the screen. Word order was manipulated as follows: Consecutive stimulus repetitions occurred once for each of the 16 strings and no more than 6 times in each block. All other stimuli appeared at random and no other doubles were allowed.

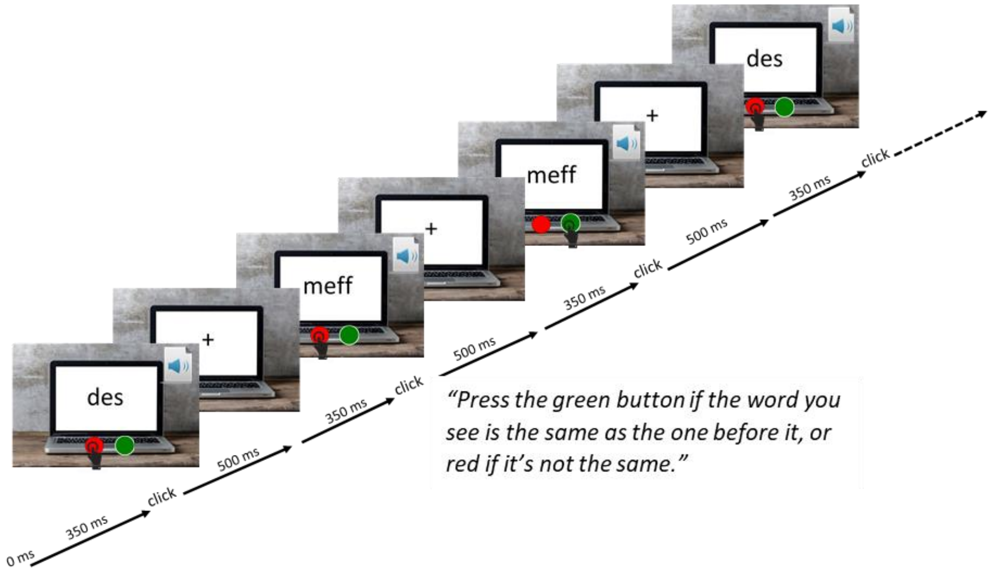


Figure 3.2: Schematic Representation of the Exposure Task

***Fill-in-the-blanks* task**. A fill-in-the-blanks task (illustrated in Figure 3.3) was devised to measure pattern sensitivity as reflected in participants' ability to choose the appropriate context (vowel) to create legal words in Zorib language. In each trial, participants saw (a) novel word frames consisting of a consonant, an underlined blank space for the missing middle vowel, and a word-final consonant (e.g., *d_t*) and (b) below the frame, the two vowels used during exposure (*e*, *o*). The experimenter explained that their task was to drag the vowel and fill the blank to make a word that they thought possible in Zorib language. They were encouraged to use their gut feeling and were allowed to change their mind once they saw the word in full. Stimuli (*n* = 16) were presented one at a time in random order. Note that choosing correct responses made the 16 legal unseen items used in the legality judgment test.

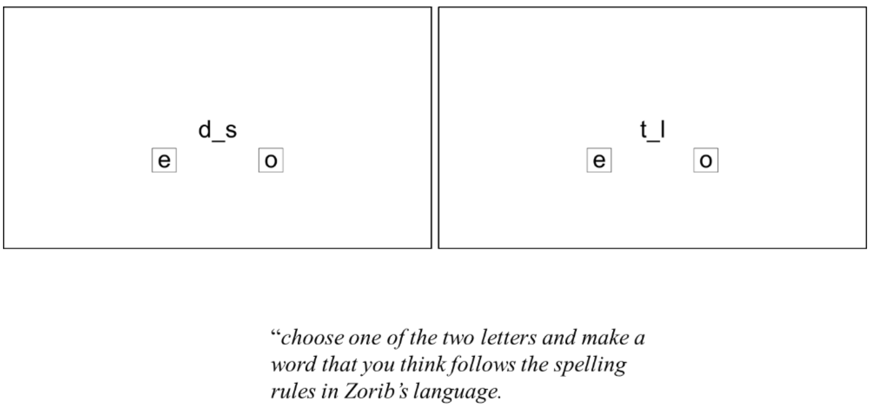


Figure 3.3: Schematic Representation of the Fill-in-the-blanks Task

***Legality judgment* task.** In the legality judgment task (illustrated in Figure 3.4), participants were presented visually only with novel legal unseen (*n* = 16) and illegal (*n* = 16) strings in randomized order and were asked to decide if each of the words could/could not exist in Zorib language and press a corresponding button accordingly. A sticker with the word YES was attached to key *Q* and a sticker with the word NO was attached to key *P*. If unsure, they were encouraged to trust their intuition or "gut feel". Each string was presented in the middle of the screen and remained until a response was given. A fixation point (black cross) appeared for 500ms at the center of the computer screen after each trial. A total of 32 items were presented in a single block.

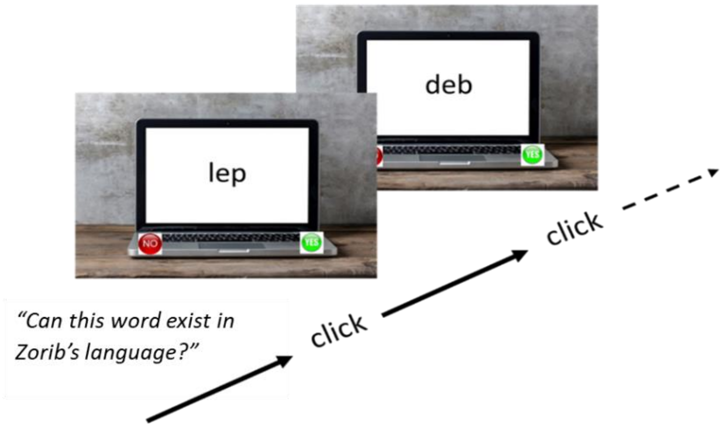


Figure 3.4: Schematic Representation of the Legality Judgment Task

Awareness questionnaire. A brief aural questionnaire was administered to assess whether participants were able to verbalize the graphotactic constraints governing Zorib words. Participants were asked: "Did you notice the patterns/rules that exist in Zorib orthography?" If a participant reported that they noticed patterns before they were informed regarding their presence, further questions probed what patterns they thought they noticed and how they made their choices in each of the two tests: For the fill-in-the-blanks task: "How did you decide which letter to choose to make a Zorib word?"; and for legality judgment: "How did you decide which words can exist in Zoribs' language and which cannot?". The examiner recorded by hand the answers verbatim on a response sheet.

Literacy measures. Participants' reading and spelling skills were assessed using the two relevant subtests of the WRAT-IV (Wilkinson & Robertson, 2006, Green form).

3.2.2 Results

The design and hypotheses of this study were not pre-registered; however, the H_1 for the comparison against chance (for children) was modelled using the same estimate as in the following pre-registered studies (Experiment 3, 4, & 5) that use the same methods as Experiment 1. The data is available at https://osf.io/8fv9d/ for children and https://osf.io/r9yzf/9 for adults, and data analyses are available at https://rpubs.com/DSingh/Phono_Grapho. The priors used to model the theory are detailed in Table 2.1 and Table 2.2.

3.2.2.1 Children

Figure 3.5 shows the mean proportion of children's correct responses in the fill-in-the-blanks task. Our data provided support for H_1, that is, there was evidence that children were better than chance (50%) at choosing the correct vowel, $BF_{(0,0.16)}$ = 38.13, *RR* [0.03, > 4.59] (model intercept: β = 0.39, *SE* = 0.12, *z* = 3.32, *p* = .001). The robustness regions suggest that H_1 is preferred over the null for any minimal value used to model H_1 and beyond the maximum effect we can reasonably expect to observe (99% accuracy, i.e., 4.59 in log-odds space).

Figure 3.5 also shows the mean proportion of children's correct legality judgments. Here as well, there was evidence for above (50%) chance learning, $BF_{(0,0.16)}$ = 46.08, *RR* [0.03, > 4.59] (model intercept: β = 0.29, *SE* = 0.09, *z* = 3.2, *p* = .001), that is, children were better than chance at discriminating between legal and illegal items.

---

[9] See also Appendix B

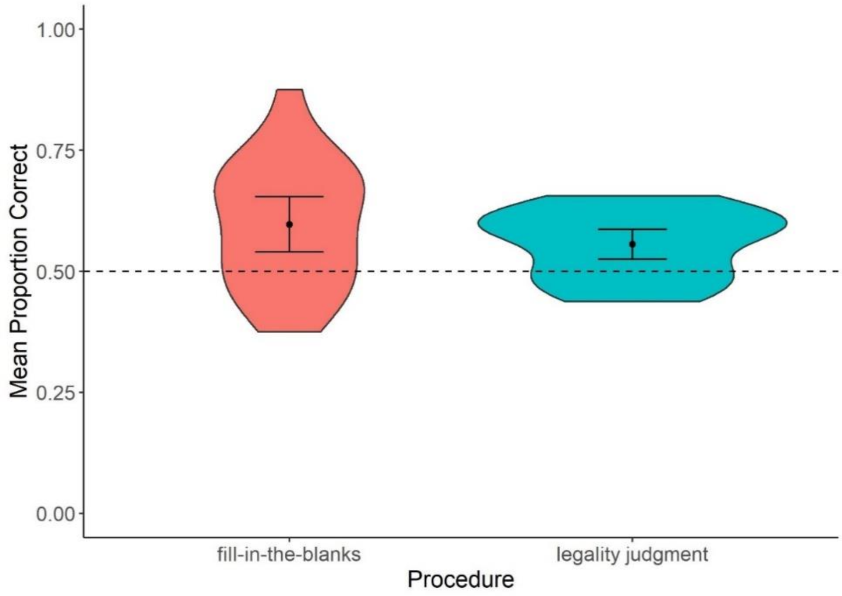


Figure 3.5: Children's Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 1. The Dashed Line Represents Chance-level Performance (50%).

3.2.2.2 Associations between the two tests of graphotactic learning

Our data was inconclusive, that is, we cannot say whether or not children's performance at fill-in-the-blanks task positively correlated with their performance at legality judgment task, $BF_{(0,0.58)}$ = 1.83, *RR* [0, 4.06] (*r*(18) = .35, z_r = 0.36, *SE* = 0.24, *p* = .13).

3.2.2.3 Adults

Figure 3.6 shows the mean proportion of adults' correct responses in the fill-in-the-blanks and legality judgment task. In the fill-in-the-blanks task, we found that adults were above (50%) chance at creating permissible generalization stimuli, $BF_{(0,0.20)}$ = 8491, *RR* [0, >4.59], (β = 1.02, *SE* = 0.18, *z* = 5.68, *p* < .001).

In the legality judgment task, adults were above (50%) chance at discriminating between legal and illegal items, $BF_{(0,0.20)}$ = 462, *RR*[0.04, > 4.59] (model intercept: β = 0.65, *SE* = 0.15, *z* = 4.34, *p* < .001).

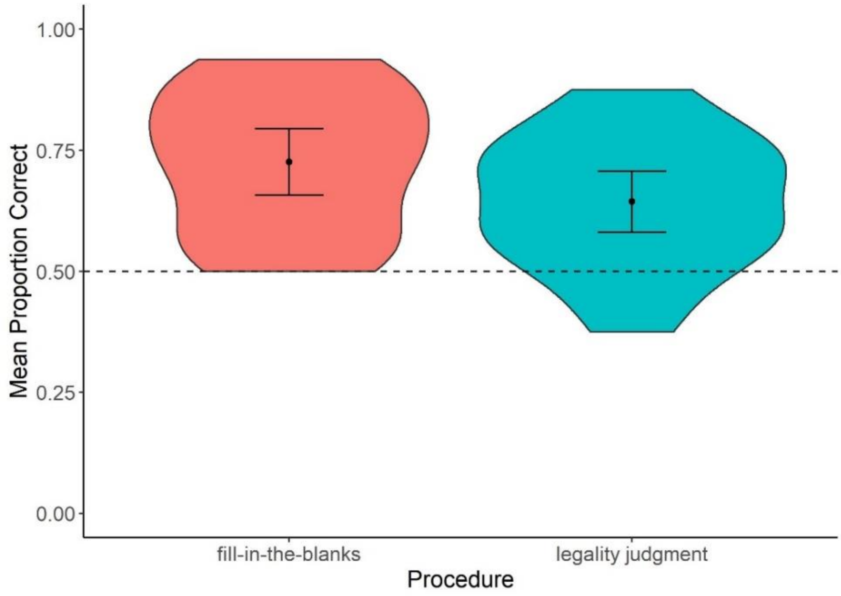


Figure 3.6: Adults' Mean Accuracy (Violin Plots with 95% Confidence Intervals) In the Fill-in-the-blanks and Legality Judgment Task in Experiment 1. The Dashed Line Represents Chance-level Performance (50%).

3.2.2.4 Associations between the two tests of graphotactic learning[10]

Our data provided substantial evidence for H_1, that adults' performance at fill-in-the-blanks task positively correlated with their performance at legality judgment task, $BF_{(0,0.58)}$ = 6.00, *RR* [0.17, 1.89] (*r*(16) = .52, z_r = 0.58, *SE* = 0.26, *p* = .03).

3.2.2.5 Comparison between age groups: children and adults

The results of child and adult phono-graphotactic learning study consistently suggest that both age groups pick up on novel graphotactic constraints from brief incidental exposure. Samara and Caravolas (2014) found that adults outperformed children in contextual constraints learning. However, they found that this pattern of results did not hold when the few real English words (e.g., *nod*, *mop*) were excluded from analyses[11]. Therefore, we compared

[10] See Table 2.1 and Table 2.2 in Chapter 2 for details on how H_1 was modelled

[11] Note that, although the pattern of results in the comparison between age groups did not hold when lexical items were excluded, the overall learning in each age group was not influenced by the presence of real words. As I note in Section 3.2.1, for this reason, the lexical items were not excluded from our other analyses.

children's and adults' performance on both tasks, with and without lexical items, to provide further evidence for an effect of age, even when complex, contextual rather than positional patterns are learned (Cassar & Treiman, 1997; Hayes et al., 2006; Treiman & Kessler, 2006).

3.2.2.5.1 Analyses including lexical items (i.e., real English words[12])

In the fill-in-the-blanks task, we found substantial evidence for better learning in adults compared to children, $BF_{(0,0.11)}$ = 35.43, *RR* [0.08, >4.59], (effect of age: β = 0.6, *SE* = 0.2, *z* = 2.91, *p* < .001). In the legality judgment task, once again, we found evidence for H_1, $BF_{(0,0.11)}$ = 11.54, *RR* [0.08, 2.43 (effect of age: β = 0.38, *SE* = 0.15, *z* = 2.59, *p* = .01).

3.2.2.5.2 Analyses excluding lexical items (i.e., real English words)

The pattern of results seen when analyses were carried out with all items, held when lexical items were excluded: substantial evidence for better learning in adults compared to children in both (i) fill-in-the-blanks task: $BF_{(0,0.11)}$ = 18.61, *RR* [0.08, >4.59], (effect of experiment: β = 0.56, *SE* = 0.2, *z* = 2.77, *p* = .01); and (ii) legality judgment task: $BF_{(0,0.11)}$ = 10.15, *RR* [0.09, 2.34 (effect of experiment: β = 0.42, *SE* = 0.17, *z* = 2.51, *p* = .01).

3.2.2.6 Comparison between studies: Experiment 1 and Samara and Caravolas (2014)

To investigate whether the changes in the current study have indeed boosted learning relative to Samara and Caravolas' (2014) study, we compare performance in the two studies. As detailed in Section 2.2.3, for between-experiment comparisons, we set a constraint on a likely maximum value from the data itself (Method C). Analyses were run for both children and adults, but only for the legality judgment task, since Samara and Caravolas' (2014) study did not have the fill-in-the-blanks task. Figure 3.7 shows children's and adults' performance at the legality judgment task, in the two studies.

[12] See Section 3.2.1 for examples.

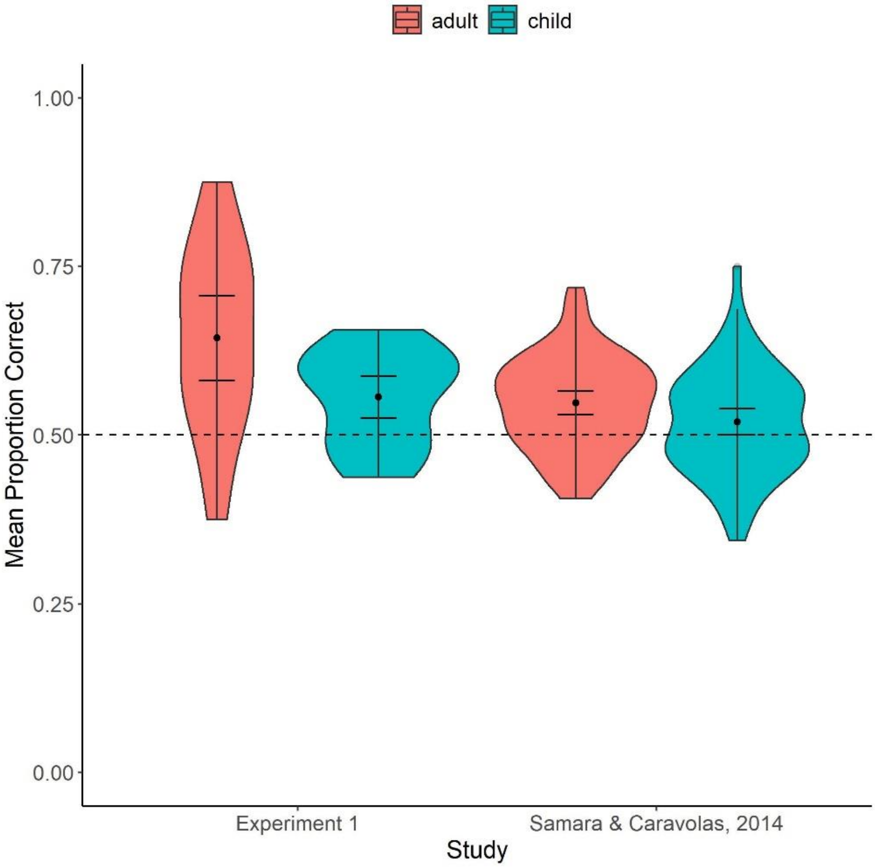


Figure 3.7: Children and Adults’ Mean Accuracy (Violin Plots with 95% Confidence Intervals) In the Legality Judgment Task in Samara and Caravolas (2014) Study and Experiment 1. The Dashed Line Represents Chance-level Performance (50%).

3.2.2.6.1 Children

The evidence was inconclusive, $BF_{(0,0.11)}$ = 2.51, *RR* [1.96, >4.59] (effect of experiment: β = 0.15, *SE* = 0.09, *z* = 1.64, *p* = .1). We do not have substantial evidence that children were better at discriminating between legal and illegal items in Experiment 1, when the phonology was provided, relative to the same constraints in Samara and Caravolas (2014), where phonology was not provided (though also no evidence for the null).

3.2.2.6.2 Adults

There was substantial evidence for H_1, $BF_{(0,0.29)}$ = 888, *RR* [0.03, > 4.59], that is, adults were more accurate in discriminating between legal and illegal items in Experiment 1 relative to Samara and Caravolas' (2014) study (effect of experiment: β = 0.40, *SE* = 0.10, *z* = 4.02, *p* < .001).

3.2.2.7 Associations between learning and literacy performance

We looked for relationships between statistical learning and reading and spelling ability using BF analyses[13] and a novel method of estimating the theory (outlined in Section 2.3.1: Method D), to quantify evidence for both H_1 (positive associations such that those who performed better in the learning tasks were also better readers/spellers) and the null (no relationship between statistical learning and literacy skills).

Results from our correlation analyses are presented in Table 3.1 To sum up the results, in Experiment 1, the evidence for the association between learning performance and literacy was inconclusive in all but two occasions, where we demonstrated the null (no relationship between task performance and literacy): Between children's performance in our legality judgment task and (i) reading and (ii) spelling ability.

In sum, when conclusive patterns of correlations emerged, these were evidence of no relationship between learning performance and literacy.

[13] See Table 2.1 in Chapter 2

Table 3.1: Correlations between Accuracy in Experiment 1 and WRAT Reading and Spelling Raw Scores (by Procedure)

Experiment & Procedure	Statistics	Reading	Spelling
Experiment 1: Children			
fill-in-the-blanks	*BF* [RR]	0.38 [0, 0.51]	0.71 [0, 1.07]
	p	.69	.62
	z_r (SE_{Zr})	-0.09 (0.24)	0.12 (0.24)
legality judgment	*BF* [RR]	**0.27**[a] [0.33, >4.59]	**0.29**[a] [0, >4.59]
	p	.30	.37
	z_r (SE_{Zr})	-0.25 (0.24)	-0.21 (0.24)
Experiment 1: Adults			
fill-in-the-blanks	*BF* [RR]	2.23 [0, >4.59]	0.74 [0, 1.15]
	p	.12	.63
	z_r (SE_{Zr})	0.40 (0.26)	0.12 (0.26)
legality judgment	*BF* [RR]	1.03 [0, 1.76]	0.47 [0, 0.65]
	p	.39	.89
	z_r (SE_{Zr})	0.22 (0.26)	-0.04 (0.26)

[a] substantial evidence for H_0
[b] substantial evidence for H_1

3.2.2.8 Awareness Questionnaire

Participants' responses were coded such that, if any response described at least one aspect of the patterns embedded, such as "*o is followed by d",* that response was coded as *explicit*. Other responses that were vague and described a pattern that was not the actual manipulation, such as "there were more *d*-s in the language" were coded as *other*. Those responses that did not refer to any patterns at all, e.g., "I don't know" or "I guessed", were coded as *nothing relevant*. Only the responses coded *explicit* were considered to represent awareness.

No adults reported awareness or were able to provide a description of an aspect of the patterns. 75% of the responses were "I guessed" or "I don't know" and 25% described patterns that did not correspond to the ones in the input. While children were more detailed in their responses, their answers were not relating to the patterns, e.g., "I figured out the reason for the game: to help in guessing the yes or no words". No participants were excluded on this basis.

3.2.3 Discussion

Previous research demonstrated that graphotactic learning can be induced under brief incidental experimental conditions (Chetail, 2017; Nigro et al., 2016; Samara & Caravolas, 2014; Samara et al., 2019). The work of Samara and colleagues showed that children and adults generalize over relatively complex, contextual constraints embedded at the beginning and/or end of pronounceable made-up words (e.g., *t* can be followed by *o*, never *e* and/or *o* can be followed by *f*, never *t*, as in *tof*). Evidence of a relationship between learning in the lab and literacy attainment has shown mixed results, with some studies reporting weak to medium correlations (e.g., Arciuli and Simpson, 2012) or no correlations at all (West et al., 2017). Issues with measures of implicit learning were recently highlighted (Siegelman et al., 2017), resulting in tasks with low reliability and small (near chance) effects, as seen in Samara and colleagues' studies. For this reason, when looking for correlations between learning in Samara et al.'s (2019) study and literacy measures, the evidence was inconclusive. Experiment 1 builds on Samara and colleagues' work with the key goal of finding stronger learning effects in a modified version of their studies. A crucial cue that was added to boost learning was the overtly presented phonology to the stimuli in Samara and Caravolas (2014). Because the stimuli in their study formed pronounceable nonwords, they provided additional phonotactic cues (e.g., "/t/ can only follow /o/, never /ɛ/") that could have been covertly accessed and used. However, the overt pronunciation was not encouraged, and the stimulus was presented in written form, therefore, it is not clear if these cues contributed to the learning effect. In Experiment 1, we provided audio input with phonology matched to the orthography, to children and adult participants, alongside identical visual stimuli used in the contextual condition in Samara and Caravolas (2014). Research from the statistical learning literature shows that learning of visual patterns is facilitated by correlated auditory structure that is overtly presented (cross-modality facilitation) (Robinson & Sloutsky, 2007). Thus, we hypothesized that participants would benefit from the overt phonology and use their sensitivity to phonotactics to generalize over the graphotactic constraints. We also made other changes that we hoped might boost learning: A different cover task was used which we hoped would focus attention on details of the word form; pattern-embedding stimuli were exposed over two session in consecutive days, rather than a single session and participants' generalization was tested in both legality judgment task (as in Samara & Caravolas, 2014) and a more naturalistic production (fill-in-the-blanks) task where participants were presented with generalization word frames (legal unseen) and were asked to fill in a missing vowel, choosing from two alternatives. To measure the boost provided by these changes, performance in the experimental condition in the current study was compared to chance as well as to that observed in Samara and Caravolas (2014). The positive association between performance in the two test tasks was also explored to find whether both measure similar abilities. Finally, we tested the relationship between incidental phono-graphotactic learning and standardized measures of reading and spelling ability. Bayes factor analyses were employed throughout, to evaluate the evidence for and crucially, against hypotheses.

Consistent with our hypothesis, children and adult participants learned the phono-graphotactic patterns: They reliably selected the appropriate letters and formed pattern conforming novel strings, and discriminated between legal and illegal items with better than chance accuracy. We interpret performance as reflecting implicit learning, given that none of the participants (children and adults) reported awareness of any of the patterns in the questionnaire. For both age groups, performance was numerically higher in both test tasks than in Samara and Caravolas (2014): They report 52% (children) and 55% (adults) in the legality judgment task, whereas here, we see 56% (children) and 64% (adults) in the same task, and 60% (children) and 73% (adults) in the novel, fill-in-the-blanks task. We were able to quantify this increase statistically but only for the legality judgment task (since this was the only comparable task), with those in Samara and Caravolas (2014): We found substantial evidence of higher performance with adults, but the evidence for children was inconclusive (supplementary power analyses revealed that we needed 4 additional participants to find substantial evidence for H1). As in Samara and Caravolas (2014), we observed a developmental effect, with adults performing better than children in the contextual constraints condition, although, unlike in that study, we found that this held even when real lexical items (such as *met* or *pod)* were excluded from analyses. Finally, we also investigated whether incidental phono-graphotactic learning was related to literacy attainment. We hoped that our new paradigm might lead to strong enough performance, that we would see evidence for these relationships, or if not, that we might have clear evidence for the null. In fact, the evidence was inconclusive in all but two occasions. Specifically, we found evidence for the null, that is, no relationship between children's performance at legality judgment and both (i) reading and (ii) spelling ability. We discuss these findings below.

3.2.3.1 Phono-graphotactic learning of contextual constraints

We replicate Samara and Caravolas (2014) finding that children and adults can pick up on context-based graphotactic conditional constraints embedded in word-like visual strings from brief incidental exposure. In the test that was comparable across studies, for adults, we found evidence of stronger learning. We acknowledge the fact that, since we adapted multiple aspects of the paradigm, any conclusions about which aspects led to stronger learning effects in adults are speculative. We also note that we did not have substantial evidence that these same changes made a difference for children. Nevertheless, since, at least for adults, we do see a benefit, apart from the addition of overt phonology (discussed further below), we retain the methodological changes for later experiments (except that the one-back-task is not used in the final two experiments (Experiment 6 and Experiment 7) since we explore learning under a different type of training).

Turning to the fill-in-the-blanks task, this was validated as a new measure of artificial pattern sensitivity. In both age groups, we saw higher performance than in the legality judgment, although this difference was not large. Unlike the legality judgment task (which we also employed for consistency with previous work), production performance simulates more closely what children and adults do in naturalistic situations. Thus, our study goes beyond previous work by showing that children's knowledge of novel orthographic constraints generalize, to some extent, to their own (partial) written productions. Beyond seeing larger effects, having two measures allow us to determine the robustness of learning. Our two tasks correlated for adult data, but again, for children, data were insensitive to test this hypothesis. Nevertheless, at least for adults, the correlation suggests that the tests may both measure similar abilities. This task was retained for the rest of the thesis, with the exception of Experiment 2, where this was not possible due to methodological issues of implementing it (see Section 4.3 for further details).

Before turning to discuss the (lack of) correlations in the next section, I briefly consider limitations of the stimuli used in the current study, which apply equally to Samara and Caravolas (2014). While we show generalization of patterns consistent with rules on letter co-occurrences, we cannot distinguish whether the knowledge obtained is best characterized as rules about letter co-occurrences or as chunks (e.g., Perruchet, 2019; Reber, 1967; see Section 1.5 for details). Since this question is relevant to all the studies reported in this thesis, I will leave it here and return to it in the general discussion (Chapter 7). Another limitation is that the combination of letters that form the CVC strings resulted in a number of real words (lexical items). It is possible that learning was affected by familiarity with these words, and decisions at test were made based on their legality as English words rather than novel, alien words. In fact, the awareness questionnaire revealed that some participants rejected the lexical items on the basis that they could not be in Zorib language, if they were English words. Importantly, however, the list allocation was such that each list contained a small number of lexical items and these were counterbalanced between participants so that stronger/weaker learning for real word items could not drive the pattern of results reported. Note that this type of list allocation can also control for the possibility that some participants find certain bigrams more familiar than others, although every effort was made so that the frequency of these in the English corpus was similar. Another potential limitation in the design of the stimuli was that the patterns were embedded in both beginning and end of strings: Both CV and VC contingencies were presented equally often and thus, participants could have relied on cues from either onset, rime, or both units for learning. As noted in the General Introduction (Section 1.1) such redundant cues are not generally present in naturalistic spelling. However, this limitation was addressed in Samara et al. (2019), who demonstrated learning of similar constraints based on onset/rimes alone. Our choice to use the redundant cues here was in order to make a direct comparison with Samara and Caravolas (2014).

3.2.3.2 Relationship between phono-graphotactic learning and reading and spelling skills

A key question in this chapter concerned the widely hypothesized link between statistical learning processes and literacy performance. A few previous studies have empirically demonstrated that variations in lab-based statistical or implicit learning skill are related to literacy performance among typically developing children and adults (most notably, Arciuli & Simpson, 2012; Frost et al., 2013; Misyak & Christiansen, 2012; Spencer et al., 2015), but others have not (Schmalz et al., 2019; West et al., 2017). Although, as hoped, at least for adults, we saw stronger learning effects in the current experiment than in Samara and Caravolas (2014), we did *not* find any association with literacy measures. Using Bayes factors, we were able to conclude that there is no relationship here, at least for the learning as measured by the legality judgment task and in children. However other correlations—that is, with the production task for children and with both tasks for adults—were inconclusive (i.e., are neither evidence for H_1 or for H_0). This highlights that nonsignificant correlations do *not* necessarily constitute evidence for the null, as often interpreted in the literature. For example, a recent study by Qi, Sanchez Araujo, Georgan, Gabrieli, and Arciuli (2019) reported significant associations between auditory but not visual statistical learning ability and reading fluency and took this as evidence that "hearing is more important than seeing" for literacy development. In line with Dienes (2014), we caution against these statements on the basis of frequentist results alone.

The fact that we continue to see many inconclusive results in this study may be because the reliability of our tests was low[14], in line with the concerns raised by (Siegelman et al., 2017) and (West et al., 2017). In addition, although at least for adults we did see higher performance in this experiment, still, performance was quite weak (around 6% higher than chance for children, 14% for adults). This suggests that we did not succeed in our goal of getting sufficiently strong learning in the graphotactic task in order to be able to detect correlations with other tasks. On reflection, we conclude that getting stronger learning would likely require more training sessions (i.e., beyond the two we use) and this is impracticable for the rest of the thesis. Therefore, instead of looking for stronger effects per se, the rest of the thesis focusses on other investigations that I explain at the end of this section.

In sum, our experiment investigated whether stronger effects can be measured in a new version of Samara and Caravolas' (2014) task, where overt phonological information was provided along with the visual stimuli, as well as other methodological changes designed to boost performance. We found that, although both children and adults learned the graphotactic context-based

[14] See Appendix H for reliability calculations

patterns, only for adults was there evidence that performance was stronger in this experiment than the previous experiment and, critically, even for adults we did not see evidence of correlations with tasks of literacy.

As noted above, we cannot say with certainty whether improved performance in this experiment, which was seen at least in adults, was due to the overt availability of phonology. However, this was certainly the most substantial change to the experiment and it is likely that this is the case. This would indicate that overt phonology was useful and used (at least by adults), which is in line with research from statistical learning literature that showed cross-modality facilitation from the auditory to visual modality when the underlying statistics are matched (Mitchel & Weiss, 2011; Robinson & Sloutsky, 2007). If so, it raises the intriguing possibility that participants in Samara and Caravolas (2014) did not access or use the covert phonology at all, in which case they may have, indeed, engaged in purely graphotactic learning, with no phonotactic support. This raises the interesting question of whether "purely" graphotactic learning can occur in both child and adult learners. This is an important question since it is well established that phonotactic knowledge about the native language is acquired in infancy (e.g., Jusczyk et al., 1993) and brief listening experience (Chambers et al., 2003, 2010; Endress & Mehler, 2010; Onishi et al., 2002) is sufficient to induce learning of purely phonotactic constraints in the auditory domain. Does graphotatic learning *rely on* this ability in the auditory domain? This is the key question for the remainder of this thesis. Instead of augmenting stimuli to add phonological cues, we address this question in further experiments by *removing* the correlated phonotactics altogether from the stimuli, and by creating rules over either (i) novel graphemes (new symbols) which have no associated phonemes, or (ii) familiar graphemes which share the same phonology. I turn to this in the following chapters.

4 Incidental Learning of Novel Semi-artificial Positional Patterns

4.1 Introduction

Humans' ability to pick up on regularities from the environment over time (Schapiro & Turk-Browne, 2015) is not restricted to auditory linguistic stimuli, be it syllable sequences (Saffran et al., 1996) or phonotactics (e.g., Onishi et al., 2002), but extends also to non-linguistic auditory (tones; e.g., Saffran et al., 1999) and visual stimuli (geometrical shapes and spatial positions; e.g., Fiser & Aslin, 2002; Kirkham et al., 2002; see also Conway & Christiansen, 2006; Perruchet & Pacton, 2006 for more examples). Grahotactics, the topic of the current thesis are visual statistical regularities in written language. As we saw in Section 1.3, these can restrict either the position or context of letter distributions in a deterministic (absolute, all-or-none, legal vs. illegal) or probabilistic (frequency) way, and, as with sound-based phonotactics (Mattys et al., 1999), they are language specific (e.g., *ll* is frequent in Welsh and Spanish in word beginnings but not in English). Sensitivity to graphotactic regularities has been shown to emerge early on, with some patterns easier to learn, such as restrictions on the position of certain letters in words (e.g., *ck* or doublets cannot start English words). Knowledge of more complex patterns such as context-based regularities in the orthography (e.g., the word-final sound /z/ is spelled as *s* when preceded by *g*, as in *bags*) takes longer to develop and reaches adult-like levels only in more advanced spellers (Hayes et al., 2006; Treiman & Kessler, 2006). Sensitivity to graphotactic regularities has been tested predominantly using patterns that exist in participant's own orthography (Kessler, 2009; Pollo et al., 2007; Steffler, 2001; Treiman & Kessler, 2013; see also Section 1.3). As discussed in Section 1.6, Samara and Caravolas (2014) provided a first demonstration of graphotactic learning under incidental experimental conditions, extending the methodology used in studies that explored phonotactic learning. However, as discussed in Chapter 3, using pronounceable strings meant that the graphotactic regularities all had phonotactic counterparts so it does not exclude the possibility that the learning effects were fully or partially driven by phonotactic sensitivity.

Using participants' native alphabet may introduce other confounds, in addition to their pronounceability: Regularities manipulated in experimental studies using familiar letters co-vary with those in natural languages, such as frequency and orthographic neighborhood. Such confounds may have been the reason behind inconsistent findings in the visual word processing literature, in regards to the role of graphotactic learning (Chetail, Balota, Treiman, & Content, 2015; Gernsbacher, 1984; Keuleers, Lacey, Rastle, & Brysbaert, 2012). Vidal, Content, and Chetail (2017) identified these confounds and devised a set of characters (BACS-1, The Brussels Artificial Character Sets) to allow researchers to have control over manipulations when examining the mechanisms involved in visual word recognition. When aiming to understand the processes involved in activation of letter representations when recognizing words (e.g., McClelland & Rumelhart, 1981), researchers can use such simple objects (characters) while removing their associated characteristics, such as shape or sound. Artificial or unknown scripts have been used in manipulations with print-to-sound mappings with some studies using scripts unfamiliar to participants (e.g., Greek or Thai script for monolingual English or French participants) as well as ancient, unused scripts, e.g., Phoenician Moabite alphabet, and newly designed characters. For example, one study of J. Taylor, Plunkett, & Nation (2011) used an artificial script to investigate print-to-sound consistency effects by creating an artificial orthography using a set of symbols (also used in the current study) that were considered letter-like in terms of configuration of strokes, and manipulating the consistency of vowel character to phoneme mapping in CVC monosyllabic novel nonwords. While each consonant sound was represented consistently by one character (aptly referred to in the study as *letter*), the vowels could either be consistent (i.e., pronounced the same way in all items), inconsistent-conditioned (i.e., pronounced one way when preceded by a particular consonant) or inconsistent-unconditioned (i.e., pronounced in a different way than the conditioned one when preceded by any other consonant). Adult participants familiarized themselves with the items (i.e., viewed each item, listened to the pronunciation and repeated it once), then, in a training phase, they were asked to pronounce them while looking at the visual items. One more correct pronunciation was then provided as feedback. In a lexical decision task, participants were able to discriminate between trained and untrained items, and, in a generalization task (i.e., read aloud items that conformed to the training set), they read accurately items that consistent and/or high frequency during training. The researchers also manipulated the frequency of the vowels and responses (accuracy and reaction times) to items containing these showed that the participants were sensitive to the statistical properties of the mappings between phonemes and graphemes in their learning environment. The effects of consistency and frequency resembled those in natural languages effects, and thus, J. Taylor et al. (2011) validated the use of an artificial orthography paradigm for investigating the factors important in learning written language. Similar results were found by Brooks, Rosch, and Lloyd (1978) and Yoncheva, Blau, Maurer, and McCandliss (2010), who both found that adults can learn novel print-to-sound mappings and use them to read words in an unfamiliar script; and explicit teaching facilitated learning, supporting a phonic method of instruction (i.e., systematic teaching of print-to-sound mapping). Most relevant to the current thesis are studies that use artificial orthographies to explore graphotactic learning using visual stimuli to mimic orthography, but which are unpronounceable, thus avoiding the confounds with phonotactics.

To some extent, the studies discussed above, which used geometric shapes (e.g., Kirkham et al., 2002) and abstract shapes (e.g., Fiser & Aslin, 2002) to look at visual statistical learning over spatial elements (e.g., Fiser & Aslin, 2001), explored learning of graphotactic-like patterns. However, these only tested infants and adults' ability to extract patterns using trained items, and the ability to generalize, which is critical to graphotactic learning, was not tested (see Section 1.5.2). Two studies that are particularly relevant and that have tested generalization over graphotactic patterns are Chetail (2017) and Nigro et al. (2016), which were briefly introduced in Section 1.6. I will review these studies in more detail below.

Chetail (2017) conducted a study looking at whether adult participants could learn and generalize context-based patterns in a novel script, using 22 characters from the Phoenician Moabite alphabet to embed pairs of symbols (bigrams occurrences, also referred to as "critical item") either at the beginning or at middle of five-character strings. Following exposure to these pattern-embedding stimuli containing the critical items, they tested generalization with a wordlikeness task, whereby participants were asked to judge which of two novel stimuli was more like the words seen at exposure. The test stimuli were pairs of strings: (i) One embedding the trained frequent bigrams in their corresponding position (critical item) while the rest of the characters were random, and (ii) the other containing two frequent characters in their corresponding position but that never occurred together (non-critical item), with the other characters being random. Adults discriminated above chance between the patterned and non-critical items, showing sensitivity to the *context-based* patterns. An important methodological aspect of this study was that, before starting the exposure phase, the participants were familiarized with the novel symbols by asking the participants to hand-copy each of them on a sheet of paper followed by an independent study of all symbols. J. Taylor et al. (2011) had a similar methodological approach when they included an initial, exposure phase before training, as described above. This type of exposure may be necessary for studies using novel scripts, but might also be particularly important since in Chetail (2017), the patterns were relatively difficult to learn (i.e., context-based patterns) even for adults, compared to the positional patterns, as noted in Section 1.6 (e.g., Samara & Caravolas, 2014).

A similar experiment to Chetail's (2017), that is, looking at learning of context-based graphotactics, has not been conducted with children, probably due to recognition of the fact that children show weaker learning in these experiments (Samara & Caravolas, 2014; see Section 1.6). Furthermore, with additional difficulty of using unfamiliar, artificially created symbols, teaching more complex context-based constraints might not be possible in the context of a short experiment. However, Nigro et al. (2016) have conducted an experiment looking at learning of *positional* constraints in an artificial script. An additional feature of this experiment was that both typically developing and dyslexic 8–9-year-old children were included to determine whether dyslexic children differ from typically developing ones in their implicit learning ability. They used 10 abstract shapes (from Fiser & Aslin, 2001) to form four-shape strings embedding constraints at the beginning and penultimate position in the string: One of three specific shapes could appear in first and fourth position, while the other two positions could be filled by any of the four remaining shapes. They exposed the participants to pattern embedding stimuli in one session of a child-friendly game and then tested them on the same wordlikeness task as in Chetail (2017). The test items were either some of those seen at exposure (trained, also referred to as legal seen) or novel, combinations of unseen symbols that were placed around the two set positions (legal unseen). They found that, while both groups correctly identified the training (legal seen) items, and dyslexic children were not significantly worse that typically developing children, the dyslexic children's performance as a group was not above chance for novel items (legal unseen), taken to suggest that transfer of such knowledge is more challenging for them (although, as only frequentist *p*-values were used, it is not established that they showed no generalization).

While Nigro et al. (2016) study shows that children can learn purely visual patterns, the shapes did not have the letter-like features that distinguish the characters of a writing system, even if unfamiliar (C. Vidal et al., 2017), which means that the relevance to a writing system is limited, since children may not have viewed them as orthography. In the current study (Experiment 2), we use artificial symbols that are more letter-like, borrowed from J. Taylor et al. (2011). Importantly, we use a combination of these symbols and real English letters (consonants), as this has the benefit of both (i) having less "novel" symbols for the children to learn (and thus avoiding a lengthy pre-training as in Chetail (2017) and J. Taylor et al. (2011), but also, and critically, (ii) making it clearer to the learners that these are intended to be graphemes, whilst removing the phonotactic confound (since they are still unpronounceable). In addition, using real letters allows an additional manipulation: To assess generalization not only using re-combinations of the graphemes (English and artificial) used in training, but also including English consonants not included in exposure stimuli. Seeing generalization to these items can further establish that children see these as graphotactic constraints that govern the positions of graphemes, so that they are willing to generalize them across other graphemes not included in exposure.

4.2 The current study

The key goal in Experiment 2 was to investigate if children and adults can learn purely visual, positional graphotactic constraints embedded in a semiartificial orthography, incidentally, via statistical learning processes. We addressed this by using semiartificial stimuli where the "word"-like stimuli were not pronounceable and no pronunciation was provided. Specifically, we used three-character strings, where the middle character was a familiar English consonant, while the edge characters were unfamiliar grapheme like symbols (taken from J. Taylor et al. 2011).

As in the Samara and Caravolas (2014) study, the embedded pattern was deterministic (rather than probabilistic): Symbols from one set of novel symbols (set 1) were restricted to onset (beginning of string) positions and never appeared as codas (end of strings). Inversely, Symbols from set 2 were restricted to codas and never appeared as onsets. The word-medial consonants varied and were randomly drawn from a subset of English consonants. Note that, since both constraints on onsets and codas appeared simultaneously, participants were given both cues to legality. For example, participants could infer legality by attending to positional constraints on onsets, codas or both (as in Samara & Caravolas, 2014).

Both children and adults were exposed to stimuli embedding this pattern in two brief sessions. As in Experiment 1 (Section 3.2), we modified the exposure task used in Samara and Caravolas (2014) to encourage participants to attend to the form of the nonwords as a whole rather than their irrelevant properties such as color. Here, this was achieved by replacing the letter color-detection task (where one letter changed color) to a whole-word color detection task. Following exposure, learning generalizations were tested in a legality judgment task (used in Experiment 1 and Samara & Caravolas, 2014). Participants were required to provide yes/no answers to unseen stimuli that either conformed to or violated the learned patterns.

In contrast to Nigro et al. (2016), the test items were constructed such, that they were all novel and had not occurred in the exposure test. However, two types of generalization items were included: For one set of test items (*trained consonant* items), the English consonant that appeared in the middle of the three symbol string was drawn from the set of consonants which occurred at exposure, although it had never occurred in this particular frame; in a second set of *untrained consonant* test items, the consonant in the middle of the three symbol string was an English consonant not used at all at exposure. Our analyses looked at whether there was above chance generalization for both types of test items—that is, trained consonant and untrained-consonant—separately, and also whether there was a difference between learning these two types of consonants. We expected to find above chance performance in both cases, but poorer performance with untrained consonants.

As in the other studies presented in this thesis, this study was run with both children and adults. A further comparison between their performance was carried out (predicting a benefit for adults).

Finally, as in Experiment 1, we explored associations between graphotactic learning ability and literacy performance by administering standardized tests of English word reading and spelling ability to all children (but not adults[15]).

For all participants, as in previous experiments, we also administered a post-experiment Questionnaire: For children, the questionnaire was verbal, while adults were required to write down their answers.

4.3 Experiment 2

4.3.1 Method

4.3.1.1 Participants

The sample size was estimated using the procedure outlined in Section 2.6, however, this plan was not pre-registered. Nevertheless, the details of this plan have already been established within our lab, and therefore, we recruited as many participants as practically possible (in excess of the 25 required for a first look at the data) and stopped if we found substantial evidence for above-chance learning in the graphotactic tests, with this first sample.

Thirty-five typically developing Year 2 children (19 female, 16 male; mean age = 6.7 years, *SD* = 0.27) took part in the study. They were all recruited using an opt-out procedure from a primary school in London, had no known language, hearing or vision impairments and no history of learning difficulties. As expected in a London primary school, there was a mix of language backgrounds: 4 children reported English as their second language (with Polish, Portuguese, French, and German as first languages); 11 children reported being bilingual, with English as a first language (with French, Italian and Polish as second languages); while 20 were monolingual English speakers, and all 35 had received the same amount of formal literacy tuition (2 years). Children were rewarded with stickers and a

---

[15] Adults were tested online and due to copyright and data protection, standardized tests could not be administered.

certificate. As in previous work (Samara et al., 2019), the mean reading and spelling performance in our sample was above average (mean reading efficiency as measured by TOWRE = 116, *SD* = 14.45; mean reading as measured by WRAT = 119.48, *SD* = 10.69; mean spelling as measured by WRAT = 117.97, *SD* = 13.8), which is relatively typical in experimental studies with child participants.[16]

Twenty-seven adult participants (17 female, 10 male; mean age = 32.6 years, SD = 11.04) also participated in the study. They were recruited via Prolific (https://www.prolific.co/) online participant recruitment platform and performed the experiment on their own devices. Twenty-three of the adults reported being monolingual native speakers of English or bilingual with English as first language, while four reported being bilingual, with English as second language (e.g., Polish and Portuguese as first language). They reported no language, hearing, vision impairments or any learning difficulties. They all provided informed consent and were paid for their participation.

4.3.1.2 Materials

Semiartificial Positional Graphotactic learning task. Following the paradigm of Samara and Caravolas (2014) but using symbols instead of the consonants and consonants instead of the vowels in their CVC strings, the zero-order positional constraints were maintained, so that certain symbols could begin "words" (i.e., were possible 'onsets') and other symbols could end "words" (i.e., were possible 'codas'). The symbols were taken from J. Taylor et al. (2011). Two sets of symbols (Set 1: symbol1 and symbol2, and Set 2: symbol3 and symbol4; see Figure 4.1 for an illustration) were combined into eight semiartificial stimuli Symbol_Symbol frames. Two pairs of lists were created by using Set 1 symbols as onsets and Set 2 as codas for the first pair of lists (List 1 and List 2, e.g., symbol1_symbol3), and the reverse for the second pair of lists (List 3 and List 4, e.g., symbol3_symbol1). Four English consonants (*c, g, s,* and *v*) were used to

[16]TOWRE-2 and WRAT-IV standardization is drawn following normative data collected from the US, where formal literacy instruction begins one year later relative to the UK. Thus, the standard scores reported here may overestimate the reading level of our participants relative to their age group in England (see Marinus et al., 2013).

fill in the frames, giving rise to a total of thirty-two semiartificial (i.e., Symbol-Consonant-Symbol) unpronounceable made-up “alien” words. The presentation of these lists was counterbalanced within each experimental group (children and adults) such that, in one counterbalanced list conditions, one Set 1-C-Set 2 list (e.g., List 1) served as exposure items, whereas the other list (e.g., List 2) served as legal unseen items at test. The Set 2-C-Set 1 items (List 3) served as illegal items during test phase. Each counterbalanced list consisted of twenty-four sequences. In other words, for half of the participants Set 1 symbols were onsets and Set 2 symbols were codas, while for the other half of the participants, the reverse was true.

In addition to the thirty-two items, another eight test items were created by adding four new consonants to fill the symbol frames (e.g., *r*, *k*, *t*, and *d*), to create legal unseen with untrained consonants and illegal unseen untrained consonants.

Four different versions of each list were created, in order to control for any effect of middle consonant familiarity. Each version used a different set of consonants for the exposure, legal unseen and illegal groups (e.g., Version 1, as described above: *c*, *g*, *s*, and *v*; Version 2: *d*, *k*, *r*, and *t*; Version 3: *b*, *f*, *j*, and *n*; and Version 4: *l*, *m*, *p*, and *x*) and new consonants for legal and illegal new items (e.g., Version 1 used Version 2 consonants for the new items, and so on). Version assignment was also counterbalanced between participants. Figure 4.1 lays out the structure of training and test items as seen by one particular group of participants (assigned to the counterbalance *Version 1 List 1*, and the full set of items used in this experiment is shown in Appendix D.

Exposure items (n = 8), legal unseen items (n = 8), and legal new items (n = 4) conformed to the following zero-order statistics: In two (List 1 and List 2) of the four lists, for example, there was an equal probability of appearance of any Set 1 symbol as an onset [e.g., p(symbol 1) = p(symbol 2) = .50], whereas the probability of appearance of any Set2 symbol as an onset [e.g., p(symbol 3) = p(symbol 4)] was zero. The reverse was true for the other sets of symbols, that is, there was an equal probability of appearance of any Set 2 symbol as a coda [e.g., p(symbol 3) = p(symbol 4) = .5], whereas the probability of appearance of any Set 1 symbol as a coda [e.g., p(symbol 1) = p(symbol 2)] was zero. As a result, participants could benefit from zero-order positional contingencies in both onset and coda portions of the stimuli and these contingencies occurred systematically and with the same statistical probabilities throughout the task. No other statistics were predictive of legality: The probability of appearance of any middle consonant was equated [e.g., p(c) = p(g) = .25], and the probability of any Set 1 symbol preceding and any Set 2 symbol following each middle consonant was equated [e.g., p(symbol 1, c) = p(c, symbol 3) = .25].

Illegal items (n = 8) and illegal new items (n = 4) violated the graphotactic rule: In the same two (List 1 and List 2) of the four lists, Set 1 symbols appeared in coda position and Set 2 symbols appeared in onset positions, and these were not permissible (i.e., had zero probability) during exposure.

Literacy Measures. The Word and Nonword subtests of the Test of Word Reading Efficiency (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012), along with the Word Reading and Spelling subtests of the Wide Range Achievement Test-Forth Edition (WRAT-IV; Wilkinson & Robertson, 2006; green forms) were used to measure children's reading and spelling achievement.

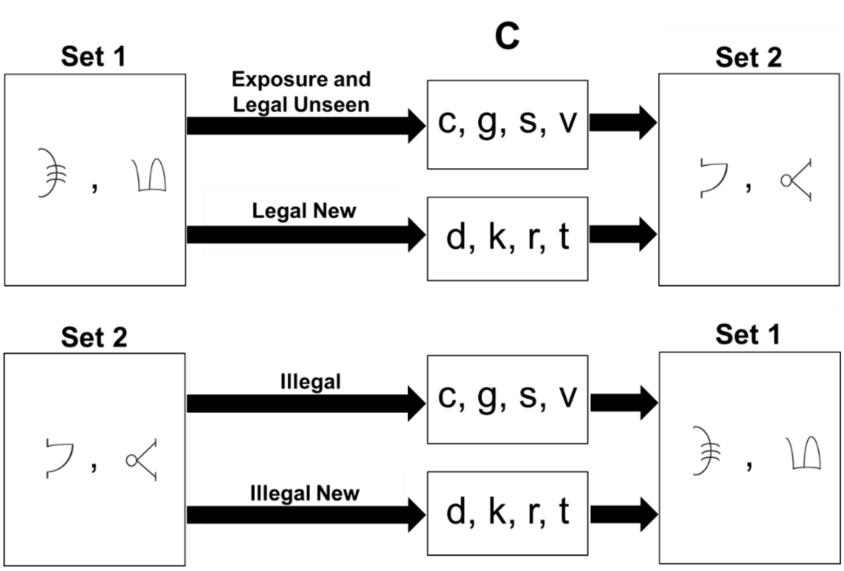


Figure 4.1: Schematic Representation of the Underlying Graphotactic Restrictions in Version 1 List 1

4.3.1.3 Apparatus

The experiment was designed in Gorilla (Anwyl-irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019) and conducted with children using a touch-screen iPad, while adult testing was carried out online using a link distributed via Prolific (https://www.prolific.co/) and was run on participants' own devices at home.

4.3.1.4 Procedure

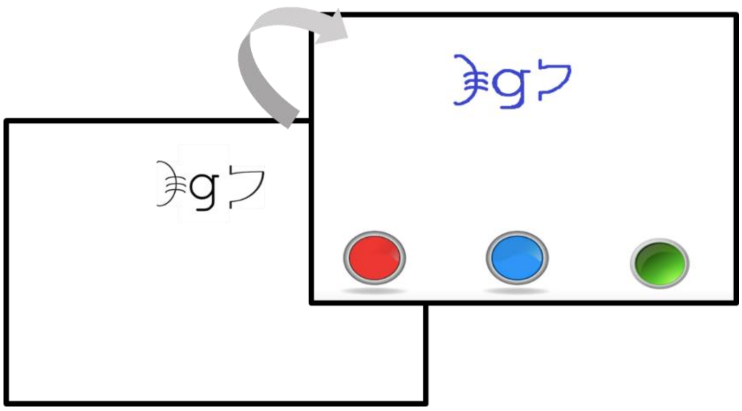
Child testing was carried out individually in school and the experiment ran on experimenter's touch-screen iPad using a web link to the Gorilla study. All children were seen in two 30-minute sessions over two consecutive days. Adults completed the tests independently, by following the instructions on screen (same as those provided to children face-to-face).

Semiartificial Positional Graphotactic learning task. The child-appropriate learning task, as in Chapter 3, was modelled based on Samara and Caravolas (2014) and Samara et al. (2019). At the beginning of the experiment, children were told that they will be seeing some words from an alien language, called Zorib, and they would have to play games with Zorib words. In session 1, the game (cover practice and exposure task) was to detect the color of the words when they turned from black to either blue, green or red. In session 2, further (covert) exposure was given, followed by disclosure of the patterned nature of Zorib words (i.e., informing participants of the existence of patterns in the words seen). In a subsequent game, they classified new words as possible/not possible Zorib words ("legality judgment test"). Procedures for each task are detailed below.

Practice task. Participants were presented visually with real simple English words (*cat*, *dog*, *cow,* and *hen*) and were asked to detect the color of the words when changing from black to either green, blue or red, by pressing a corresponding button (procedure detailed in the section below: Exposure task).

Exposure task. A total of 288 Zorib "words" (144 presented in 3 blocks in each session; 18 repetitions/string in each session) were shown in the context of a color detection task (illustrated in Figure 4.2). Each of the thirty-two items in each version and list had a colored version: blue, green, and red, and each stimulus changed from black to each of these three colors with equal frequency but in random order. Participants were instructed to look at each word and, as soon as they saw the word change color, to press (touch the screen) the button that was the same color as the word: either blue, green or red. No other instructions or feedback was given. Stimuli were presented in black in the middle of a white background and, after 350ms, they turned blue, green or red. As soon as the stimuli changed color, three buttons appeared at the bottom of the screen: red button on the left, blue button in the middle and green button on the right.

Thus, a response was only allowed after 350ms. While participants were asked to respond as soon as they saw the word change color, the stimulus remained on the screen until a response was given. A fixation point (black cross, presented for 500ms) followed the response, in the middle of the screen. All stimuli were randomized for each participant.



"Press the button of the same color as the letters"

Figure 4.2 Schematic Representation of the Exposure Task

Legality judgment **task.** In the legality judgment task (illustrated in Figure 4.3), participants were presented visually with novel legal unseen (n = 8), legal new (n = 4), illegal (n = 8) and illegal new (n = 4) strings in randomized order and were asked to decide if each of the words could/could not exist in Zorib language and press a corresponding button accordingly. A red button with "no" written on it appeared on the bottom left of the screen, while a green button with "yes" appeared on the bottom right of the screen. The participants were encouraged to trust their intuition or "gut feel" if unsure. Each string was presented in the middle of the screen and remained until a response was given. A fixation point (black cross) appeared for 500ms at the center of the computer screen after each trial. A total of 24 items were presented in a single block.

"Can this word exist in Zorib language?"

Figure 4.3: Schematic Representation of the Legality Judgment Task

Awareness questionnaire. As in all experiments in this thesis, a brief questionnaire was administered to assess whether participants were able to verbalize the graphotactic constraints governing Zorib words. In this experiment, instead of asking the questions aurally (see Section 3.2.1.4 for specific questions), these were presented on screen and responses were recorded on Gorilla (see Appendix E for an illustration).

Literacy measures. Child participants' reading and spelling skills were assessed using two relevant subtests of the TOWRE-2 (Torgesen et al., 2012) and WRAT-IV (Wilkinson & Robertson, 2006, Green form).

4.3.2 Results

As in Experiment 1, the design and hypotheses of this study were not pre-registered; however, the H_1 for the comparison against chance (for children) was modelled using the same estimate as in the following pre-registered studies (Experiment 3, 4, & 5) that use the same methods as Experiment 1.

The data is available at https://osf.io/c2hqk/ for children and https://osf.io/b3x5n/ for adults[17], and data analyses are available at https://rpubs.com/DSingh/SemiartificialPositional. The priors used to model the theory are detailed in Table 2.3.

I present separately the results from tests with novel semiartificial stimuli that embedded *trained* and *untrained* word-medial consonants (i.e., consonants seen—trained, or not seen—untrained, at exposure) and compare children's and adults' performance with these two types of stimuli.

4.3.2.1 Children

4.3.2.1.1 Trained consonants

Figure 4.4 shows the mean proportion of children's correct legality judgments for semiartificial stimuli containing *trained* consonants. There was evidence for above chance learning, $BF_{(0,0.16)}$, = 152, *RR* [0.04, > 4.59] (model

[17] See also Appendix B

intercept: β = 0.64, *SE* = 0.15, *z* = 4.29 *p* < .001), that is, children were better than chance at discriminating between legal and illegal items.

4.3.2.1.2 Untrained consonants

Figure 4.4 also shows the mean proportion of children's correct legality judgments when the consonants were *untrained*. There was evidence for above (50%) chance learning, *BF* = $4.57_{(0,0.16)}$, *RR* [0.10, 1.04] (model intercept: β = 0.15, *SE* = 0.33, *z* = 2.26 *p* = .02), that is, children were better than chance at discriminating between legal and illegal items.

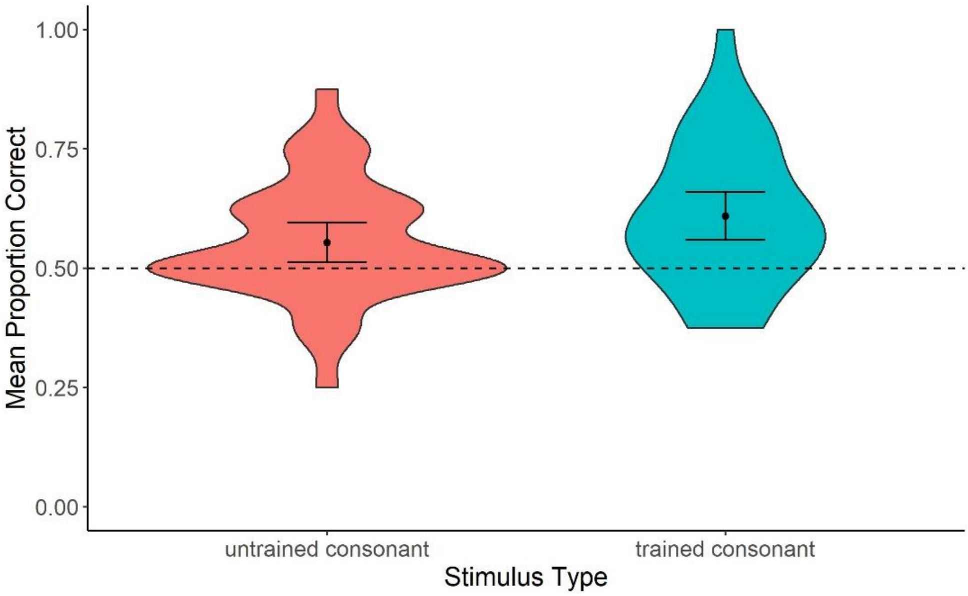


Figure 4.4: Children's Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Legality Judgment Task in Experiment 2, by Type of Consonant (Trained and Untrained). The Dashed Line Represents Chance-level Performance (50%).

4.3.2.1.3 Comparison between trained and untrained consonants

We found no conclusive evidence for H_1, that is, that children would be better with *trained* than *untrained* consonant test items $BF_{(0,0.49)}$ = 2.77, *RR* [0, 0.18] (effect of type of consonant: β = 0.31, *SE* = 0.17, *z* = 1.84 *p* = .07).

4.3.2.2 Adults

4.3.2.2.1 Trained consonants

Figure 4.5 shows the mean proportion of adults' correct legality judgments for semiartificial stimuli containing *trained* consonants. There was evidence for above chance (50%) learning, $BF_{(0, 0.19)}$ = 184, *RR* [0.06, > 4.59] (model intercept: β = 1.30, *SE* = 0.26, *z* = 5.07 *p* < .001), that is, adults were better than chance at discriminating between legal and illegal items.

4.3.2.2.2 Untrained consonants

Figure 4.5 also shows the mean proportion of adults' correct legality judgments when the consonants were *untrained*. There was evidence for above (50%) chance learning, $BF_{(0, 0.20)}$ = 10.66, *RR* [0.07, > 4.59] (model intercept: β = 0.50, *SE* = 0.18, *z* = 2.73 *p* = .01), that is, adults were better than chance at discriminating between legal and illegal items.

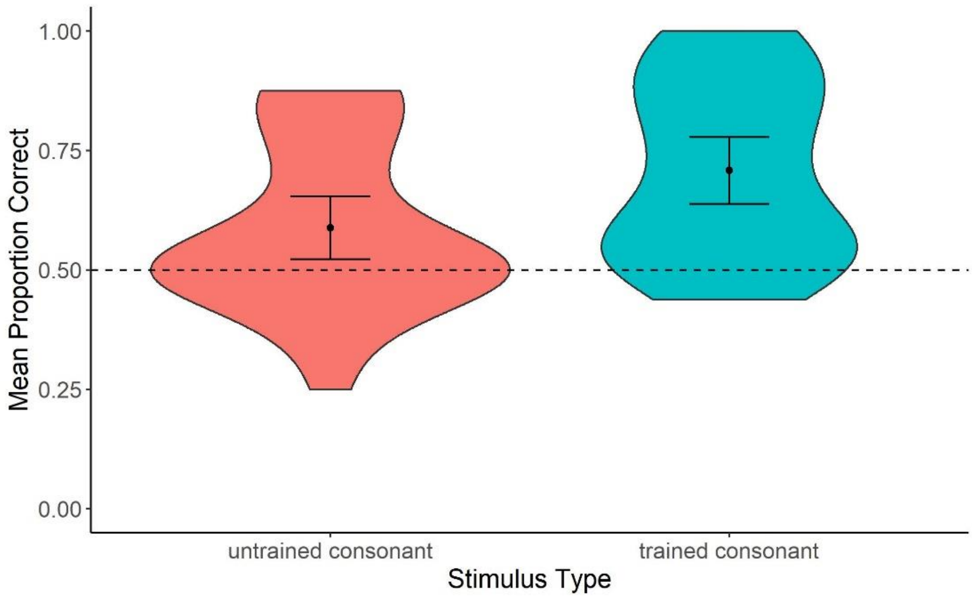


Figure 4.5: Adults' Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Legality Judgment Task in Experiment 2, by Type of Consonant (Trained and Untrained). The Dashed Line Represents Chance-level Performance (50%).

4.3.2.2.3 Comparison between trained and untrained consonants

We found conclusive evidence for H_1, $BF_{(0, 0.92)}$ = 245, *RR* [0.63, > 4.59] (effect of type of consonant: β = 0.73, *SE* = 0.20, *z* = 3.60 *p* < .001).

4.3.2.3 Comparison between children and adults

We showed that both children and adults pick up on the novel graphotactic positional constraints in semiartificial stimuli both when they are tested on novel items embedding trained and untrained "word"-medial consonant. We now turn to look at whether adults outperform children in learning these visual positional constraints.

4.3.2.3.1 Trained consonants

We found conclusive evidence for H_1, that is, adults (*M* = .71, *SD* = 0.18) outperformed children (*M* = .61, *SD* = 0.15), $BF_{(0,\ 0.42)}$ = 11.30, *RR* [0.14, > 4.59] (effect of age group: β = 0.69, *SE* = 0.27, *z* = 2.60 *p* < .01).

4.3.2.3.2 Untrained consonants

The evidence was inconclusive, $BF_{(0,\ 0.42)}$ = 0.91, *RR* [0, 0.93] (effect of age group: β = 0.17, *SE* = 0.22, *z* = 0.80 *p* = .42). Thus, we do not have substantial evidence that adults (*M* = .58, *SD* = 0.17) were better than children (*M* = .55, *SD* = 0.13) at accurately judging the legality of novel strings embedding untrained consonants (though also no evidence for the null).

4.3.2.4 Associations between learning and literacy performance

As in Experiment 1, we explored associations between learning performance and accuracy on standardized (WRAT-4 and TOWRE) reading and spelling performance. We used BF[18] analyses and a novel method of estimating the theory (outlined in Section 2.3.1: Method D), to quantify evidence for both H_1 (positive associations such that those who performed better in the learning tasks were also better readers/spellers) and the null (no relationship between statistical learning and literacy skills).

Results from our correlation analyses are presented in Table 4.1. For each participant, their raw scores for each standardized test were correlated with the mean average accuracy across all test items. To sum up the results, in the current semiartificial positional study, the evidence for the association between learning performance and literacy was inconclusive in all but one occasion where we demonstrated the null (no relationship between task performance and literacy): children's legality judgment performance and spelling ability.

In sum, when conclusive patterns of correlations emerged, these were evidence of no relationship between learning performance and literacy.

[18] See Table 2.3 in Chapter 2

Table 4.1: Correlations between accuracy in Experiment 2 and TOWRE word and nonword, and WRAT reading and spelling raw scores

Statistics	TOWRE word	TOWRE nonword	WRAT reading	WRAT spelling
BF [RR]	0.40 [0, 0.63]	0.52 [0, 0.87]	0.40 [0, 0.63]	**0.20**[a] [0.30, >4.59]
p	.79	.56	.78	.50
z_r (SE_{Zr})	0.05 (0.18)	0.10 (0.18)	0.05 (0.18)	-0.12 (0.24)

[a] substantial evidence for H_0

[b] substantial evidence for H_1

4.3.2.5 Awareness Questionnaire

Participants' responses were coded in a similar way as in Experiment 1: If any response described patterns that described the features of a symbol and its correct position, e.g., "the reversed e with three lines could only appear at the beginning of words", that response was coded as *explicit*. See Section 3.2.2.8 for the coding of all other responses.

12 adults (44%) responded "yes" to the question of whether they had awareness of any patterns, 8 (30% said "I don't know" and 7 (26%) said "no". Of those who said "yes", none of the answers to the follow-up question matched the description of a correct pattern. Their answers included: "If all the letters seemed familiar to me, I assumed that I've seen them before", "I tried to identify the words that exist in Zoribs' language by trying to remember which shapes I had seen during the color matching game and which I hadn't. I also tried to think of 'English alphabet letters' I could recall seeing in the game and which I couldn't", "I just looked for the English letter being present in the middle of two swirly Zorib style letters". Many participants reported that they guessed or the choices were subconscious, even though they responded "yes" to the question whether they were aware of patterns.

13 of children's responses (37%) included details regarding what they thought the patterns were, such as "Some had three lines, trying to remember the ones that were in the game" or "*n*, *j* and *f* were Zorib words"; while all others (63%) said "I don't know" or "it felt right". No adults and child participants were excluded on this basis.

4.3.3 Discussion

Learning effects seen in previous studies (Samara & Caravolas, 2014; Samara et al., 2019) have been interpreted as graphotactic in nature. However, it is unclear if this is underpinned by learning of correlated phonotactics. To address this confound, the current study introduced positional graphotactic constraints in semiartificial stimuli using unpronounceable symbols (from J. Taylor et al., 2011) at the edges of three-character strings, with English consonants in the medial position. Positional learning was induced by introducing constraints on the allowable position of symbols (e.g., in one list, set 1 symbols could only appear in word-initial positions; and set 2 symbols could only appear in word-final positions). As in Experiment 1, following incidental exposure to the pattern-embedding stimuli, children's and adults' learning was tested with legality judgments about conforming/nonconforming strings which were always novel—that is, the string had not been seen during exposure—and thus tested generalization. However, two types of generalization test items were included: two thirds of the test items contained a trained word-medial English consonant (i.e., seen at exposure although not with the combination of onset and coda symbols used in the test item) and one third contained an untrained English consonants (i.e., not seen in exposure at all). Further to the experimental learning task, child participants were administered measures of literacy (reading and spelling task), in order to explore associations between incidental/explicit learning and literacy performance. The key result was in line with Nigro et al. (2016) and Samara and Caravolas (2014), who demonstrated that children (6–7-year-olds and 8–9-year-olds, respectively) learned the positional graphotactic patterns following brief incidental exposure to novel pattern-embedding stimuli. In our study, both children and adults learned the positional constraints from semiartificial stimuli both when the word-medial consonants were trained and untrained, with Bayes factor analyses revealing substantial support for H_1 for both types of stimuli in both age groups. While adults were better at learning the semiartificial positional patterns when the consonants were trained, compared to untrained ones, for children this difference was inconclusive. We tested whether adults were better than children for each type of test item: This was the case when the medial consonants were trained, but the evidence was inconclusive where the medial consonant was untrained. Finally, we found no evidence of a relationship between incidental learning performance and participants' reading/spelling ability for any of our TOWRE and WRAT literacy measures, with evidence for the null in one case. I discuss each of these findings below.

We provide a strong demonstration that purely visual patterns on letter positions that exist in natural languages can be learned via statistical learning mechanisms by developing readers. The semiartificial strings embedding the positional patterns were designed to be unpronounceable and thus, the symbols did not provide phonological cues. Unlike previous studies that used pronounceable letters (Samara & Caravolas, 2014; Samara et al., 2019 and Experiment 1), we can determine that participants were sensitive to the graphotactic and not phonotactic constraints. This study goes further than Nigro et al. (2016) in that it shows learning using stimuli which are much more letter-like than the symbols used in their study, and, therefore, it is clearer that the learning in the current study relates to learning over orthographic scripts. Moreover, the fact that the participants generalize over English consonants, further demonstrates that their learning is orthographic in nature.

One potential limitation that requires consideration is the fact that we cannot rule out that the children (and to some extent adults) might have attempted to associate symbols with sounds, either as an association with a similarity with Latin alphabet letters or by using their imagination. This may have been possible despite our efforts to create unpronounceable "words". While we deliberately used symbols that are more letter-like in order to mimic naturalistic written language features, such design could introduce more similarities between our symbols and familiar letters, as demonstrated by Vidal et al. (2017), who pointed that despite the variation between scripts, some features such as configurations of stokes are shared.

Moreover, in the current study, the pronounceability of the stimuli have been exacerbated by the use of real English consonants, which of course are pronounceable. If participants do form such covert phonological representations of the string, this could have led to phonotactic learning, which we were aiming to avoid. However, it is unlikely that this happened in the current study: Given the brief exposure, it would have been extremely difficult to come up with a pronunciation for each novel symbol, especially under cover task conditions with no instruction regarding the presence of patterns. However, in the end, it is impossible to guarantee that stimuli are completely unpronounceable, whist ensuring they are sufficiently "orthographic-like" to be relevant to graphotactic learning.

An important question in whether the learning in these experiments was truly implicit, we tentatively suggest that this is the case, for two reasons. Firstly, the exposure phase involved a cover task (color detection) that was intended to distract the participants from explicitly searching for regularities, and they were not explicitly instructed regarding the presence of patterns. Secondly, the awareness questionnaire suggested that neither child nor adult participants were aware of the existence of patterns in the stimuli, or their reports did not match the actual rules. We acknowledge, however, that verbal awareness reports from children or written reports from adults may not be reliable measures of actual awareness (Brewer, 1974; Dienes, Broadbent, & Berry, 1991), and some explicit knowledge might either not be articulated or miscommunicated. While participants, especially adults, could have verbalized some patters as "the turned C with two lines only begins words", none have explicitly reported such strategy. If, however, they did verbalize the patterns but did not (were not able to) report this, they could have, at most, been able to partially achieve this; that is, apply this rule to one symbol occurring in one position. We assume that children would not be able to apply such covert rules. In sum, our results strongly indicate that, even when the phonology is not provided, available or helpful, children and adults can extract positional graphotactic constraints through implicit statistical learning mechanisms and generalize them onto new semiartificial strings. Developing spellers as young as 7 years old are sensitive to these visual constraints, despite not yet being explicitly aware of the existence of more complex patterns beyond letter-sound correspondences, such as graphotactics. This is in line with experimental evidence from studies that show that children are sensitive to graphotactic patterns governing doublet use in natural language stimuli (Cassar & Treiman, 1997; Lehtonen & Bryant, 2005).

As noted above, a key finding in this chapter was that both children and adults showed generalization to strings using English consonants which did not occur at all at exposure. This suggests that participants view these semiartificial strings as orthographic, so that they allow generalization within English orthography. Including these strings also demonstrates that learning was truly cued by the position of symbols and not their relationship with the adjacent consonant. For example, in the trained consonant strings, participants were able to take advantage of the whole bigram (e.g., "*Symbol 1 is followed by consonant 'c'*"; Symbol 1-c) in learning the graphotactic constraints. In the untrained consonant strings, however, participants could only rely on information regarding the position of the symbols alone (e.g., "*Symbol 1 can only occur as an onset*").

The untrained consonant introduced a chunk novelty, that is, bigrams that did not appear during the training phase (Meulemans & Van Der Linden, 1997), allowing assessment of sensitivity beyond context. For example, when learning that *ck* cannot begin words, children do not need to consider the vowel that follows (e.g., **ckat*, **ckut*). We found evidence for a sensitivity to the positional constrains even when the context was completely unfamiliar: Both children and adults discriminated between legal and illegal test strings with untrained word-medial consonants, with significantly better than chance accuracy.

To investigate the extent to which participants' performance with trained consonants was better than with untrained consonants, we ran a comparison between these conditions. Note that, although other studies have compared fully trained versus novel test items—e.g., Onishi et al. (2002) and Samara and Caravolas (2014), who found no significantly differences; Nigro et al. (2016) who found that transfer of knowledge from trained items to untrained ones was challenging at least for dyslexic children—none have compared different types of generalization items, as we do here, so the predictions are unclear. In fact, we found substantial evidence for a difference in adult participants, that is, they were better at discriminating legal strings when the embedded consonant was trained, however, the evidence was ambiguous for children. We also found that adults were better than children with trained consonant items, but the evidence was ambiguous for the items with the untrained consonants. Because we cannot interpret ambiguous data, we can only conclude that, at least for adults, the items containing the consonant seen at exposure (trained) are easier to learn. This may be because they find it hard to accept legal strings which use a consonant they are not sure they have seen before. Some adults reported that, when presented with stimuli containing untrained consonants, they tended to reject them as not being in the language they were trained on, purely because they have not seen them before. It is also possible that adults' performance in the test items is boosted by learning of knowledge of bigrams, that is, they show both positional and contextual learning. From our data, we cannot differentiate between these possibilities, or determine whether they are relevant for children, given the ambiguous findings. A critical finding in this study, however, was that we have clearly established that both children and adults generalize with both types of stimuli (trained and untrained consonants), that is, their performance was above chance with each type of stimuli.

Finally, as in Experiment 1, we did not find evidence for a positive association between statistical learning processes—as measured in our legality judgment task—and literacy performance. While most of the correlations were inconclusive (i.e., are neither evidence for H_1 or for H_0), where data was conclusive, it was against H_1, that is, it showed that children's variations in learning skill (as measured by legality judgment) were unrelated to variations in reading and spelling ability (see Section 1.7 in Literature Review and Section 7.3 in General Discussion for further comments).

In conclusion, we found clear evidence of learning of positional graphotactic constraints when the phonology was not available and not provided. Children and adults learned the nonphonological novel graphotactic constraints and none of the participants reported any awareness of the patterns. The current study goes beyond previous research by demonstrating purely graphotactic learning that cannot be explained by phonotactic sensitivity. One potential limitation, however, is that the patterns in this experiment were positional, and thus relatively simple compared to the contextual ones that were shown to be learned by 7-year-olds (Samara & Caravolas, 2014; and Experiment 1). While the positional constraints we use do mirror naturalistic evidence that children's invented spellings often conform to positional constraints (e.g., "words begin with *ck*") (Treiman, 1993), there remains the important question of whether children are also able to learn context-based graphotactic patterns when these have no phonotactic counterpart. We could potentially explore contextual learning using the current, semi-artificial paradigm. However, the mean performance here was relatively low (59% for children and 71% for adults, with trained consonants), and Samara and Caravolas (2014) did not find stronger learning with positional constraints (64% for children and 69% for adults). Therefore, it is likely that, in order to find learning effects of contextual constraints with these stimuli, we would need longer training, as has been done in spoken language experiments using artificial or semi-artificial language (e.g., Hudson Kam & Newport, 2009; Samara, Smith, Brown, & Wonnacott, 2017). Furthermore, the fact that we used only eight symbols to create our stimuli may have limited the real-life validity of our results: Chetail (2017) argued that, in order to create more naturalistic experimental stimuli, it is best to have closer to the number found in real written languages. With these limitations in mind, in Chapter 5 we moved to use a different paradigm and assess graphotactic sensitivity to context-based conditions using rules governing the use of single and double consonants. The rules governing the use of doublet versus singlet consonants are a good example of purely visual patterns in that they can give rise to homophones (e.g., *d* is pronounced the same as *dd*). The work of Treiman and colleagues (Hayes et al., 2006; Treiman & Boland, 2017; Treiman & Wolter, 2018) has demonstrated that children are sensitive to visual contextual constraints on double consonants in their own orthography.

5 Implicit and Explicit Learning of Graphotactic Contextual Patterns with no phonological counterpart

5.1 Introduction

As argued in Section 1.3, patterns found in the visually presented words of written languages can be purely visual (graphotactic) in nature, with no phonological counterpart. I also demonstrated that double consonants (doublets) are particularly relevant examples of graphotactic constraints because, generally, differentiating singlets and doublets cannot be done on a phonological basis. Because this chapter introduces the use of doublets to create homophone[19] stimuli (i.e., words that have the same pronunciation but different spellings), I discuss the situations in which they do and do not differ in the acoustic properties compared to their singlet form, and how they are used in psycholinguistic research.

5.1.1 Homophones in psycholinguistic research

Different spellings of homophones often are not arbitrary: The sound-spelling divergence in most homophones has a phonological (70%[20]) or morphological (23%) justification. This is not surprising, considering that, in most orthographies, there are many more spelling options for a phoneme than there are phonemic options for a grapheme, as noted in Section 1.2.2 (Treiman, 2018). In English, for example, the word *beat* has only one possible pronunciation, while /biːt/ can be spelled as *beet*, *beete*, *beat*, *beate*, *biet*, etc. (Crystal, 1995). For this reason, English spellings create homophones such as *cent*/*scent*; few/*phew*; *dew*/*due*; *gilt*/*guilt*; *great*/*grate*; *idol*/*idle*. Another phonological explanation is the

[19] This thesis does not focus on the other terms used in describing homophony, and which are used interchangeably, such as *homonyms*, i.e., words that have the same pronunciation and same spelling but different meaning (e.g., *bark*, *bear*); or *homographs*, i.e., words that have the same spelling but different pronunciation and meaning (e.g., wind, pronounced as /wɪnd/, meaning air in motion, or /waɪnd/ as in wind down, meaning to draw gradually to an end).

[20] Data from an analysis of a corpus of 750 homophone sets (Hwayed, 2010)

elision (i.e., orthographic presence but phonological absence) of certain sounds (e.g., *rite*/*write*, *dam*/*damn*, *no*/*know*, *our*/*hour*, *by*/*bye*). Other influences are morphological, such as suffix formation (e.g., *tide*/*tied*, *guest*/*guessed*, *red*/*read*, *mince*/*mints*) and contraction (e.g., *it's*/*its*; *he'd*/*heed*). The spelling differences can also reflect both phonology and morphology (e.g., the sound /k/ represented as *c* or *ck*, in addition to suffix formation in: *pact*/*packed*). As seen in the examples above, homophones are realized through transposals of letters in words of the same length (e.g., *tide*/*tied*), a change in one letter (e.g., *groan*/*grown*) or a change in the length of the word (e.g., *threw*/*through*) (Thorpe, 2017). While the classification above shows systematic linguistic factors that make the homophones less arbitrary, some patterns such as double consonants remain difficult to explain.

In some cases, doublets do not preserve the same pronunciation in both singlet and doublet form, as they do in some languages (e.g., in English, *n* in *in* and *nn* in *inn* represents the single sound /n/). In some languages, doublets stand for long consonants (a process known as "gemination"), while in others they represent a completely different sound. In Finnish, for example, the word *illan* is the genitive form of *ilta* (meaning evening), where the doublet *ll* is formed by the single *t* assimilating to the preceding *l,* resulting in a phonetically longer consonant. In Spanish, the doublet *ll* is pronounced as /ɟʝa/ in *llamar* (/ɟʝaˈmaɾ/, meaning 'to call') while the single *l* is pronounced as /l/ in *leche* (ˈle.tʃe, meaning 'milk'). Often, however, grapheme doubling does not have an obvious phonological explanation (Hwayed, 2010) but is used to differentiate words (e.g., *ad*/*add*; *in*/*inn*; *finish*/*Finnish*; *canon*/*cannon*), albeit in a relatively low proportion (7% in the English corpus). For example, they may reflect a historical sound difference that is now lost: In Old English, double consonant letters represented a long sound (Lass, 1994), and this acoustic feature was lost during the later Middle English period, while the doublet spelling was retained. In a similar way, Dutch consonants double in closed syllables due to spelling reform and thus, without sensitivity to the phonological context (e.g., *sterren*, meaning stars) (Kemper et al., 2012; Nunn, 1998).

The apparently unconditioned variation in spelling doublets described above causes difficulties for children learning to spell. However, one potential source of regularity can come from graphotatctic regularities, that is, as explained in Section 1.3, the usage of single versus double forms might be conditioned/predictable on the basis of aspects of the orthography, even when they are not phonologically or morphologically conditioned (e.g., doublets often occur after single rather than double-letter vowel spelling, as in *bedding*) (Hayes et al., 2006). As reviewed in Section 1.3.4 and 1.6, there is evidence that children are sensitive to the statistical influences that regulate consonants doubling, such as frequency, position and context. This sensitivity was seen in preliterate children's naturalistic spellings (e.g., Treiman, 1993), and was shown in experimental designs that tested sensitivity to patterns in participants' own language (e.g., Cassar & Treiman, 1997; Pacton et al., 2001), or in a novel, experimenter designed, artificial language (Nigro et al., 2015; Samara & Caravolas, 2014; Samara et al., 2019). A goal of this thesis was to explore experimentally whether children as young as 7-year-olds can learn these types of constraints. In the previous two empirical chapters, we have seen that children can learn phono-graphotactic contextual constraints (Chapter 3) and positional constraints that are fully graphotactic, since they were embedded in a semiartificial language (Chapter 4). However, we have not yet looked at learning of graphotactic contextual patterns. The main goal of the current chapter is to explore whether the *more complex* context-based graphotactic constraints with no phonotactic counterpart can be learned by children (and adults) under brief incidental conditions. This was achieved by designing novel homophonic stimuli that retain the familiarity of letters but remove the phonological confounds.

Artificially created homophones (pseudohomophones) have been extensively used in psycholinguistic research, particularly in the reading literature, and they have helped inform models of word recognition (Coltheart, Davelaar, Jonasson, & Besner, 1977; Rubenstein, Lewis, & Rubenstein, 1971; Seidenberg, Petersen, MacDonald, & Plaut, 1996; Yates, Locker, & Simpson, 2003; Ziegler, Jacobs, & Klüppel, 2001) and can help assess the effects of sublexical orthographic processing more directly. They are nonwords that sound like real words but differ in spelling (e.g., *brane* for *brain*, or *fownd* for *found*). While in the reading literature the pseudohomophones share a phonological or semantic code with lexical items, thus helping assess knowledge of patterns that exist in participants' own language, they can also be completely artificial: A target homophone would share the properties of a newly trained item. This resembles second language learning, when both phonology and orthography are learned together, making prior knowledge of the phonological property of words redundant. Completely novel homophonic pairs have been used to test the orthographic learning in studies exploring the role of self-teaching theory in early reading acquisition: In Cunningham, Perry, Stanovich, and Share (2002), for example, 7–8-year-olds were asked to read aloud a short story that embedded one of the homophonic pseudowords from a pair (e.g., *yait*), and their orthographic learning was assessed using three tasks: (i) an orthographic choice task (with four options: the target homophone, the homophonic alternative (e.g., *yate*), as well as letter substitution (e.g., *yoit*) and letter transposition (e.g., *yiat*) alternatives); (ii) a spelling task; and (iii) a homophone naming task. Three days following exposure to the novel word form, children were faster and more accurate at identifying, reproducing (i.e., spelling), and naming the target homophones, compared to the alternatives. An interesting outcome in this study came from a series of hierarchical regression analyses: While the self-teaching theory affirms that target decoding (i.e., the accuracy at decoding the target word during reading aloud) predicts the entire effect in such paradigm (i.e., orthographic learning relies solely on accurate print-to-sound decoding), Cunningham et al. (2002) found that, when decoding was partialled out, prior orthographic knowledge (measured by the orthographic choice of real homophones) predicted a significant amount of variance. Neither general ability (e.g., picture vocabulary, digit span, and word attack tests), nor rapid automatized naming (RAN), reached significance levels in hierarchical analyses. This was surprising, considering that RAN, for example, has been shown to be particularly powerful predictor in reading development (Caravolas, Lervåg, Mikulajová, Defior, & Hulme, 2019). While these findings show that the development of orthographic knowledge is not “entirely parasitic on decoding ability” (Cunningham et al., 2002, p. 196), and that prior successful orthographic skill predicts success with novel orthographic input, such paradigm does not specify what are the processes underlying this success. In the current chapter, a goal was to provide further evidence that children and adults can learn different homophone-based spellings based purely on orthography, under incidental learning conditions. We aimed to find whether statistical learning processes underlie learning of the more difficult, *context-based* graphotactic patterns. We achieved this in a language learning design by using artificial homophonic pairs created by adding consonants in their single and double form at the end of monosyllabic (CVC/C) strings, where the final consonant doubles depending on the identity of the preceding vowel.

Another way that children master homophone spellings is via explicit teaching in the classroom. In the current experiment, therefore, we also included an explicit condition, which we compare to incidental learning. I review relevant literature in the following section.

5.1.2 Explicit learning in spelling

The literature reviewed in Chapter 1 and evidence presented in Chapter 3 and 4, suggest that aspects of spelling can be acquired without deliberate effort to learn and that knowledge acquired in this manner may be unavailable to retrospection (or "awareness", by some definitions of implicit learning; Frensch & Rünger, 2003). Spelling patterns and rules are, however, also taught explicitly in schools (e.g., "*i* before *e*, except after *c*"; "change *y* to *i* when adding suffix endings"; "*z*, never *s,* spells /z/ at the beginning of a base word") and learning by this process conceivably captures some of the advanced knowledge exhibited by proficient readers and spellers. While explicit instruction does not explain all aspects of how children learn to spell (see Section 1.8), it is important to understand how it contributes to proficiency in spelling, over and above implicit learning from incidental exposure to print. Questions regarding the differential effectiveness of instructional approaches are both theoretically important and practically relevant for spelling instruction. To date, little work has directly investigated whether learning of the same patterns is more effective under explicit than incidental conditions, or whether patterns that are hard to articulate, thus, not explicitly taught as part of the curriculum, are better learned incidentally.

Lab-based work carried out under the rubric of "implicit learning" research provides mixed early insights into these questions. In early artificial grammar learning work[21] (Reber, 1967) (experiments whereby participants perform a surprise grammaticality test following implicit exposure to stimuli generated from a finite-state grammar), warning participants of the rule-embedding nature of the

---

[21] See the introduction to Section 1.5 for a discussion about the difference and similarities between the two literatures of implicit and statistical learning; and Section 1.5.3 for examples of studies using artificial grammar learning (AGL) paradigm.

stimuli was shown to improve learning (Gebauer & Mackintosh, 2007; Reber, Kassin, Lewis, & Cantor, 1980), impair learning (e.g., Brooks et al., 1978; Reber, 1976), or make little difference in terms of learning effects (Dienes et al., 1991; Dulany, Carlson, & Dewey, 1984). Reber (e.g., Reber, Walkenfeld, & Hernstadt, 1991) had notably argued that what is learned implicitly often appears to exceed what can be learned explicitly with deliberate effort: By this view, searching for complex grammatical rules and linguistic patterns impedes learning.

A few studies from the statistical learning tradition have also compared the degree of learning exhibited by participants who were intentionally searching for patterns relative to participants trained under standard (incidental) conditions, but these have also yielded conflicting results. Arciuli, von Koss Torkildsen, Stevens, and Simpson (2014) for example, found nonsignificant differences in adults' incidental and explicit visual statistical learning ability under conditions of short stimulus presentation, whereas Kachergis, Yu, and Shiffrin (2010) demonstrated an explicit condition advantage in a within-subjects design: Participants were better when they were explicitly instructed to count co-occurrence statistics obtained in a cross-situational learning task relative to when they performed the task incidentally.

Graham and Santangelo (2014) analyzed evidence from 53 studies that used various explicit spelling instruction methods (e.g., teaching students specific word spellings, strategy building activities, systematic word-study activities) and conclusively demonstrated literacy-related gains that were long lasting and held across age and literacy ability groups. While this finding is not entirely unexpected, it contradicts views put forward by proponents of the input hypothesis (e.g., Krashen, 1989), according to which implicit exposure (e.g., used in free-reading school programs) is efficient for spelling and vocabulary acquisition.

Turning to orthographic pattern learning, results have been also mixed (Bosman et al., 2006; Butyniec-Thomas & Woloshyn, 1997; de Bree, Geelhoed, & van den Boer, 2018; Kemper et al., 2012; Nunes et al., 2003). Sobaco et al. (2015) directly compared their participants' ability to learn isolated graphotactic patterns that are either legitimate or illegitimate in French spelling, from nonwords, under implicit conditions (whereby the instructions were to simply read the words aloud) and explicit conditions (whereby the nonword spellings were to be memorized) and found better learning recall across spellings in the latter group of participants (75.7% relatively to 57.6 % in the implicit condition). An advantage for explicit learning was also found in a study with 7 year old Dutch children (Kemper et al., 2012), which compared implicit and explicit learning of a vowel-degemination rule (long vowels spelled with double letters in a final closed syllable lose one of their vowels when the noun is made plural), controlling for previous knowledge with a pre-test. Children were either explicitly taught the rule (followed by practice) or were shown one singular-plural example and subsequently produced plurals from singulars (implicit condition). Although no significant difference between implicit and explicit instruction was found for trained items, only explicitly taught children's knowledge showed generalization over transfer items. Bosman et al. (2006) also reported an explicit learning advantage: Dutch 9-year-olds were explicitly taught to read aloud inconsistent loan words such as *computer* and *bureau* phonetically to cue the unpredictable spellings. They subsequently tested children' ability to spell these words and compared their performance against that of children who were taught the conventional pronunciations. It was found that they not only benefitted from this strategy but were also better at generalizing on novel, untrained items. These results suggest that explicit instruction of strategies when reading (letter-to-sound) facilitate spelling (sound-to-letter) of inconsistent orthography, at least in the short term. Using comparable incidental and explicit tasks, Rastle et al. (2021) showed that, when explicit instruction over symbol-to-sound and symbol-to-meaning mappings in two artificial systems was given, the vast majority of English-speaking adult participants generalized well (i.e., could accurately read aloud untrained words and draw spelling-to-meaning mappings within untrained words). On the other hand, only 21% of their participants reached the same high levels of performance in the discovery (i.e., incidental) condition.

de Bree et al. (2018) on the other hand, found an implicit learning advantage: They studied learning of the same Dutch vowel-degemination rule among Dutch-second graders (7–8-year-olds, who were explicitly taught the spelling rule in school) relative to first graders (6–7-year-olds, not yet taught the rule) using a spelling to dictation task that featured words and nonwords that varied in terms of morphological and semantic properties as well as orthographic familiarity. Children who had received explicit instruction of the rule in the classroom (and were also one year older), did not demonstrate an overall better performance than the younger ones who had not; in fact, older children were shown to exploit various untaught (implicit) spelling cues in choosing between single and double *ee* spellings.

While studies with typically developing children find, at best, modest benefits of explicit learning compared to implicit, studies with dyslexic adolescents (11–14-year-olds) have found strong evidence in favor of explicit and systematic instruction of literacy and literacy-related skills (Wanzek et al., 2006). Bhattacharya and Ehri (2004), for example, showed that teaching adolescents with dyslexia how to break down a multisyllabic word into its constituent syllables (an area of difficulty for poor spellers; Ehri & Saltmarsh, 1995) significantly enhanced their decoding and spelling ability for familiar, unfamiliar, and novel words. Poor readers receiving whole-word reading or no instruction did not improve in performance, taken to suggest that explicit instruction that targets directly poor skills is necessary for remediation in dyslexic individuals.

In sum, results of the few studies that have directly evaluated learning effects under comparable implicit and explicit conditions do not converge. In addition, evidence from Dutch (a consistent orthography) may not generalize to English; some of the studies use a naturalistic design that does not control for age differences in the implicit and explicitly taught groups of participants; and explicit instruction has been operationalized differently across studies. A comparison between strictly comparable incidental and explicit conditions for spelling patterns that are shown to be extracted from incidental reading exposure is overdue. One of the aims of this chapter, therefore, was to address the question of whether learning in our experiments occurs without awareness (incidental learning) and whether explicit awareness provided through instruction is beneficial.

As in the previous chapters, we also included tests tapping individual differences in children’s literacy skills, and here, we also looked at how these relate to learning spelling patterns under both implicit and explicit conditions in the laboratory (see Section 1.7 for a literature review).

5.1.3 The current study

In the current study, stimuli embedding contextual graphotactic patterns were created by using homophones spelled with single versus double letters, where spellings (e.g., *dd*, *d*) map to the same sound (/d/), thus, learning when letters double is a purely graphotactic effect. Two different types of constraints of differing levels of difficulty were tested in two experiments (with different groups of participants). In both, singlet/doublet usage was predicted by the identity of the preceding vowel: In Experiment 3, one medial vowel always predicted single consonants and one always predicted doublets, whereas in Experiment 4, each of the two possible vowels predicted some singlets/doublets.

Both experiments were run with both child and adult participants. Across all groups and experiments, participants were exposed, under the same task procedures as in Experiment 1 (Section 3.2.1.4), to the relevant patterns in two brief sessions, and learning generalizations were tested in two tasks (also the same as in Experiment 1): a *fill-in-the-blanks test*, whereby participants were asked to construct conforming generalization nonwords by choosing one of the possible vowels to "fill-in" a consonantal frame (C_C/C); and a *legality judgment*, which required yes/no answers to unseen stimuli that either conformed to or violated the learned patterns.

We also ran an additional experiment (Experiment 5) with children, identical to Experiment 3, with an additional explicit instruction that preceded the incidental exposure in both sessions: The children were told what the rule underlying the patterns in Zorib language were, before starting the exposure game.

Finally, as in Experiment 1 and 2, we explored associations between graphotactic learning ability and literacy performance by administering standardized tests of English word reading and spelling ability to all children (but not adults[22]).

For all participants, as in previous experiments, we also administered a post-experiment awareness questionnaire: For children, the questionnaire was verbal, while adults were required to write down their answers.

---

[22] Adults were tested online and due to copyright and data protection, standardized tests could not be administered.

5.2 Experiment 3

5.2.1 Method

5.2.1.1 Participants

The sample size was estimated using the procedure outlined in Section 2.6 and optional stopping was used, as specified in our optional stopping plan. This plan was pre-registered and can be found at https://osf.io/mn254. For child participants, we looked at the data after 25 (as per our optional stopping plan) and, because we did not find substantial evidence for above-chance learning, we recruited a further sample of 10 participants. For adults, because we expected a drop-out rate due to the two-session study and no possibility to guarantee that they return for the final (crucial) session, we recruited in excess of the 25 required for a first look at the data. Because we found substantial evidence for above-chance learning, we stopped recruiting after the first sample.

Using this approach, thirty-five typically developing Year 2 children (19 female, 16 male; mean age = 6.6 years, *SD* = 0.31) took part in Experiment 3. They were all recruited using an opt-out procedure from a primary school in London, had no known language, hearing or vision impairments and no history of learning difficulties. All children were monolingual English speakers and had received the same amount of formal literacy tuition (2 years). Children were rewarded with stickers and a certificate. As in previous work (Samara et al., 2019), the mean reading and spelling performance in our sample was above average (mean reading = 123.4, *SD* = 9.77; mean spelling = 121.4, *SD* = 12.8), which is relatively typical in experimental studies with child participants.

Twenty-nine adults (18 female, 11 male; mean age = 31.5 years, *SD* = 8.87) also participated in Experiment 3. They were recruited via Prolific (https://www.prolific.co/) and were tested online. They reported being monolingual native speakers of English with no language, hearing, vision impairments or any learning difficulties. They all provided informed consent and were paid for their participation.

5.2.1.2 Materials

Graphotactic learning task. We manipulated the joint probability of middle vowels and word-final consonants in monosyllabic words to induce a purely graphotactic constraint: Word-ending consonants always doubled in one context (i.e., following one of two possible vowels) and never doubled in another context (i.e., following the other possible vowel). Note that learning this constraint does not involve phonotactic sensitivity, in that word-final singlets and doublets (e.g., *d*, *dd*) map onto the same sound (/d/). Thus, vowel context is only predictive of word-ending letters, not sounds.

Stimuli were 64 pronounceable monosyllabic letter strings shown in Appendix F. They were created by combining one of four word-initial consonants (C_1: *d*, *g*, *m*, *r*), one of two vowels (Vs: *e*, *u*) and one of four consonants (codas), either single or doubled (C_2: *f*, *ff*, *l*, *ll*, *s*, *ss*, *t*, *tt*). Six of the stimuli were real English words (*mess*, *dull*, *get*, *gull*, *gut, met*) but, as in Experiment 1, these were spread equally across types of test items, which controls for any bias they might bring to performance. The 64 stimuli were arranged into four lists, two of which conformed to– and two of which violated a novel purely graphotactic constraint. Nonwords from the pattern-conforming lists served as exposure and legal unseen test items, and nonwords from one of the two pattern violating lists served as illegal test items. Importantly, list assignment was counterbalanced between participants such that, for half of the participants word-middle *u* predicted doublets (i.e., *guff*, at test, was illegal) and for the remaining half, the same vowel predicted singlets (i.e., *guff*, at test, was legal). This counterbalancing mitigates the concern that our effects reflect children's sensitivity to English statistical patterns.

Exposure items (n = 16) and legal unseen items (n = 16) conformed to the following graphotactic rule: One vowel was only followed by single consonants (e.g., in one counterbalanced list, *u* was always followed, with equal (.25) probability by *f*, *l*, *s*, *t*) and the other vowel was only followed by double consonants (i.e., in the same counterbalanced list, *e* was only followed with equal (.25) probability by *ff*, *ll*, *ss*, *tt*) (Figure 5.1). There were two stimuli for each legal coda (e.g., *duf, muf* and *deff*, *meff*). No other statistics were predictive of legality: Word-beginning consonants (C_1s) co-occurred with both vowels (Vs) with .25 probability [e.g., $P(d, e) = P(d, u)$].

Illegal items (n = 16) violated the graphotactic rule: The vowels were followed by single/doublet consonants that were not permissible (i.e., had zero probability) during exposure.

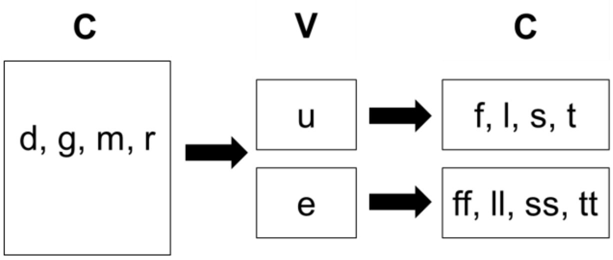


Figure 5.1: Schematic Representation of the Underlying Graphotactic Restrictions in List 1 of Experiment 3

5.2.1.3 Apparatus

For children, the experiment was run on a Windows 7 Enterprise PC with a 13.3-inch CRT color monitor. Visual stimulus presentation and millisecond accurate response registration was achieved using PsychoPy 1.82.01. Participants responded on a standard QWERTY keyboard.

As in Experiment 2, adult testing was carried out online using a link distributed via Prolific (https://www.prolific.co/). The experiment was designed on the Gorilla.sc platform (Anwyl-irvine et al., 2019) and was run on participants' own devices at home.

5.2.1.4 Procedure

Child testing was carried out individually in school and adult testing was carried out online. All child participants were seen in two 30-minute sessions over two consecutive days, except for six child participants who completed the sessions with a one-day gap. All adult participants completed each 15-minute session on two consecutive days.

Graphotactic learning task. As in the previous experiments (Experiment 1 and 2), at the beginning of the experiment, children were told that they were going to see written words from an alien language, called Zorib, and they would have to play games with the alien words. In session 1, the game (cover exposure task) was to detect consecutive word repetitions. In session 2, further (covert) exposure was given, followed by disclosure of the patterned nature of Zorib words (i.e., informing participants of the existence of patterns in the words seen). Subsequently, two new games (administered in fixed order) were (i) a task where children were asked to produce Zorib words by filling in a missing letter ("fill-in-the-blanks" test), and (ii) a game where they classified new words as possible/not possible Zorib words ("legality judgment" test). Procedures for each task are detailed below.

Exposure task. A total of 288 Zorib "words" (144 presented in 3 blocks in each session; 9 repetitions/string in each session) were shown in the context of a one-back cover task: Participants were instructed to look at each word and press a button when repetitions occurred consecutively (16 in each session). No other instructions or feedback was given. Stimuli were presented in black in the middle of a white background and remained there until a response was given. A response was allowed only after 350ms. A fixation point (black cross, presented for 500ms) followed the response, in the middle of the screen. Word order was manipulated as follows: Consecutive stimulus repetitions occurred once for each of the 16 strings and no more than 6 times in each block. All other stimuli appeared at random and no other doubles were allowed.

***Fill-in-the-blanks* task**. The fill-in-the-blanks task measured pattern sensitivity as reflected in participants' ability to choose the appropriate context (vowel) to create legal words in Zorib language. In each trial, participants saw (a) novel word frames consisting of a consonant, an underlined blank space for the missing middle vowel, and a word-final consonant (e.g., *r_ll*) and (b) below the frame, the two vowels used during exposure (*e*, *u*). The experimenter explained that their task was to drag the vowel and fill the blank to make a word that they thought possible in Zorib language. They were encouraged to use their gut feeling and were allowed to change their mind once they saw the word in full. Stimuli (*n* = 16) were presented one at a time in random order. Note that choosing correct responses made the 16 legal unseen frames used in the legality judgment test.

***Legality judgment* task.** In the legality judgment task, participants were presented with novel legal unseen (*n* = 16) and illegal (*n* = 16) strings in randomized order and were asked to decide if each of the words could/could not exist in Zorib language and press a corresponding button accordingly. If unsure, they were encouraged to trust their intuition or "gut feel". Each string was presented in the middle of the screen and remained until a response was given. A total of 32 items were presented in a single block.

Awareness questionnaire. As in all experiments in this thesis, a brief questionnaire was administered to assess whether participants were able to verbalize the graphotactic constraints governing Zorib words. Details of specific questions used are provided in 3.2.1.4.

Literacy measures. Child participants' reading and spelling skills were assessed using the two relevant subtests of the TOWRE-2 (Torgesen et al., 2012) and WRAT-IV (Wilkinson & Robertson, 2006, Green form).

5.2.2 Results

The design and hypotheses of this study were pre-registered at https://osf.io/mn254; however, due to an oversight, the estimates for comparisons against chance for adults were not specified in the pre-registered plan. In addition, the correlational analyses were exploratory and thus, the estimates for these were not included in the pre-registered plan—and these analyses, for all three experiments reported in this chapter, will be presented together at the end of the chapter (Section 5.4.1.4.2). For all three experiments reported in this chapter, the data is available at https://osf.io/tsxf3/; and data analyses are available at https://rpubs.com/DSingh/Graphotactics. The priors used to model the theory are detailed in Table 2.4.

5.2.2.1 Children

Figure 5.2 shows the mean proportion of children's correct responses in the fill-in-the-blanks task. Our data provided substantial evidence for the hypothesis, that children were better than chance (50%) at choosing the correct vowel, $BF_{(0,0.39)}$ = 8.72, *RR* [0.05, 1.40] (model intercept: β = 0.25, *SE* = 0.1, *z* = 2.47, *p* = .01).

Figure 5.2 also shows the mean proportion of children's correct legality judgments. There was evidence for above (50%) chance learning, $BF9_{(0,0.16)}$ = 21.5, *RR* [0.03, 2.21] (model intercept: β = 0.19, *SE* = 0.07, *z* = 2.80, *p* = .01), that is, children were better than chance at discriminating between legal and illegal items.

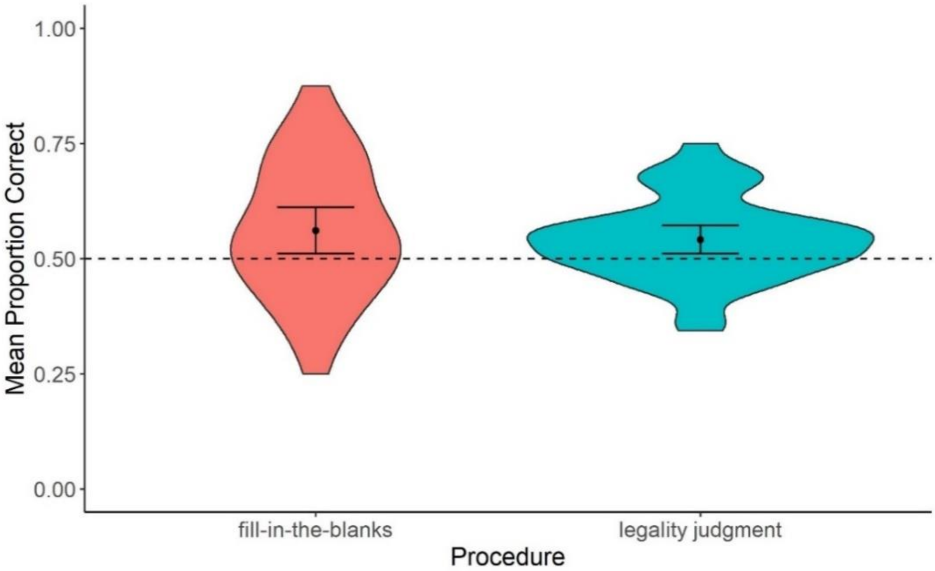


Figure 5.2: Children's Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 3. The Dashed Line Represents Chance-level Performance (50%).

5.2.2.2 Adults

Figure 5.3 shows the mean proportion of adults' correct responses in the fill-in-the-blanks and legality judgment task. For fill-in-the-blanks performance, we found that adults were above (50%) chance at creating permissible generalization stimuli, $BF_{(0,0.51)}$ = 12,455, *RR* [0.05, >4.59], (β = 1.14, *SE* = 0.24, *z* = 4.83, $p < .001$).

For legality judgments, we confirmed that adults were above (50%) chance at discriminating between legal and illegal items, $BF_{(0,0.32)}$ = 298.85, *RR*[0.06, > 4.59] (model intercept: β = 0.91, *SE* = 0.22, *z* = 4.06, $p < .001$).

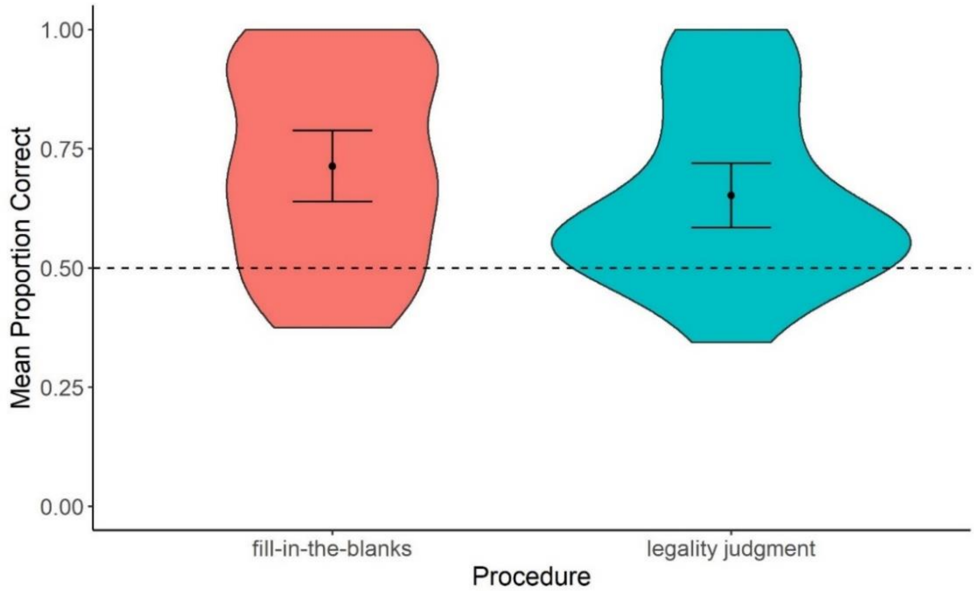


Figure 5.3: Adults' Mean Accuracy (Violin Plots with 95% Confidence Intervals) In the Fill-in-the-blanks and Legality Judgment Task in Experiment 3. The Dashed Line Represents Chance-level Performance (50%).

5.2.2.3 Awareness Questionnaire

Participants' responses were coded in a similar way as in Experiment 1, to identify if they could accurately describe the novel graphotactic pattern on doubling and vowel co-occurrence. See Section 3.2.2.8 for the coding of all other responses. None of the children reported awareness of the patterns embedded: 76% of children responded with "no" or "I don't know" to the question whether they were aware of patterns in Zorib language, while 24% provided uninformative replies, such as "Zorib language: doubles at the end". Thus, no further analyses were carried out. Eight adult participants, on the other hand, were classified as aware on the basis of accurate descriptions such as "*Doubled end letter were preceded by the letter e. Single end letter were preceded by the letter u*"; "*Four letter words could only have e and three letter words could have u*", "*Zorib language has an e when the last two letters are a double f*". To investigate whether task performance was driven by this subset of aware adults, we excluded them and repeated all analyses as above.

The mean correct performance of aware and unaware adults' in the fill-in-the-blanks and legality judgment task is shown in Figure 5.4. From a visual inspection, it is clear that the two groups performed qualitatively differently (there is no overlap in confidence intervals between groups), and that aware participants' accuracy was close to ceiling.

For fill-in-the-blanks performance, there was evidence of above-chance learning $BF_{(0,0.51)}$ = 83.64, *RR*[0.07, > 4.59], and so was for legality judgment performance, $BF_{(0,0.32)}$ =19.74, *RR* [0.09, 1.68] (model intercept: β = 0.31, *SE* = 0.11, *z* = 2.76, *p* = .01).

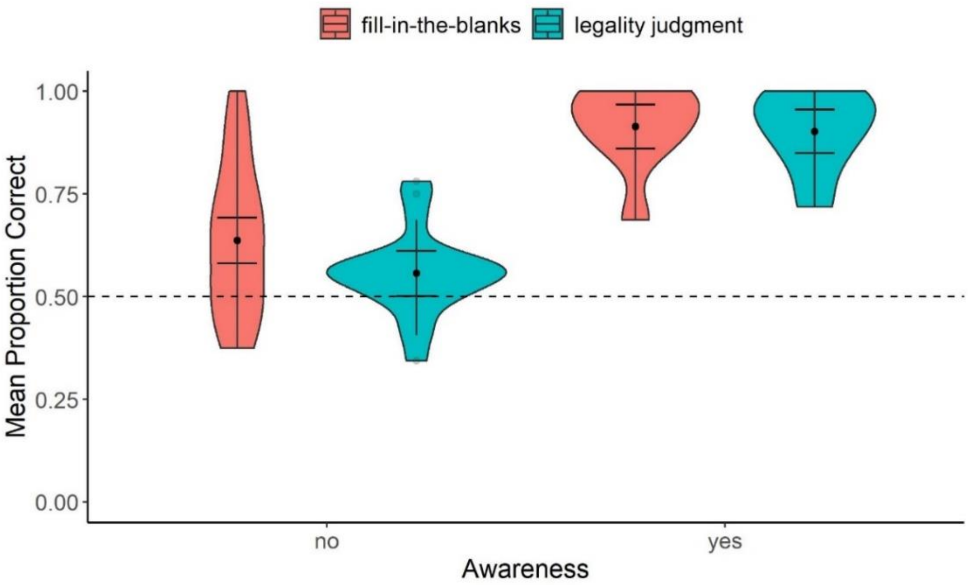


Figure 5.4: Adults' Mean Accuracy (Violin plots with 95% Confidence Intervals) by Awareness Status (Aware, Unaware Participants) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 3. The Dashed Line Represents Chance-level Performance (50%).

In sum, some adults demonstrated explicit awareness of the patterns embedded in the stimuli, and these participants did show markedly better performance in both tests of performance. Nevertheless, even when these participants were removed from the analysis, the remaining unaware participants were above chance as a group in both tests.

5.2.3 Discussion

In Experiment 3, we investigated children and adults' ability to learn novel purely graphotactic constraints incidentally. We created stimuli ending with either single or double letters (homophones) and manipulated the context in which they occurred such that all consonants always doubled following one of the two vowels and never following the other. We tested generalizations over this pattern by asking participants to discriminate between novel legal and illegal items (as in previous work by Samara and colleagues: Samara & Caravolas, 2014; Samara et al., 2019), and in a more naturalistic production task: Participants were presented with generalization (unseen) word frames and were asked to fill in a missing vowel, choosing from two alternatives, in order to create permissible (generalization) words in Zorib language.

In both tasks, we found clear evidence of learning. Children and adults learned the nonphonological novel graphotactic constraints, which they used to judge and produce novel unseen test items, that is, performance cannot reflect the ability to memorize the 16 exposure items. None of the children reported any awareness of the patterns, but some adults (*n* = 8) did (i.e., they were able to describe them when prompted at the end of the experiment). However, repeating the analysis without them confirmed that above-chance performance in the task was not driven solely by aware adults.

In sum, Experiment 3 goes beyond previous research by demonstrating purely graphotactic learning that cannot be explained by phonotactic sensitivity. One potential limitation, though, is that the patterns in this experiment were relatively simple compared to those, for example, studied by Hayes and colleagues (e.g., Hayes et al., 2006). It is, therefore, unclear whether our learning demonstration scales up to the challenge of learning real orthographic patterns. The somewhat simplified nature of pattern in our study might also explain why almost a third of adults could clearly articulate them. We address this limitation in a second experiment whereby we investigate children's and adults' ability to learn novel purely graphotactic constraints that are more complex in that both vowels can be followed by singlets and doublets.

5.3 Experiment 4

5.3.1 Method

5.3.1.1 Participants

Using the same sampling plan as in Experiment 3, described in Section 2.6 and pre-registered at https://osf.io/kz26g, twenty-five typically developing Year 2 children (16 female; 9 male; mean age = 6.8 years, SD = 0.39) and 35 monolingual adults native speakers of English (15 female; 18 male; mean age = 31.7 years, *SD* = 9.08) took part in Experiment 4. Decisions regarding optional stopping, recruitment, consent, and compensation processes were as in

Experiment 3. The mean reading standard score for children was 120 (*SD* = 9.15) and the mean spelling standard score was 121 (*SD* = 15.3). All participants completed two experimental sessions over two consecutive days.

5.3.1.2 Materials and Procedure

We manipulated the joint probability of middle vowels and word final consonants within homophonic CVC/C strings, but, unlike in Experiment 3, both vowels were followed by double/single letters (Figure 5.5). Specifically, Vowel1 was followed (i) by two of the consonants as doublets (i.e., in one counterbalanced condition, *e* was always followed with equal (.50) probability by *ff*, *ll*, but not by *ss*, *tt* (ii) and by the other two consonants as singlets (i.e., in the same counterbalanced condition, *e* was always followed with equal (.50) probability by *s*, *t*, but not by *f*, *l*. The opposite held true for Vowel2 (Figure 5.5). No other statistics were predictive of legality: Word-beginning consonants (C_1s) co-occurred with both vowels (Vs) with .25 probability.

As in Experiment 3, the 64 pronounceable monosyllabic letter strings shown in Appendix G (28 nonwords; 4 English words: *met*, *gut*, *get*, *dull*) were arranged into four lists whose order was counterbalanced across participants.

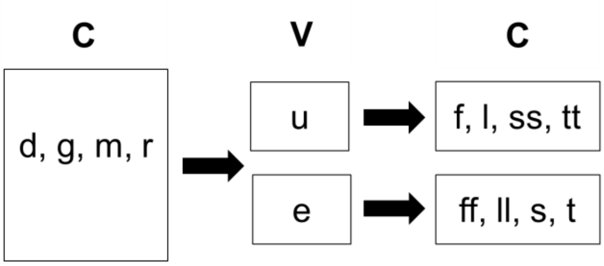


Figure 5.5: Schematic Representation of the Underlying Graphotactic Restrictions in List 1 of Experiment 4

All other aspects of the experimental task and our procedure were the same as in Experiment 3.

5.3.2 Results

The design and hypotheses of this study were pre-registered at [https://osf.io/kz26g.](https://osf.io/kz26g) The data and analyses, as well as the priors used to model the theory are under the same links listed in the results section of Experiment 3 (Section 5.2.2).

5.3.2.1 Children

Figure 5.6 shows the mean proportion of children's correct responses in the fill-in-the-blanks task. The BFs confirmed learning was above chance (i.e., children chose the expected vowel with better than chance (50%) accuracy), $BF_{(0,0.39)}$ = 3.13, *RR* [0.08, 0.42] (model intercept: β = 0.2, *SE* = 0.1, *z* = 2, *p* = .05).

Figure 5.6 also shows children's mean correct performance in the legality judgment task. The evidence supported H_1 (above chance discrimination), $BF_{(0,0.16)}$ = 5.20, *RR* [0.05, 0.43] (model intercept: β = 0.16, *SE* = 0.08, *z* = 2.12, *p* = .03).

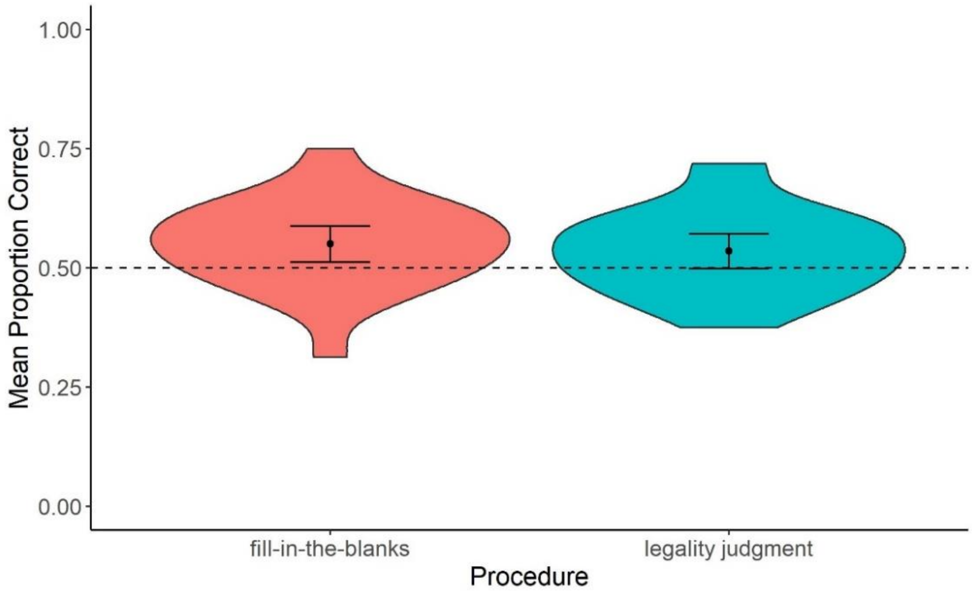


Figure 5.6: Children's Mean Accuracy (Violin Plots with 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 4. The Dashed Line Represents Chance-level Performance (50%).

5.3.2.2 Adults

Adults' fill-in-the-blanks and legality judgment task performance is shown in Figure 5.7. There was evidence of above chance learning (i.e., adults chose the correct vowel with above chance accuracy), $BF_{(0,0.51)}$ = 3.97, *RR* [0.05, 0.70] (model intercept: β = 0.2, *SE* = 0.09, *z* = 2.26, *p* = .02).

For the legality judgments, we again found evidence for H_1 (i.e., above chance discrimination between legal and illegal items), $BF_{(0,0.32)}$ = 5.18, *RR* [0.04, 0.62]) (model intercept: β = 0.16, *SE* = 0.07, *z* = 2.31, *p* = .02).

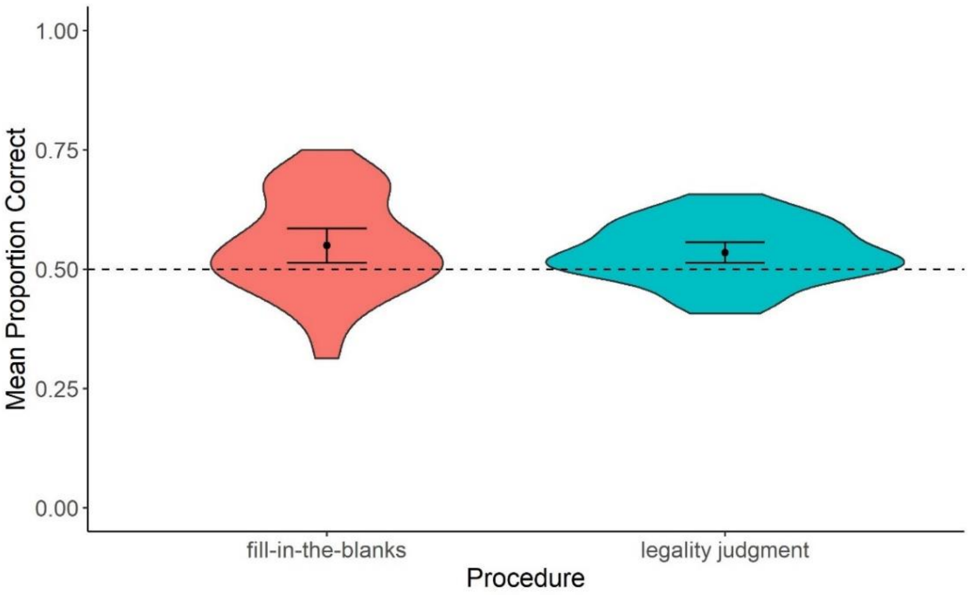


Figure 5.7: Adults' Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 4. The Dashed Line Represents Chance-level Performance (50%).

5.3.2.3 Awareness Questionnaire

None of the children responded "yes" when asked if they were aware of the patterns in Zorib language, 64% said "I don't know", while 20% said "I guessed" or "I remembered from the previous game", and 16% gave uninformative responses, such as "all *u*'s" or "*res* Zorib spelling, *met* not Zorib". 45% of adults replied "yes" to the awareness question, 29% replied "no" and 26% said "I don't know". Of those who said "yes", none provided further comments that were informative, e.g., "from game 1 I noticed Zorib language had plenty of "u's" therefore when playing the concluding games it felt right to select this letter over most" or "I noticed to some extent a pattern between the first letter and the one that came after it, e.g., r followed by e". In sum, none of the children, and, unlike in Experiment 3, none of the adults gave informative descriptions of the patterns that were introduced in the exposure phase of the experiment, thus, they were all included in the analyses reported below.

5.3.2.4 Comparison between Experiment 3 and Experiment 4

The results of Experiment 3 and 4 consistently suggest that children and adults pick up on novel graphotactic constraints from brief incidental exposure. Above-chance learning was seen when vowel identity predicted singlet/doublet occurrence (i.e., one vowel was consistently followed by doublets and the other was not) but also when each vowel predicted both word-ending singlets and word-ending doublets (half of each type). We hypothesized that learning the former pattern would be easier and sought to directly compare whether pattern complexity mediated the effects seen in Experiment 3 and 4. As detailed in Section 2.2.3, for between-experiment comparisons, we set a constraint on a likely maximum value from the data itself (Method C). Analyses were run both for children and adults (after removing those in Experiment 3 who were able to verbalize the patterns and were possibly engaging explicit learning processes).

5.3.2.4.1 Children

For fill-in-the-blanks task performance, our hypothesis could neither be accepted nor rejected, that is, the evidence for H_1 was inconclusive, $BF_{(0, 0.23)}$ = 0.63), *RR* [0, >4.59] (effect of experiment: β = 0.04, *SE* = 0.13, *z* = 0.33, *p* = .74).

For the legality judgment task, once again, data were insensitive, that is, we could not conclusively reject the H_1, $BF_{(0, 0.17)}$ = 0.59, *RR* [0, >4.59] (effect of experiment: β = 0.02, *SE* = 0.10, *z* = 0.22, *p* = .82).

In sum, we found no conclusive evidence that children were better at learning the graphotactic simpler patterns relative to more complex ones, but also no evidence for the null.

5.3.2.4.2 (Unaware) adults

We conducted our analyses here excluding those participants in Experiment 3 who were coded as aware. In the fill-in-the-blanks task, there was substantial evidence for H_1, $BF_{(0, 0.35)}$ = 7.23, *RR* [0.10, 1.43], that is, adults were more accurate in choosing the correct vowel in the simpler relative to the complex pattern condition (effect of experiment: β = 0.37, *SE* = 0.16, *z* = 2.30, *p* = .02). However, for legality judgments the evidence was inconclusive, $BF_{(0, 0.21)}$ = 1.18, RR [0, >4.59] (effect of experiment: β = 0.12, *SE* = 0.11, *z* = 1.06, *p* = .29).

In sum, the evidence on adults' ability to learn the two types of patterns was mixed. They were better at selecting allowable word-medial vowels in the easier (Experiment 3) relative to the harder fill-in-the-blanks condition (Experiment 4), but there was no evidence that pattern complexity affected their ability to correctly discriminate between legal and illegal items.

5.3.3 Discussion

Experiment 4 introduced graphotactic constraints that were more complex and harder to articulate relative to those of Experiment 3. Specifically, word-medial vowels predicted the occurrence of both doublets and singlets but certain bigrams or trigrams (e.g., *uf, *uff*, *uss*, **us*, in list 1) were never allowed. We investigated children's and adults' ability to learn these constraints following the same procedure as in Experiment 3 and found significant evidence of learning in both participant groups. This result confirms that both age groups are capable of learning purely visual context-based patterns without any explicit instruction and after only a few minutes of exposure.

We also compared learning effects in Experiment 3 and Experiment 4, in order to investigate differences in generalization ability depending on the complexity of the pattern to be learned. For children, there was no substantial evidence that they learned the simple patterns better than the complex ones, but the *BF* was inconclusive: Thus, we cannot conclude that pattern complexity does not mediate learning performance. For adults, there was substantial evidence that easier patterns were learned better than the hard ones in one of the two tasks (fill-in-the-blanks) and inconclusive evidence in the other task (legality discriminations). Note that participants who showed explicit awareness (all in the easier Experiment 3) were excluded from these analyses. This mitigates the concern that they would drive the difference between conditions and the result suggests that the pattern tested in the current experiments is indeed harder for implicit learning mechanisms. Note, however, that removing unaware participants does reduce power (we provide power awareness for inconclusive results in the General Discussion of this chapter, Section 5.5).

While no children were able to describe Zorib patterns, the subset of adults (28%) who reported them in the questionnaire of Experiment 3 reached close-to-ceiling performance (90% accurate) in both of our tasks. This finding tentatively suggests a positive link between the ability to verbally articulate (i.e., being aware of) the patterns and strong performance in the task. It does not, however, settle how and when precisely explicit awareness emerged. One possibility is that awareness was a by-product of learning, only emerging in the end or even only when participants reflect on their performance at test. Alternatively, aware participants may have begun engaging deliberate hypothesis testing processes during learning. Given that performance was so much higher for explicit learners, the possibility that an explicit process may have been at work during the learning process raises a question with important implications for learning spelling patterns. Rather than relying on incidental learning alone, is learning more efficient when you are given explicit instructions about spelling patterns prior to print exposure? We investigate this in a final experiment (in this chapter) with child participants (instead of adults; to counteract potential ceiling effects and make results more relevant to literacy development) in which we provide explicit instruction as to the simpler spelling patterns of Experiment 3 (since these can be most straightforwardly verbalized as a rule).

5.4 Experiment 5: Explicit Learning of Graphotactic Patterns

5.4.1 Method

5.4.1.1 Participants

Using the same sampling plan as in Experiment 3 and 4, described in Section 2.6 and pre-registered at https://osf.io/m76ck, twenty-five typically developing Year 2 children (10 female; 15 male; mean age = 7.20 years, *SD* = 0.60) completed two sessions on two consecutive days. The same recruitment, consent, and compensation processes were used as in Experiments 3 and 4. As in our previous samples, the mean reading and spelling performance was above average (mean reading = 118.80, *SD* = 11.90; mean spelling = 114.40, *SD* = 10.60).

5.4.1.2 Materials

The stimuli were identical to those in Experiment 3.

5.4.1.3 Procedure

We replicated all aspects of the experimental procedure used in Experiments 3 and 4, except that, before exposure, children were explicitly told that written words in Zorib language adhered to a set of rules, which were described as follows: (In one counterbalanced list condition) “in Zorib language, the letter *e* is always followed by single letters (as in *rel, det*); and the letter *u* is followed by double letters (as in *rull, dutt*)”. Participants were then invited to perform the one-back task used in Experiments 3 and 4 over two sessions, followed by the fill-in-the-blanks and legality judgment post-tests.

5.4.1.4 Results

The design and hypotheses of this study were pre-registered at https://osf.io/m76ck. The data and analyses, as well as the priors used to model the theory are under the same links listed in the results section of Experiment 3 (Section 5.2.2).

Figure 5.8 shows the mean proportion of children’s correct responses in the fill-in-the-blanks and legality judgment task. On inspection, both appear to be above chance (50%), and this was statistically confirmed.

In fill-in-the-blanks task, we found substantial evidence for H_1, $BF_{(0,\ 0.39)}$ = 18.25, *RR* [0.17, > 4.59] (model intercept: β = 2.21, *SE* = 0.57, *z* = 3.87, *p* < .001). Children selected the correct vowel to form legal strings with better than chance accuracy.

Turning to children’s legality judgments, again, we found evidence of above-chance accuracy, $BF_{(0,0.16)}$ = 4.61, *RR* [0.112, > 4.59] (model intercept: β = 0.74, *SE* = 0.26, *z* = 2.85, *p* = .004), despite the fact that none of the children were able to report the rules they were explicitly taught.

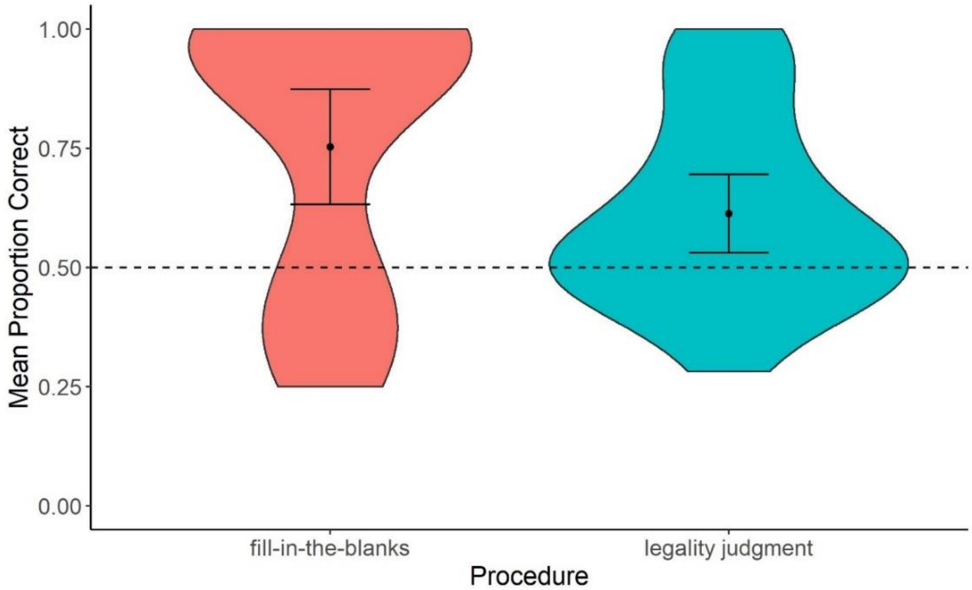


Figure 5.8: Children's Mean Accuracy (Violin Plots with 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment task in Experiment 5. The Dashed Line Represents Chance-level Performance (50%).

5.4.1.4.1 Comparison between Experiment 5 and Experiment 3

In Experiment 5, we investigated the effect of explicit instruction on 6–7-year-olds' ability to learn the novel graphotactic constraints of Experiment 3. Children were invited to play the same games as before, but were now told, at the beginning of the study, that all Zorib words were spelled according to a rule: This was explicitly taught and further illustrated via exposure to pattern-embedding instances. Results (both in the fill-in-the-blanks and legality judgment tasks) supported H_1, thus, performance was reliably better than chance, and in fact, numerically higher relative to the condition where patterns were learned from mere exposure to the pattern-embedding stimuli. To directly investigate whether there is a relative advantage for explicit instruction above incidental learning, we ran comparisons between Experiment 3 and Experiment 5, given that the stimuli in both experiments are identical.

In the fill-in-the-blanks task, we found substantial evidence for H_1, $BF_{(0,\ 0.83)}$ = 389 *RR* [0.104, >4.59] (effect of experiment: $\beta = 1.33$, $SE = 0.35$, $z = 3.83$, $p < .001$). Similarly, for the legality judgment task, we found substantial evidence for H_1, $BF_{(0,\ 0.38)}$ = 5.93, *RR* [0.13, 1.32] (effect of experiment: $\beta = 0.43$, $SE = 0.19$, $z = 2.19$, $p = .03$). In sum, children's performance was stronger under explicit instructions relative to comparable incidental exposure.

5.4.1.4.2 Associations between literacy performance and Experiments 3, 4, and 5

There is ongoing debate regarding the relationship between learning via statistical learning processes and written (as well as spoken) language ability, due to mixed significant and nonsignificant patterns of association shown in previous work. In our studies, a secondary goal was to explore associations between learning performance and accuracy on standardized (WRAT-4) reading and spelling, as well as TOWRE word and nonword reading performance. As noted in Section 5.2.2, these analyses were exploratory and therefore not pre-registered. We addressed this question among children and used BF analyses to quantify evidence for both H_1 (positive associations such that those who performed better in the learning tasks were also better readers/spellers) and the null (no relationship between statistical learning and literacy skills)[23].

Results from our correlation analyses are presented in Table 5.1. To sum up the results, in the incidental Experiments 3 and 4, the evidence for the association between learning performance and literacy was inconclusive in 9 out of the 16 occasions, while in five occasions, we demonstrated the null (no relationship between task performance and literacy): Fill-in-the-blanks performance in Experiment 3 and WRAT reading ability; fill-in-the-blanks performance in Experiment 4 and WRAT spelling, TOWRE word nonword reading ability; and legality performance in Experiment 4 and WRAT spelling ability. In order to maximize the available evidence, we also pooled across the two experiment datasets and explored correlations: We found substantial evidence for the null in half of the occasions: Fill-in-the-blanks performance and WRAT reading and spelling ability as well as TOWRE word reading efficiency; and legality performance and WRAT spelling ability. In the other occasions, the evidence remained inconclusive. Turning to Experiment 5 (explicit learning), we found substantial evidence for a positive association (better learning scores in better readers) between all but two occasions: Fill-in-the-blanks performance and

[23] See Section 2.3 for a description of the methods used and Table 2.4 for the priors used in this chapter

WRAT spelling ability and TOWRE nonword reading efficiency, where the evidence was inconclusive.

In sum, when conclusive patterns of correlations emerged, for incidental learning, these were evidence of no relationship between learning performance and literacy; for explicit learning, they were evidence for a positive association between explicit learning and literacy (H_1).

Table 5.1: Correlations between Accuracy in Experiment 3, 4 & 5 and WRAT Reading and Spelling Raw Scores (by Procedure)

Experiment & Procedure	Statistics	WRAT Reading	WRAT Spelling	TOWRE word	TOWRE nonword
		Experiment 3			
fill-in-the-blanks	*BF* [RR]	**0.17**[a] [0.24, >4.59]	**0.33**[a] [0.5, >4.59]	0.46 [0, 0.72]	0.61 [0, 1.01]
	p	.29	.95	0.67	0.46
	z_r (SE_{Zr})	-0.19(0.18)	0.01(0.18)	0.07(0.18)	0.13(0.18)
legality judgment	*BF* [RR]	0.76 [0, 1.28]	0.39 [0, 0.61]	0.56 [0, 0.91]	0.67 [0, 1.11]
	p	.35	.80	.52	.46
	z_r (SE_{Zr})	0.16(0.18)	0.05(0.18)	0.11(0.18)	0.14(0.18)
		Experiment 4			
fill-in-the-blanks	*BF* [RR]	0.58 [0, 0.94]	**0.27**[a] [0.41, >4.59]	**0.25**[a] [0.37, >4.59]	**0.29**[a] [0.44, >4.59]
	p	.61	.61	.54	.71
	z_r (SE_{Zr})	0.11(0.21)	-0.11 (0.21)	-0.13(0.21)	-0.08 (0.21)
legality judgment	*BF* [RR]	0.54 (0, 0.88]	**0.21**[a] (0.32, >4.59]	0.52 (0, 0.84]	0.41 (0, 0.64]
	p	.66	.37	.68	.91
	z_r (SE_{Zr})	0.09(0.21)	-0.19(0.21)	0.09(0.21)	0.02(0.21)
		Experiment 3 and 4, pooled			
fill-in-the-blanks	*BF* [RR]	**0.16**[a] [0.24, >4.59]	**0.21**[a] [0.32, >4.59]	**0.26**[a] [0.42, >4.59]	0.35 [0, 0.53]
	p	.55	.82	.92	.68
	z_r (SE_{Zr})	-0.08(0.13)	-0.03(0.13)	0.01(0.13)	0.05(0.13)
legality judgment	*BF* [RR]	0.65 [0, 1.05]	**0.18**[a] [0.42, >4.59]	0.50 [0, 1.07]	0.45 [0, 1.01]
	p	.32	.65	.44	.51
	z_r (SE_{Zr})	0.13(0.13)	-0.06(0.13)	0.10(0.13)	0.09(0.13)
		Experiment 5			
fill-in-the-blanks	*BF* [RR]	**3.46**[b] [0.20, 0.62]	1.40 [0, 2.30]	**4.87**[b] [0.16, 1.09]	1.41 [0, 2.32]
	p	.06	.22	.04	.22
	z_r (SE_{Zr})	0.40(0.21)	0.26(0.21)	0.44(0.21)	0.26(0.21)
legality judgment	*BF* [RR]	**9.42**[b] [0.12, 2.67]	**5.74**[b] [0.14, 1.38]	**21.52**[b] [0.10, >4.59]	**5.93**[b] **[0.14, 1.38]**
	p	.02	.03	.01	.03
	z_r (SE_{Zr})	0.52(0.21)	0.46(0.21)	0.54(0.21)	0.44(0.21)

[a] substantial evidence for H_0

[b] substantial evidence for H_1

5.5 General Discussion of Experiments 3, 4 and 5

In three comparable learning experiments, we assessed children and adults' ability to pick up on purely visual graphotactic constraints between word-final single/double consonants and letter-vowel context. Across experiments and age groups, learning was induced—over two brief sessions—by exposure to pattern-embedding stimuli that were to be attended as part of the requirements of a cover one-back task. Upon training completion (session 2), post-tests were used to measure whether participants had learned the patterns and could generalize over them in a legality judgment and production (fill-in-the-blanks) task. There were two key manipulations across the three experiments presented here: First, learning was induced either incidentally (no instruction to learn the patterns; Experiment 3 & 4) or explicitly (explicit rule teaching followed by the same cover task; Experiment 5); and second, patterns varied in complexity, such that, in one experiment, vowel identity predicted word-final singlet/doublet usage (Experiment 3 and 5: “Easy” graphotactic pattern), whereas in Experiment 4 the restriction concerned specific word-middle vowel + word-ending singlet/doublet combinations, rather than doubling per se. Further to the experimental learning task, child participants were administered measures of literacy (reading and spelling task), in order to explore associations between incidental/explicit learning and literacy performance.

There were three key results: First, both children and adults incidentally learned the novel graphotactic constraints that had no phonological counterpart and consistently generalized over the test stimuli: They produced permissible spellings by selecting allowable word-medial vowels and discriminated between permissible and impermissible strings with better-than-chance accuracy. Moreover, they were able to do this for more complex rules. Second, adults who picked up on the task explicitly did better; no children could report the untaught rules but a final experiment showed that explicitly teaching children a graphotactic rule yielded significant learning effects that were stronger relative to those induced incidentally. Finally, there was only evidence of significant associations between test generalization performance and performance on reading and spelling measures under explicit conditions. Incidental learning performance was either unrelated to participants’ reading/spelling ability, or data were inconclusive. I discuss these findings, in turn, below.

5.5.1 Implicit learning of visual graphotactic spelling patterns

The experiments in Chapter 5 provide the first strong demonstration that visual statistical learning processes underlie the learning and generalization of purely graphotactic *context-based* spelling patterns in children, controlling for the possibility that sensitivity to phonology accounts for this effect. Previous work measuring children’s sensitivity to spelling patterns within a familiar language has shown that children prefer stimuli that conform to such regularities. For example, Hayes et al. (2006) showed that English-speaking children prefer nonwords embedding doublets after single-vowel than double-vowel spellings (e.g., *saff* > **saf*, following *Jeff*/*bedding*/*pull* versus *deaf*/*heading*/*sail*) and have explained these effects in terms of statistical sensitivity to orthographic properties. However, the learning mechanisms at play were postulated, not directly shown; it is also hard—as with all studies in this line of work—to establish whether developmental patterns of improvement were due to maturational differences in learning ability or differences in the amount of children's exposure to print.

We addressed these limitations by manipulating patterns in a well-controlled learning experiment, maintaining the use of pronounceable stimuli to draw close parallels to how people learn existing writing systems. The same approach was taken by Samara and colleagues (Samara & Caravolas, 2014; Samara et al., 2019) who manipulated patterns of letter co-occurrence within consonant-vowel-consonant strings and interpreted their learning effects as evidence for graphotactic sensitivity. However, their stimuli and patterns were all pronounceable, thus, children's apparent graphotactic sensitivity could be underpinned by phonotactic learning (Chambers et al., 2003; Onishi et al., 2002).

It is also possible to use semi- or fully-artificial languages to this effect as seen in Chetail (2017) and (Nigro et al., 2016) and in Experiment 2 (Chapter 4), where I showed that children and adults learn positional graphotactic patterns following brief exposure to novel, pattern-embedding semiartificial stimuli. However, assessing learning of more complex, contextual constraints is expected to pose more challenging methodological constraints, as I discussed in Chapter 3, and so we decided instead to use letters that are familiar to children, but combined to generate homophone sets of letters (e.g., *dd* and *d* map onto the same sound).

One concern with using the children's native alphabet is that children may bring knowledge from English patterns which interfere with the novel graphotactic learning task. For example, some of the stimuli violate an existing English pattern—*f* would be doubled in these stimuli so those stimuli with ff may be easier to learn than those with *f*, where an existing pattern has to be unlearned. Critically, however, as explained under Procedure and section 5.2.1.2, such items are counterbalanced as legal and illegal between participants, so that such relationship with English stimuli may add noise but cannot underpin our effects.

One further potential concern was that, even though homophone letters are pronounced the same by skilled readers, children might have covertly pronounced them differently, implicating phonotactic learning in their performance. We therefore conducted a brief online study and clearly demonstrate that this was not the case: 14 children (6–7 years old) read aloud the full set of nonwords (presented one at a time) that comprised the stimuli in Experiment 2. Subsequently, two native English speakers who were blind to the purpose of the experiment, listened to all of the recordings that were grouped (unbeknown to them) into ‘homophone’ pairs (e.g., a child’s responses for *det* and *dett*) versus foil pairs (within-participant recordings were mixed at random). Of the 448 homophone trials (32 per participant), 85% of them were identified by both coders as the ‘same’ word, 11% were poor quality recordings (e.g., one of the recordings cut off prematurely or was missing), and only 4% were cases where the child actually produced two different words. Importantly, these were all idiosyncratic pronunciations made by individual children and mainly reflected differences in onset pronunciations (e.g., *ges* vs. *yes*) rather than in different pronunciations for the codas, mitigating the concern that homophone letters may have been pronounced differently by the children in the main experiments. One final point we acknowledge is that even if the difference between the conditioned graphotactic patterns has no phonological counterpart, the conditioning graphotactic environment—that is, the vowel—is itself pronounceable. As noted above, future experiments could address this by using an artificial lexicon of fully unpronounceable stimuli. In terms of the current study, our contribution is to establish that children can learn constraints on the occurrence of different graphotactic patterns even when the graphotactic differences are not supported by correlated phonological differences, and thus learning of the constraints cannot depend on learning phonotactics. In sum, we have thus demonstrated above-chance learning of visual graphotactic constraints which cannot be underpinned by phonotactic learning. We provide the first direct evidence for this in relative beginner spellers tested on context-based graphotactic patterns.

We further investigated the effect of complexity on generalization ability in childhood and adulthood and expected that learning to predict word-final singlet/doublet usage from previous vowel context would be easier than learning specific combinations between word middle vowels and subsequent single/double letters. This is because, while the underlying joint probabilities of vowels and word-ending letters were the same across conditions (1 in 4 for each vowel and final (single/double) letter combination), there is an additional (higher-level) joint probability statistic underlying the stimuli used in Experiment 3. Namely, participants could pick up on the joint probability of 1 between Vowel1 and singlets, as opposed to the joint probability of 0 between the same vowel and doublets. In Experiment 4, the joint probability of Vowel1 and singlets was .50 and so was the joint probability of Vowel1 and doublets.

For adults, we found substantial evidence that easier patterns were learned better than hard in one of the tasks, namely, the fill-in-the-blanks task. Note that our analyses included only unaware participants, indicating that this difference is not due to the rule in Experiment 3 being easier to articulate and, thus, picked up explicitly: Notably, this is easily verbalized as "*e* is always followed by singlets, never doublets". The evidence regarding adult performance in legality judgments and children's performance in both tasks was, on the other hand, inconclusive. Thus, we cannot interpret null findings for children as demonstrating that they learn simple and complex patterns similarly. Supplementary analyses (assuming that the error term would reduce in proportion to $\sqrt{SE}$; see https://osf.io/tsxf3/) suggest that, to establish H_1 (i.e., demonstrate better learning for the easier patterns) 4000 more child and 130 more adult participants would be needed, which we deem impracticable, suggesting that the current paradigm is not sufficiently sensitive to detect small differences in learning. Thus, in sum, our work tentatively suggests that in adults, the easier patterns were better learned by implicit learning mechanisms than the harder patterns, but we can draw no conclusions for children. What the current results clearly establish, is that the complex patterns *can* be learned, as well as the simple ones, even under incidental, implicit learning conditions by children and adults.

5.5.2 Explicit learning of visual graphotactic spelling patterns

In Experiment 5, we assessed children's ability to learn novel graphotactic constraints under explicit task instructions, and contrasted—for the first time, to our knowledge—the relative effectiveness of explicit and incidental processes involved in orthographic knowledge acquisition using artificial language methods. Much of previous work that has sought to address questions related to the effectiveness of implicit and explicit orthographic pattern learning suffers from methodological weaknesses. For example, only two orthographic learning studies

(Kemper et al., 2012; Sobaco et al., 2015) have controlled for idiosyncratic task differences by using implicit and explicit versions of the same tasks.

In our well-controlled experiments, learning patterns explicitly was clearly advantageous for performance in both of our tests. This demonstration converges with the practice of teaching spelling patterns from early on in literacy instruction.

5.5.3 Associations between learning performance and literacy skill

As in the previous chapters, we did not replicate the evidence for positive association between statistical learning and literacy performance under incidental learning conditions (Experiment 3 and 4). Some of the relevant correlations in each implicit experiment were inconclusive (i.e., are neither evidence for H_1 or for H_0), and when data was pooled across the two experiments, we conclusively showed that variations in learning skill were unrelated to variations in spelling ability. It is possible that the null findings reflect our methods, in that real graphotactic knowledge may have competed with graphotactic manipulations in our experiments. In this case, robust internalized representations of English graphotactics among better learners could cancel out learning of the experimental constraints. Further limitations regarding such results are discussed in 3.2.3.2 (Chapter 3), including lack of strong measures of validity and reliability in our tasks. The use of Bayes factors, however, allows us to say that, in half of the comparisons (for the pooled data), we do have substantial evidence that there was no relationship here. Once again, we interpret such result with caution and we suggest future directions in the General Discussion (Section 7.6).

Turning to the pattern of correlations seen with measures of explicit learning, we obtained substantial evidence for H_1 (positive association between statistical learning and literacy performance) for measures and variations in reading/spelling in all but two occasions that involved the fill-in-the-blanks task (possibly due to reduced item-based power in this task; 36 more participants needed to demonstrate $H_{1,}$ assuming that the error term would reduce in proportion to $\sqrt{SE}$; see https://osf.io/tsxf3/). These positive associations strengthen the view that explicit learning contributes to learning to read and spell. We acknowledge that our data are correlational and not evidence for causation, hence, it is, for example, possible that children with better reading skills *also* have stronger language skills that help them better understand the task instructions, rather than being better at explicit learning, leading to being a better speller. Last, while we maintain that our analyses are exploratory, the contrast in the implicit and explicit correlational results is intriguingly consistent with a general theory of implicit learning being less susceptible to individual differences than explicit learning due to its evolutionary precedence (Reber, 1989; Reber et al., 1991). This has been most strongly shown in terms of associations with memory and intellectual ability (Gebauer & Mackintosh, 2007; Kaufman et al., 2010).

Returning to the results from the implicit learning experiments, the results add to those of Experiment 2 and demonstrate that visual statistical learning processes support the learning of graphotactic patterns even when they do not correlate with phonology. This suggests similar mechanisms may operate in naturalistic contexts, that is, developing readers may acquire knowledge of spelling rules simply from incidental exposure to text. However, one common factor in all experiments so far may limit the relevance for learning in naturalistic contexts: Although participants were not instructed to focus on spelling (other than by implication in the explicit conditions), task was such, that the participants were nevertheless likely to focus on form, unlike real-life learning situations, when children encounter novel written words and spelling patterns in a *meaningful* context. Therefore, in the next chapter (Chapter 6), the methodology is adapted to incorporate the learning of homophonic letter strings into a word learning task.

6 Implicit and Explicit Learning of Graphotactic Contextual Patterns in a Word Learning Paradigm

6.1 Introduction

In Chapters 4 and 5 of this thesis, I have shown that both adults and children can learn purely graphotactic patterns, via implicit learning processes. I have presented evidence for generalization of both (i) unconditional, positional constraints embedded in semiartificial strings (Experiment 2), as well as (ii) constraints conditioned by context with a simple (i.e., easy to verbalize; Experiment 3) and more complex (Experiment 4) structure, embedded in homophonic nonwords. These findings are in line with literature that explains that graphotactic learning is underpinned by statistical learning mechanisms (Pacton et al., 2001) seen to operate over spoken language as well as other nonlinguistic input (see Section 1.5 for an overview of statistical learning literature). However, much of this evidence comes from experimental conditions in which participants are presented with nonwords in isolation, and are asked to make judgments about their familiarity with the previously presented novel instances (the experiments in the current thesis and those in Section 1.6) or words that exist in their own language (e.g., Cassar & Treiman, 1997; Danjon & Pacton, 2009; Pacton et al., 2001; see also Section 1.3). Such designs provide a single source of information for participants to process, namely the graphotactic information (i.e., the "form" of the words), while in natural language, children and adults learn to spell from reading words for meaning. In the current chapter, therefore, we adapted our design to investigate whether spelling patterns can still be learned implicitly under conditions where participants are not just reading the words but are also learning their meanings. Research in reading as well as spelling literature has emphasized the importance of generalizability of laboratory research to natural contexts and I present below studies that have incorporated meaning to the design assessing sensitivity to spelling patterns.

Research in support of the "self-teaching" hypothesis (Share, 1995) agrees that children and adults can teach themselves to spell through reading (Burt & Butterworth, 1996; Burt & Fury, 2000; Cunningham et al., 2002; de Jong & Share, 2007). According to this hypothesis, once unfamiliar words are translated from letters into sound, readers acquire most of the orthographic knowledge incidentally, while reading independently, and thus teaching themselves to read. Studies that tested the self-teaching hypothesis have typically required participants to read aloud a short passage of text that contains some novel pseudowords. Orthographic learning was then tested in subsequent tasks (usually several days later), such as orthographic choice, spelling, and naming of the target (studied) pseudowords. However, these studies focused on fluency during reading, without specifying how reading for meaning affected orthographic learning. To address this, some early research by (Gilbert, 1934, 1935) and Ormrod (1986) showed that, later in literacy (in high school and college years in the US), when systematic spelling instruction is no longer provided, improvement in spelling is accounted for by reading for meaning, that is, for the purpose of answering questions on the content. More relevant to the current chapter, Pacton, Borchardt, Treiman, Lété, and Fayol (2014) directly examined French undergraduate students' knowledge of graphotactic patterns, in a task where reading for meaning was an additional source of information. Specifically, they presented participants with six stories containing a trisyllabic nonword (the name of the hero in the story) embedding consonants that either frequently double (e.g., *n*, *r*, or *t*) or rarely double (e.g., *b*, *d*, or *g*) in French. The stimuli, therefore, contained either no doublets (e.g., *bagotin*), frequent doublets (e.g., *bagottin*) or rare doublets (*baggotin*), thus testing participants' sensitivity to doubling patterns that follow the probabilistic graphotactics in their own language. Other studies that have tested participants' sensitivity to patterns in their own language, e.g., Wright and Ehri (2007), discussed in Section 1.3.1, taught participants novel spellings and asked them to make decisions on items that either conformed to– or violated real patterns found in their language (i.e., English). Unlike studies such as Wright and Ehri's (2007), Pacton et al. (2014) presented doubling patterns that did not violate the French graphotactics and the participants were not specifically asked to focus on the spellings (i.e., they were not explicitly taught). At test, participants were either asked to (i) write responses (including the target word) to follow-up questions about the stories (Experiment 1), (ii) spell the target word from dictation (Experiment 2), or (iii) identify the correct spelling of the target word from three options. In all three situations, the participants better recalled the correct spelling of the target words when these contained common or no doublets, compared to when they contained rare doublets, even though the frequency of these types of patterns was equated in the stories. These results suggest that existing knowledge about graphotactics influences learning of novel spellings even in situations that are more naturalistic, such as reading for meaning. They add to literature that showed that the frequency of words encountered by readers in written text (e.g., Kreiner & Gough, 1990; Lété, Peereman, & Fayol, 2008) as well as sound-to-spelling correspondences typical of a writing system (Wang, Castles, Nickels, & Nation, 2011) influence learning of novel words. As I noted in Section 1.3.5, questions of learnability—that is, how much exposure is needed and how different patterns are learned——are hard to address in studies that test knowledge of patterns that exist in participants' actual orthography. Learning experiments such as the ones presented in the current thesis allow control over the input learners receive. Joseph, Wonnacott, Forbes, and Nation (2014) showed that adults acquire new written words in naturalistic experimental conditions: They asked participants to read meaningful sentences embedding novel word forms, and monitored their eye movements. They found that, with each repeated encounter with a novel word, reading times reduced, indicating increased familiarity with the word and a concurrent decrease in processing time. Such frequency effects were also shown in Nation, Angell, and Castles (2007) where 8–9-year-olds were shown to be more likely to learn novel words when they appeared more frequently in the input, compared to those that appeared less frequently. They also found that orthographic learning did not differ depending on context, that is, when words were presented either in meaningful text or in isolation. For graphotactics, Pacton et al.'s (2014) study clearly demonstrates the effect of frequency, that is, sensitivity to which consonants double frequently in French. However, they do not address the more complex probabilities, such as where or when these double (e.g., in French, consonants can only double in word-medial positions). As shown in Chapter 5, children and adults can also implicitly learn consonant doubling patterns of varying complexity that are conditioned by surrounding context. We therefore aim to investigate whether adults can learn contextual graphotactic patterns of a more complex structure (i.e., harder to verbalize, as in Experiment 4) when they have an additional source of information (i.e., meaning) to attend to.

While the studies presented above show that adults learn spellings from reading meaningful text, when testing children's learning of complex graphotactic patterns rather than word recognition, such methods may add noise to a set of data that already shows relatively small effects. Although the studies presented in this chapter are with adults, our goal was to develop a paradigm that could be used to investigate developing spellers' learning, as in the previous experiments in this thesis. Thus, we aimed to devise a child-friendly task. The word-picture matching task, for example, is used in psycholinguistics mainly for spoken (e.g., Kapatsinski & Johnston, 2010) rather than written words and for the purpose of testing learning rather than as a task during learning. Wonnacott and colleagues (Vujović, Ramscar, & Wonnacott, n.d.; Wonnacott, 2011; Wonnacott et al., 2017, 2008) used either pictures of familiar objects and animals, or unfamiliar objects, to teach children and adults novel vocabulary or language structures (e.g., verb argument structures and affixes) and we follow this paradigm to teach adults a novel vocabulary for novel objects, while manipulating the internal structure of words (i.e., graphotactics) that participants will be tested on. This enables us to introduce child-friendly, naturalistic learning conditions while using stimuli and test tasks comparable to our previous studies.

6.1.1 The Current Study

A key aim in this chapter was to investigate whether graphotactic patterns can be learned implicitly, under more naturalistic conditions, where participants are not just reading the words but are also learning their meanings. The stimuli were identical to those in Experiment 4, where contextual graphotactic patterns were embedded in homophone stimuli spelled with single versus double letters meaning that singlet/doublet usage was predicted by the identity of the preceding vowel, such that each of the two possible vowels predicted some singlets/doublets. Each novel stimulus was paired with an image of an unfamiliar object to create an object-meaning association.

In the exposure phase of the experiment, participants were again exposed to the relevant patterns in two brief sessions, but in this experiment this occurred while they performed a newly devised word learning task where they were asked to select one of two alien words that "goes with the alien object", followed by feedback. At first, they will be guessing, but over time, they are expected to learn the word picture associations. The second test phase of the experiment tested generalization in two tasks identical to Experiment 4: fill-in-the-blanks and legality judgment.

As in all previous experiments, we look for above-chance learning of graphotactic patterns in each of the post-exposure tests (fill-in-the-blanks and legality judgment). In addition, for the experiments in this chapter, we look at performance in the word learning task that is used at exposure, as this will provide a measure of how well participants have learned the link between the orthographic form and the associated pictures. Here, we look for both above-chance learning and investigate if learning improves over the course of exposure. As the tests in the two phases (word learning exposure and tests of graphotactic generalization) tap into different mechanisms, we also look at the relationship between the performance in these phases, that is, whether those who are better at learning the spellings are also better at learning the meanings. Note that this relationship could conceivably be in either direction, that is, it could be that, generally, good learners are better at both, but also that, learning the spellings occurs at the expense of learning word meanings, in which case the relationship would be negative. We therefore test for both directions and thus, these analyses are rather exploratory. As in Chapter 3 (Experiment 1), we also compute the association between the two graphotactic tasks, for comparison purposes.

In this chapter, another potential question of interest is whether embedding the graphotactic learning into a word learning task has affected learning compared with when we used the cover task. We therefore compare the performance in Experiment 6 against that in Experiment 4 for both the fill-in-the-blanks and legality judgment tasks, to test the hypothesis that performance will be higher in Experiment 4, where participants could focus fully on form.

As in previous studies, we also administered a post-experiment verbal questionnaire.

6.2 Experiment 6: Implicit Learning of Graphotactic Patterns in Word Learning Paradigm

6.2.1 Method

6.2.1.1 Participants

The sample size was estimated using the procedure outlined in Section 2.6, however, this plan was not pre-registered (we have, however, previously established the details within our lab). Because we expected participants to drop out after the first session, we tested in excess of the 25 required for a first look at the data (as per our optional stopping plan) and stopped if we found substantial evidence for above-chance learning in the graphotactic tests.

Using this approach, thirty-six adults (28 female, 8 male; mean age = 36.5 years, *SD* = 12.81) participated in Experiment 6. They were recruited via Prolific (https://www.prolific.co/) and were tested online. They reported being monolingual native speakers of English with no language, hearing, vision impairments or any learning difficulties. They all provided informed consent and were paid for their participation.

6.2.1.2 Materials

Written stimuli used in exposure and testing were as in Experiment 4, that is, four counterbalanced lists of 16 nonwords each, which embed a graphotactic constraint shown in Figure 5.5 (in Chapter 5). The nonwords used at exposure (legal seen list, see Section 5.3.1.2 for details) were a word tag for a picture of a novel, unfamiliar object from Horst and Hout (2016) (the full list of nonwords and the picture allocation is available in Appendix I). The stimuli presented at test (fill-in-the-blanks and legality judgment), however, did not appear with associated pictures, as only graphotactic learning was tested here.

6.2.1.3 Apparatus

As in Experiments 2, 3 and 4, with adults, the experiment was designed on the Gorilla.sc platform (Anwyl-irvine et al., 2019) and testing was carried out online using a link distributed via Prolific (https://www.prolific.co/).

6.2.1.4 Procedure

Testing was carried out online and all adults completed each 15 minute session on two consecutive days.

Practice task. Participants were presented pictures of animals (e.g., cat, dog, fox and hen) and two real simple English words under each picture, one corresponding with the animal in the picture and one of a different animal. They were asked to select the name of the animal seen on the screen (procedure detailed in the section below: Exposure task).

Exposure task. We retained our approach for a child-appropriate learning task as in our previous experiments, and, in addition, incorporated a word-learning paradigm at exposure. Participants were first told that they were going to see pictures of objects that are used on the planet Zorib along with written words from Zorib language. They were then informed that they will play three games with these words. In session 1, the game (word learning task) was to select one of the two words that name a Zorib object. In session 2, further word learning task was performed, followed by disclosure of the patterned nature of Zorib words (i.e., informing participants of the existence of patterns in the words seen).

A total of 288 Zorib objects (144 presented in 3 blocks in each session; 9 repetitions/object in each session) and 576 Zorib “words” (288 presented in 3 blocks in each session; 18 repetitions/string in each session) were shown in the context of a word learning task: Participants were instructed to look at each object and choose from the two word options, the one that names the object (see Figure 6.1). They were also informed that, since this is an unfamiliar language to them, feedback will be provided after each trial. Note that, while one word option was the target (i.e., corresponding to the picture), the other word was randomly selected from the same *legal seen* list (i.e., the nonwords conformed to the graphotactics of the language but were not the ones used at test, i.e., legal unseen). The feedback, therefore, did not target the learning of the correct spelling (graphotactics) but targeted solely the learning of word-meaning association. No other instructions were given. The objects were presented in the middle of a white background and remained there for 5,000ms (timelimit) or until a response was given. Two response buttons with the words in white letters on a black background were presented at the bottom of the screen and these appeared 350ms after the object presentation. Therefore, a response was allowed only after 350ms. If the response was correct, the green tick mark appeared on the right of the screen, and if the response was incorrect, a red cross mark appeared, while both words remained on the screen and the selected one was highlighted (i.e., the box around the word was darker) (see Figure 6.1 for an illustration). Participants were not given the option to change their mind, and their responses were recorded by the program as either correct or incorrect, trial by trial. A fixation point (black cross, presented for 500ms) followed the response, in the middle of the screen. Picture order was fully random and the word presentation was semi-randomized, such that each corresponding word appeared on the right or the left on the screen with equal probability, and all words appeared equally frequently.

Figure 6.1: Illustration of the Word Learning Task in Experiment 6 and 7

Tests. As in the other experiments in this thesis (Experiment 1, 3, 4, and 5), we used two test tasks: Fill-in-the-blanks and legality judgment (in that order), and the stimuli were identical to those in Experiment 4. Note that no picture stimuli were presented at test (as noted in Section 6.2.1.2).

Awareness questionnaire. A brief questionnaire was administered to assess whether participants were able to verbalize the graphotactic constraints governing Zorib words. If a participant reported that they noticed patterns before they were informed regarding their presence, further questions probed what patterns they thought they noticed and how they made their choices in each of the two tests.

6.2.2 Results

The design and hypotheses of this study were not pre-registered, however, the priors used to inform the theory (specified for each analysis in Table 2.5 and Table 2.6) come from the data available in this thesis, using methods described in Chapter 2, which have been established before this study was carried out. Data is available at https://osf.io/pes39/[24], and data analyses are available at https://rpubs.com/DSingh/WordLearning.

Separate analyses are presented for the data from the exposure task, which taps word learning, and the two test tasks which tap graphotactic learning

[24] See also Appendix B

(fill-in-the-blanks and legality judgment). We also look at the correlation between performance in the word learning task and the test tasks.

6.2.2.1 Word learning task

Figure 6.2 shows the mean proportion of adults' correct responses in the word learning task in the six blocks across the two exposure phases. Our data provided substantial evidence for H_1: Adults were better than chance (50%) at choosing between the correct and incorrect word labels for the novel objects, $BF_{(0,\ 0.71)}$ > 10,000, *RR* [0, >4.50] (model intercept: β = 1.46, *SE* = 0.17, *z* = 8.50, *p* < .001). We also looked for an effect of training-block (indicating increased performance through the exposure) and found an effect $BF_{(0,\ 0.20)}$ > 10,000, *RR* [0, >4.50] (effect of block: β = 0.33, *SE* = 0.06, *z* = 5.81, *p* < .001), with better accuracy in block 6 (*M* = .84, *SD* = 0.13) compared to block 1 (*M* = .63, *SD* = 0.09).

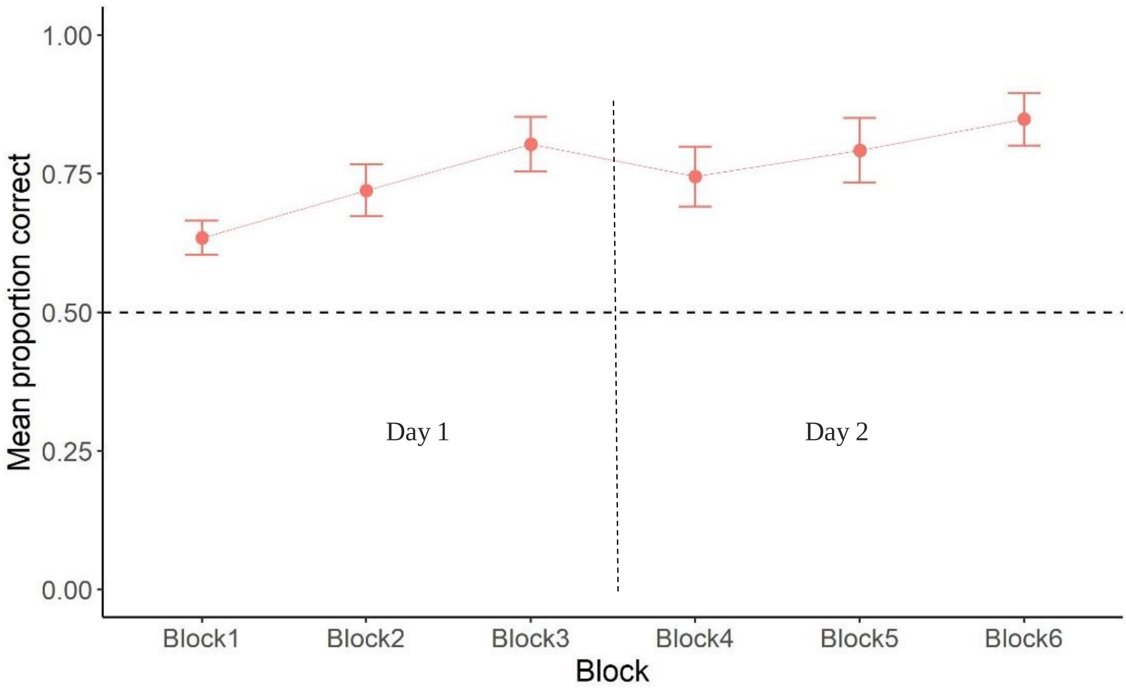


Figure 6.2: Adults' Mean Accuracy (Line Chart With 95% Confidence Intervals) in the Word Learning Task in Experiment 6, by session (day) and block. The Dashed Line Represents Chance-level Performance (50%).

6.2.2.2 Graphotactic learning test tasks

Figure 6.3 shows the mean proportion of adults' correct responses in the fill-in-the-blanks task. Our data provided substantial evidence for H_1, that adults were better than chance (50%) at choosing the correct vowel, $BF_{(0,0.19)}$ = 6.49, *RR* [0.06, 0.33] (model intercept: β = 0.18, *SE* = 0.8, *z* = 2.16, *p* = .03).

Figure 6.3 also shows the mean proportion of adults' correct legality judgments. There was evidence for above (50%) chance learning, $BF_{(0, 0.16)}$ = 7.07, *RR* [0.05, 0.60] (model intercept: β = 0.16, *SE* = 0.07, *z* = 2.23, *p* = .03), that is, adults were better than chance at discriminating between legal and illegal items.

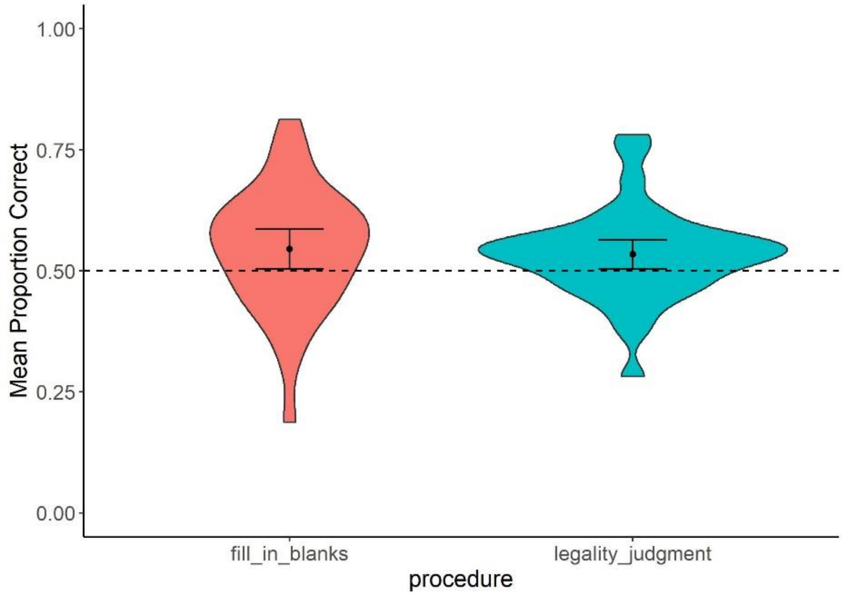


Figure 6.3: Adults' Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 6. The Dashed Line Represents Chance-level Performance (50%).

6.2.2.3 Associations between the two tests of graphotatic learning

Our data provided substantial evidence for H_1, that adults' performance in the fill-in-the-blanks task positively correlated with their performance at legality judgment task, $BF_{(0,0.58)}$ = 651, *RR* [0.05, >4.59] (*r*(34) = 0.59, *p* < .001).

6.2.2.4 Associations between word learning and graphotactic learning

The performance at word learning task in block 6 was used for correlational analyses, to capture learning following feedback received throughout the word learning task.

We did not find conclusive evidence for a positive association between performance at word learning task and fill-in-the-blanks ($BF_{(0,0.29)}$ = 1.18, *RR* [0, 1.42]; *r*(34) = 0.17, *p* = .32) or legality judgment task, $BF_{(0,0.29)}$ = 1.79, *RR* [0, 2.35]

($r(34) = 0.23$, $p = .17$). Testing for a negative association, however, we found substantial evidence for the null in both cases, that is, there was no negative correlation between performance at word learning and fill-in-the-blanks ($BF_{(0,0.29)}$ = 0.29, *RR* [0.25, >4.59] ($r(39) = 0.17$, $p = .32$)) or legality judgment ($BF_{(0,0.29)}$ = 0.25, *RR* [0.21, >4.59] ($r(34) = 0.23$, $p = .17$)) tasks.

6.2.2.5 Comparison between Experiment 4 and Experiment 6: performance in fill-in-the-blanks and legality judgment tasks

We compare performance in the current experiment (Experiment 6) at the two test tasks (fill-in-the-blanks and legality judgment, against performance in identical tests in Experiment 4. We test the hypothesis that performance would be better in Experiment 4, where adults were exposed to the patterns under a cover task, and therefore could focus all resources on form, rather than on learning the form-meaning mappings.

The mean proportion of adults' correct responses in the fill-in-the-blanks and legality judgment tasks in Experiment 4 are shown in Section 5.3.2.2, Figure 5.7, and those in Experiment 6 are shown in Section 6.2.2.2, Figure 6.3. For both tasks, the evidence was inconclusive. That is, our data does not provide substantial evidence for the hypothesis that adults were better at selecting the correct vowel to create legal strings in Experiment 4 (exposure under cover task) compared to Experiment 6 (exposure under word learning task) (fill-in-the-blanks: $BF_{(0, 0.18)}$ = 0.57 *RR* [0, 0.34] (effect of experiment: $\beta = 0.004$, $SE = 0.12$, $z = 0.03$, $p =.97$); legality judgment: $BF_{(0, 0.16)}$ = 0.51 *RR* [0, 0.26] (effect of experiment: β = -0.004, $SE = 0.09$, $z = -0.04$, $p = .97$)).

6.2.2.6 Awareness Questionnaire

Participants' responses were coded in a similar way as in previous experiments, to identify if they could accurately describe the novel graphotactic pattern on doubling and vowel co-occurrence. See Section 3.2.2.8 for the coding of responses. Participants' responses were coded such that, if any response described at least one aspect of the patterns embedded, such as "*f* and *l* double after *e"* was coded as explicit, while other vague responses such as "*g* was followed by *e*" or "double *t* and *f* were common" were coded as *other*.

Thirteen participants selected "no" when asked if they had awareness of patterns, ten selected "I don't know" and thirteen selected "yes". However, none of the participants were able to accurately describe at least one aspect of the manipulated pattern, and of those who selected "yes", responses ranged from "I realized towards the end that all their words ended in a double letter *ff* for example", to "I used similar endings to the words I saw in first game like *det* and *mutt*" and "short words, the second letter was usually a vowel (*e* or *u*). Double *t* and *f* were common."

6.2.3 Discussion

To sum up the results in Experiment 6, adults learned both (i) object-word associations in the exposure task, that is, they selected the correct words to label the objects with above-chance accuracy and their performance increased over the six blocks; and (ii) graphotactic spelling patterns—that is, they showed better-than-chance accuracy when selecting the vowel to create legal stimuli in the fill-in-the-blanks task, and when discriminating between legal and illegal strings in the legality judgment task. The awareness questionnaire indicated that none of the participants was able to describe the graphotactic patterns, which we take as evidence that learning of these patterns was implicit. We also looked for an association between the tasks. First, for the two test tasks (fill-in-the-blanks and legality judgment) we found that, as in Experiment 1, they are indeed positively correlated: Those adults who selected the correct vowel to create legal strings were also those who were more accurate at discriminating between legal and illegal items. This result further validates the novel (fill-in-the-blanks) task introduced in this thesis, and reinforces its use as a measure for the robustness of our results. It also shows that the knowledge the two tests tap into, is comparable. In contrast, the evidence for a positive association between the performance in word learning and the two graphotactic test tasks was ambiguous; however, testing for a negative association, there was evidence for the null. This indicates that we can conclude that learning in the two types of tasks did *not* impede each other.

Finally, we also compared the performance in the graphotactic learning task in Experiment 4—where participants performed a cover task at exposure, and where they could focus solely on the form of the nonwords—with that in the current experiment—where they learned the patterns while performing a word-learning task. We predicted that they could have stronger performance in the former task, since in the current experiment focus was now distributed between form and meaning. In fact, we did not see evidence for a difference here, but also no evidence for the null. The lack of difference is in line with the findings of Nation et al.'s (2007) study, which suggests that meaningful context may not play a role in learning orthographic patterns: Their participants' performance was similar regardless of whether the words were presented in meaningful text or in isolation. However, the inconclusive Bayes factors mean that we should be cautious in interpreting this results (and note that Nation et al. (2007) did not run Bayes factor analyses, so it is possible that their results, too, were insensitive rather than evidence for the null). In addition, there is another important difference between our experiments (Experiment 4 and 6) that may have led to inconclusive results here: Although the number of trials was equated across the experiments, such that each word in Experiment 6 was seen as a target equally frequently to how often it was presented in Experiment 4, because words were also seen as foils 288 times (across the two sessions), the total exposure to each word was greater in Experiment 6. It is thus possible that, did we not have the increased exposure, we would have seen better performance in the situation where the word form was the only source of information participants attended to (i.e., Experiment 4).

In sum, we have seen that participants can implicitly pick up on graphotactic patterns even when learning under meaningful conditions. However, in Chapter 5 (Experiment 5), we also showed that even in conditions where implicit learning occurred, under explicit rule instruction, graphotactic learning improved. Ormrod (1986) explored the effect of learning spellings of novel pseudowords from meaningful context, by instructing participants to either (i) focus only on the details of a story; (ii) prioritize learning the details of the story and also learn the spellings of pseudowords; or (iii) prioritize learning of spellings and *also* focus on the details of the story. She found that learning of spelling was greater when they prioritized spelling while comprehension remained unaffected. In this chapter, we contrast once again the implicit and explicit learning of adults and examine whether learning advantages in spelling occur. As in Ormrod (1986) we also investigate if learning the patterns has an impact on learning meanings.

In Experiment 7, we use the same nonword stimuli as those in Experiment 4: that is, embedding the more complex orthographic pattern. Note that, by using the more complex patterns, we provide the first test of the benefits of explicit learning of a pattern that is harder to describe as a rule. This is also contributing to the real-life validity of our study, considering the abundance of English useful vocabulary statistics that are too complex to verbalize, both in childhood and adulthood (Kessler, 2009).

Experiment 7 is identical to Experiment 6, except for the addition of the explicit instruction prior to exposure. Therefore, all the tests remain the same as in Experiment 6 (see Section 6.1.1), except the comparison with Experiment 4. Critically, we include additional analyses testing the contrast between performance in Experiment 7 against that in Experiment 6, that is, between implicit and explicit learning. Importantly, we test for a difference in graphotactic learning, which is predicted to be greater in explicit than implicit experiments, but also for a benefit in the reverse direction for word learning, that is, we predict that, if participants are focused on spelling, they might do worse in word learning.

6.3 Experiment 7: Explicit Learning of Graphotactic Patterns in Word Learning Paradigm

6.3.1 Method

6.3.1.1 Participants

Using the same method to estimate the sample size as in Experiment 6 (see Section 6.2.1.1), forty-one adults (23 female, 18 male; mean age = 34.9 years, *SD* = 13.82) participated in Experiment 7. They were recruited via Prolific (https://www.prolific.co/) and were tested online. They reported being monolingual native speakers of English with no language, hearing, vision impairments or any learning difficulties. They all provided informed consent and were paid for their participation.

6.3.1.2 Materials

The stimuli were identical to those in Experiment 6.

6.3.1.3 Procedure

We replicated all aspects of the experimental procedure used in Experiment 6 except that, before exposure, adults were explicitly told that written words in Zorib language adhered to a set of rules, which were described as follows: (In one counterbalanced list condition) “in Zorib language, the letters *f* and *l* double after *e* but not after *u*—e.g., *geff* and *guf* are possible Zorib words;

gef and *guff* are not. Also, *s* and *t* double after *u* but not after *e*—e.g., *duss* and *des* are possible Zorib words; *dus* and *dess* are not."

6.3.2 Results

As in Experiment 6, the priors used to inform the theory are specified in Table 2.5 and Table 2.6 and data is available at https://osf.io/pes39/ and data analyses are available at https://rpubs.com/DSingh/WordLearning.

Again, separate analyses are presented for the word learning task and the two test tasks. We also look at the correlation between performance in the word learning task and the test tasks. In addition, we compare the performance in these tasks in the current experiment with that in Experiment 6 (implicit graphotactic learning under word learning conditions).

6.3.2.1 Word Learning Task

Figure 6.4 shows the mean proportion of adults' correct responses in the word learning task. Our data provided substantial evidence for H_1: Adults were better than chance (50%) at choosing between the correct and incorrect word labels for the novel objects, $BF_{(0,\ 1.46)} > 10{,}000$, *RR* [0, >4.50] (model intercept: $\beta = 0.71$, $SE = 0.10$, $z = 7.35$, $p < .001$). We found an effect of block, $BF_{(0,\ 0.33)} > 10{,}000$, *RR* [0, >4.50] (effect of block: $\beta = 0.20$, $SE = 0.03$, $z = 6.17$, $p < .001$), with better accuracy in block 6 ($M = .75$, $SD = 0.17$) compared to block 1 ($M = .55$, $SD = 0.11$).

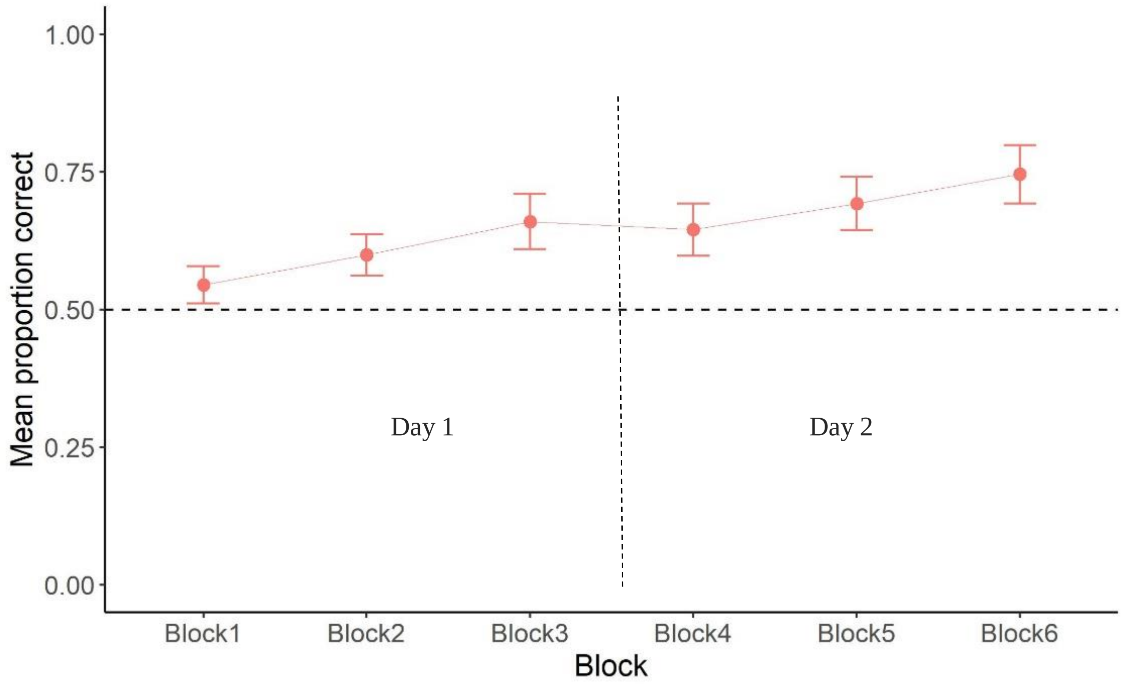


Figure 6.4: Adults' Mean Accuracy (Line Chart With 95% Confidence Intervals) in the Word Learning Task in Experiment 7, by Block. The Dashed Line Represents Chance-level Performance (50%).

6.3.2.2 Graphotactic learning test tasks

Figure 6.5 shows the mean proportion of adults' correct responses in the fill-in-the-blanks task. Our data provided substantial evidence for H_1, that is, adults were better than chance at choosing the correct vowel, $BF_{(0,0.19)}$ = 52.76, *RR* [0.06, >4.50] (model intercept: β = 0.88, *SE* = 0.22, *z* = 4.08, *p* < .001).

Figure 6.5 also shows the mean proportion of adults' correct legality judgments. There was evidence for above (50%) chance learning, $BF_{(0, 0.16)}$ = 10.94, *RR* [0.07, >4.50] (model intercept: β = 0.80, *SE* = 0.23, *z* = 3.56, *p* < .001), that is, adults were better than chance at discriminating between legal and illegal items.

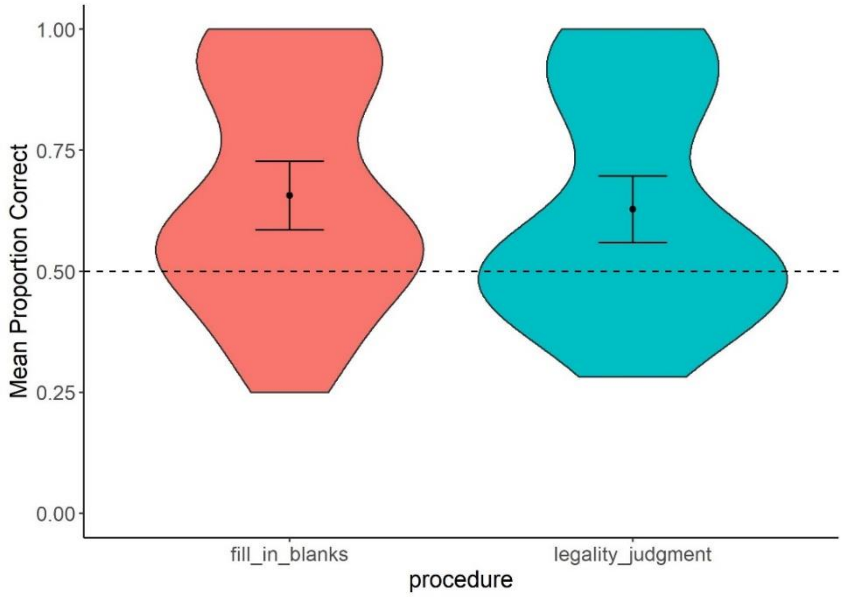


Figure 6.5: Adults' Mean Accuracy (Violin Plots With 95% Confidence Intervals) in the Fill-in-the-blanks and Legality Judgment Task in Experiment 7. The Dashed Line Represents Chance-level Performance (50%).

6.3.2.3 Associations between the two tests of graphotactic learning

Our data provided substantial evidence for H_1, that adults' performance at fill-in-the-blanks task positively correlated with their performance at legality judgment task, $BF_{(0,0.58)}$ > 10,000, *RR* [0, >4.59] ($r(39) = 0.88$, $p < .001$).

6.3.2.4 Associations between word learning and graphotactic learning

As in Experiment 6, we found ambiguous evidence for a positive correlation between performance at word learning and both fill-in-the-blanks ($BF_{(0,0.29)}$ = 1.19, *RR* [0, 1.35] ($r(39) = 0.16$, $p = .30$)) and legality judgment task ($BF_{(0,0.29)}$ = 1.09, *RR* [0, 1.22] ($r(39) = 0.15$, $p = .34$)), Testing for a negative correlation, we found conclusive evidence for the null, that is, there was no negative correlation between performance at word learning and fill-in-the-blanks ($BF_{(0,0.29)}$ = 0.27, *RR* [0.22, >4.59]; $r(39) = 0.16$, $p = .30$)) or legality judgment task, $BF_{(0,0.29)}$ = 0.28, *RR* [0.23, >4.59] ($r(39) = 0.15$, $p = .34$)).

6.3.2.5 Comparison between Experiment 6 and Experiment 7

Word Learning Task. Figure 6.6 shows the mean proportion of adults' correct responses in Experiment 6 and 7, in the word learning task over the six blocks. We found substantial evidence for H_1—that is, higher performance in Experiment 6 than Experiment 7, $BF_{(0, 1.05)}$ = 2421 *RR* [0.02, >4.59] (effect of experiment: β = 0.44, *SE* = 0.10, *z* = 4.33, *p* < .001). That is, adults were overall better at selecting the correct label for the objects in Experiment 6 (without explicit instructions) compared to Experiment 7 (with explicit instructions).

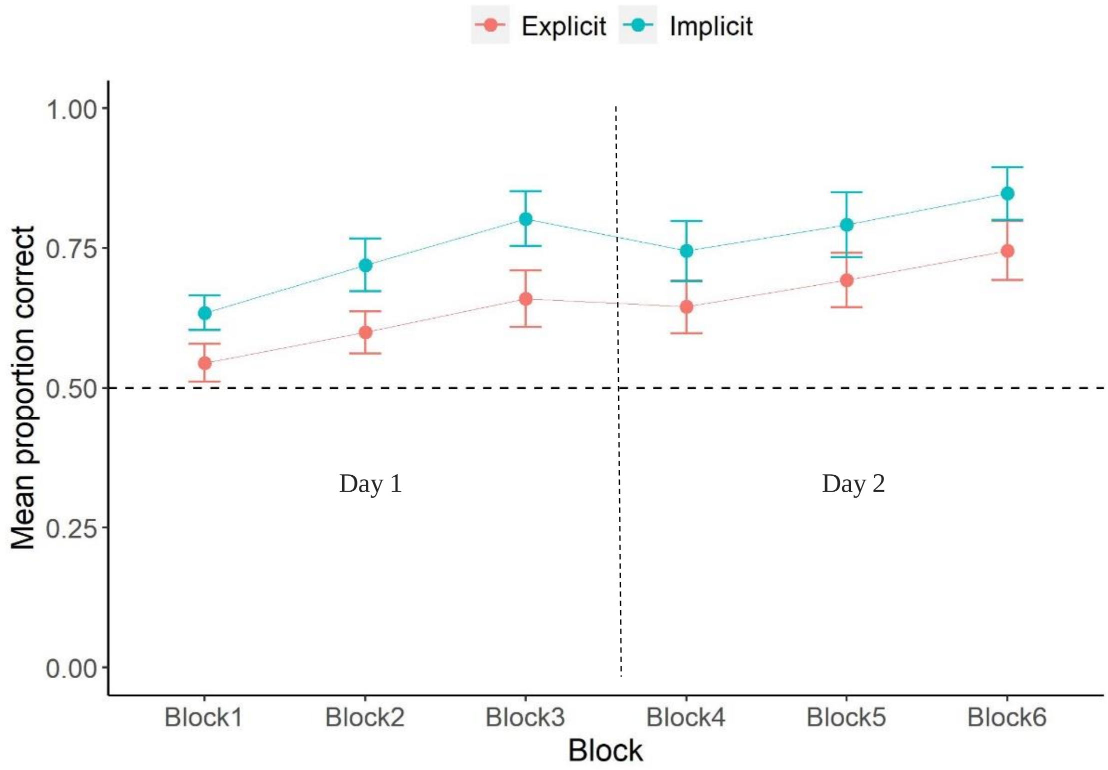


Figure 6.6: Adults' Mean Accuracy (Line Graph With 95% Confidence Intervals) in the Word Learning Task in Experiment 6 (Implicit) and 7 (Explicit) by Block. The Dashed Line Represents Chance-level Performance (50%).

Test Tasks. The mean proportion of adults' correct responses in the fill-in-the-blanks and legality judgment tasks in Experiment 6 are shown in Section 6.2.2.2, Figure 6.3, and those in Experiment 7 are shown in Section 6.3.2.2, Figure 6.5. In the fill-in-the-blanks task, we found substantial evidence for H_1 that is, greater performance in Experiment 7 than Experiment 6, $BF_{(0,\ 0.50)}$ = 19.71 *RR* [0.94, >4.59] (effect of experiment: β = 0.58, *SE* = 0.21, *z* = -2.72, *p* = .01). Adults were worse at selecting the correct vowel to create legal strings in Experiment 6 (without explicit instructions) compared to Experiment 7 (with explicit instructions). Similarly, in the legality judgment task, H_1 was supported by substantial evidence, $BF_{(0,\ 0.46)}$ = 9.44 *RR* [0.12, 2.88] (effect of experiment: β = 0.56, *SE* = 0.23, *z* = 2.45, *p* = .01). Adults were worse at correctly discriminating between legal and illegal items in Experiment 6, compared to Experiment 7.

6.3.2.6 Awareness Questionnaire

The coding was identical to Experiment 6. When asked to report whether they had awareness of the patterns, four participants selected “no”, three selected “I don’t know” and thirty-four selected “yes”. Of the ones who reported awareness, eleven were unable to describe the rules and reported that they were mostly guessing (e.g., “I was trying to remember the rules, but I mostly went by gut instinct”), ten provided vague answers (e.g., “tried to keep it similar to the words in the game”), while the remaining thirteen were able to accurately describe at least one aspect of the manipulation, such as “there is double f or s when it follows e”. This data provides an indication that, although informed about the rules governing the patterns, only 32% of the participants used this information to make their decisions at test.

6.3.3 Discussion

To sum up the results in Experiment 7, as in Experiment 6, following explicit instruction about the patterns embedded in the nonword stimuli, adults again learned both the object-word associations in the exposure task (with above-chance accuracy and their performance increasing over the six blocks) and the graphotactic patterns (with better-than-chance accuracy when selecting the vowel to create legal stimuli in the fill-in-the-blanks task, and when discriminating between legal and illegal strings in the legality judgment task). As in Experiment 6, although performance in the two tests of graphotactic learning was positively correlated, there was no evidence for a similar positive correlation between performance in either of these tasks and word learning. However when we looked for a negative correlation, there was evidence for the null, that is, evidence that learning one aspect of the stimuli did not impede each other. We then compared the performance in both (exposure) word learning and (test) graphotactic learning tasks in Experiment 6 (implicit learning) and the current experiment (explicit instruction) and found that, while adults were better at learning the graphotactic constraints when they received explicit instruction (Experiment 7), that is, they were better at both selecting the correct vowel to form legal items and judging the legality of the novel strings, in this experiment, we saw that they were worse in the word learning task. I return to discuss implications for these results in the following, general discussion of this chapter.

6.4 General Discussion for Experiment 6 and Experiment 7

In two learning experiments, we assessed adults' ability to pick up graphotactic constraints under more naturalistic conditions, rather than attending to the form of words in isolation. Across experiments, learning was induced—over two brief sessions—by exposure to pattern-embedding stimuli incorporated into a word learning task. Here, participants were presented with pictures of unfamiliar objects and were asked to identify the correct label from a choice of two pattern-embedding nonword strings. Learning of associations was induced via feedback, and performance (i.e., trial-by-trial accuracy) during this task was collected as a measure of word learning. Upon training completion (session 2), post-tests identical to those in our previous experiments, were used to measure whether participants had learned the patterns and could generalize over them in a production (fill-in-the-blanks) and legality judgment task. The patterns embedded in the stimuli were the same as in Experiment 4, that is, specific word-middle vowels predicted specific word-ending singlet/double consonants. Learning was induced either incidentally (no instruction to learn the patterns; Experiment 6) or explicitly (explicit rule teaching followed by the same word learning task; Experiment 7).

The key result of the current study was that adults were shown to be able to concurrently learn both word meaning mappings and graphotactic patterns in a task targeting word learning. Following successful selection of a novel word to match a novel object, they consistently generalized over the test stimuli: They produced permissible spellings by selecting allowable word-medial vowels and discriminated between permissible and impermissible strings with better-than-chance accuracy. We also saw that, although this concurrent learning of word meanings and spelling occurred under implicit conditions (Experiment 6), as in the previously reported experiments, adults who were explicitly instructed about the nature of the embedded patterns (Experiment 7) did better in the graphotactic learning task, than those who performed the same tasks under implicit conditions. Interestingly, however, they did worse at the (exposure) word learning task. We discuss each of these findings in turn.

6.4.1 Concurrent learning of word meanings and graphotactics

The fact that we saw substantial evidence for above-chance word learning in these experiments indicates that we were successful in developing a paradigm where word learning occurred. This is notable, since we are using a particular type of stimuli which share features (i.e., three or four letters with final consonants that double half of the time), that are harder to distinguish from each other. Before carrying out the current studies with children, a pilot study is needed to reveal whether the difficulty level is too high and find an alternative, such as reducing the number of objects to be learned, therefore reducing the number of stimuli. This, of course, would have an effect on learning of graphotactics.

Critically, despite being engaged in a task focused on word learning, adults picked up on graphotactic patterns. This evidence, from a well-controlled learning study, contributes to the existing literature that tested adults' sensitivity to graphotactic doubling patterns in their own language (e.g., Pacton et al., 2014), suggesting that the knowledge tapped in these studies was gained implicitly, from incidental exposure when reading in a meaningful context. Unfortunately, due to differences in the amount of total exposure, and potential learning from foil stimuli in this new paradigm, we were not able to conclusively determine whether the effects we found when the word learning task was added to the exposure task (Experiment 6), were less than the ones seen when the focus was solely on form (Experiment 4) (see Section 6.2.3 for a discussion). Future work could change the paradigm, so that two pictures of novel objects appear on the screen and only one nonword, requiring that participants select one of the images that correspond to the word. However, we did find evidence from the correlations, that word learning and graphotatctic leaning were not hindering each other: We tested both positive (i.e., generally good learners will do well in both types of tasks) and negative (good performance in one type of task impedes performance in the other) correlations and found substantial evidence for the null for negative correlations. Therefore, we can conclude that learning in the two types of tasks did not impede each other.

6.4.2 Implicit and Explicit Learning

As expected, based on results from Experiment 5 with children, we did find an advantage for explicit instruction, with better performance in both test tasks compared to the implicit learning. This shows that this type of instruction is useful even when learners are then engaged in a meaningful task, further establishing the relevance to more naturalistic learning. An interesting finding in this chapter was that the benefit reversed in the word learning task. This was presumably because, explicitly telling participants about the rule, has led them to focus on spelling, at the expense of learning the meanings. Interestingly, when analyzing the responses to the awareness questionnaire, many (68.3%) of the participants in the explicit experiment were still unable to report any aspect of the patterns learned (none in the implicit condition could do so), many acknowledging that they eventually used their gut feel when choosing the letter to form words in Zorib language or decide whether a word was in that language. This also reflects the fact that their group performance was not at ceiling. Nevertheless, the better performance in this experiment may have also been driven by the fact that participants were aware of the existence of patterns at the start, even though they could not verbalize the knowledge, and this is reflected in their lower performance in word learning, presumably due to the additional task of searching for the taught patterns. Conversely, when unaware of the existence of patterns, participants still learned them, but they carried out the word learning task without the additional cognitive load. The explanation in terms of searching behavior, which is not necessarily linked to better learning of word meanings, is consistent with the fact that, as in the implicit condition, we also see evidence that better word learning did *not* lead to worse graphotactic learning.

These results differ from those of Ormrod's (1986), who (as discussed in Section 6.2.3) found that learning of spellings improved when participants were explicitly asked to focus on learning the spellings encountered in a meaningful context, but their comprehension remained unchanged. One possibility for the finding, in our studies—that knowledge of graphotactic patterns had a negative impact on the ability to learn the object-word associations—is that this type of very active task of selecting words for novel images is affected more by the diverted focus to the graphotactic patterns. Another reason could be that, in our studies, the measure of word learning is more sensitive than the comprehension questions in Ormrod's study, and therefore, we picked up the difference that was otherwise unnoticed. Finally, note that the previous studies do not use Bayes factors, so the lack of effect there may not be evidence for the null.

Is a detrimental effect on learning under explicit conditions found elsewhere in the literature? The closest is Reber et al.'s (1991) claim that searching for complex grammatical rules and linguistic patterns impedes learning of those same patterns. The results here are not consistent with this, in that the graphotactic patterns themselves, even if hard to verbalize and remember, were learned with better than chance accuracy in the explicit condition and better than in the implicit condition. However, our findings confirm that searching for the patterns did indeed impede one type of learning—that is, the learning of meanings, which was the focus of the task.

7 General Discussion

This thesis presented a series of experiments that tested developing, as well as skilled spellers' ability to generalize over novel graphotactic patterns under incidental learning conditions. Statistical learning research has demonstrated humans' ability to pick up on patterns from the environment over time (Schapiro & Turk-Browne, 2015), without intention to learn and without awareness of what is learned (Reber, 1967, 1989). Statistical learning mechanisms have been shown to operate across multiple sensory modalities (e.g., auditory, visual, and tactile) and domains (e.g., language, music, vision and movement, among others; Armstrong et al., 2017) from as early as infancy. These processes have been implicated in generalizations over both sequentially presented patterns (within spoken stimuli, Saffran et al. 1996; within visual stimuli; Fiser & Aslin, 2001) as well as simultaneously presented visual (spatial) patterns (Fiser & Aslin, 2001), and are thought to play a key role in child first language acquisition (Wonnacott, 2013). Their role in literacy acquisition has been previously speculated (e.g., Kessler et al., 2013; Treiman, Kessler, Boland, Clocksin, & Chen, 2017) and more recently demonstrated in the work of Samara and colleagues (Samara & Caravolas, 2014; Samara et al., 2019). The studies presented in this thesis follow up on this work, to further probe the ability of children and adults to learn different types of spelling patterns. Certain patterns and variation found in written languages are systematically distinct from those in spoken languages (Coulmas et al., 1983) and are shaped by multiple evolutionary pressures, such as the trade-off of representing different aspects of spoken language, chance events and human cognition. This explains the existence of purely visual patterns—pure graphotactics (explained by frequency of graphemes and their probability of occurrence and co-occurrence in a certain context)—that humans learn without support from knowledge of spoken language. Evidence of children's and adults' sensitivity to graphotactics comes from studies that looked at experience with natural languages (Kessler, 2009; Pollo et al., 2007; Steffler, 2001; Treiman & Kessler, 2013), as well as studies that used well controlled experimental conditions (Chetail, 2017; Nigro et al., 2015, 2016; Samara & Caravolas, 2014; Samara et al., 2019). Because questions of learnability are hard to address in studies that test sensitivity to naturalistically acquired patterns in participants' own language, and because the graphotactic stimuli used in the few existing learning experiments were confounded by correlated phonotactics (i.e., probabilistic constraints on sound sequences), this thesis goes beyond previous work by assessing purely graphotactic learning that cannot be explained by phonotactic sensitivity. Prior to this thesis, only two studies have controlled for this confound but the experiments either featured only adults (Chetail, 2017) or very simple (positional) constraints (Nigro et al., 2016). In the experiments reported in this thesis, sensitivity to positional, as well as the more complex, context-based patterns, was tested with developing spellers (who are the relevant population for a study concerned with literacy), alongside skilled spellers (i.e., adult population, for whom the process is ongoing).

Apart from the specific purposes stated above, another question addressed in the current thesis concerned the extent to which differences in the ability measured in our experiments relate to literacy measures, the latter measured mainly by performance on the WRAT Reading and Spelling subtests and TOWRE word and nonword subtests. This link has been widely hypothesized and is important in establishing whether implicit learning mechanisms underpin spelling development. However, there have been conflicting results in the literature (e.g., Arciuli & Simpson, 2012; Frost et al., 2013; Nigro et al., 2015; Schmalz et al., 2019; West et al., 2017). Unlike in spoken language, explicit instruction plays an important role in learning the structure of written language and starts when children enter formal education. A final question explored in this thesis was whether explicit instruction in spelling rules may be beneficial over and above implicit learning (Bosman et al., 2006; Butyniec-Thomas & Woloshyn, 1997; Kemp & Bryant, 2003; Nunes et al., 2003). To explore this potential benefit, several of the studies reported in the current thesis include an explicit condition. In the next four sections, I present a summary of results for each of these broad areas, starting with the key findings of this thesis: (i) Graphotactic learning across experiments, (ii) implicit versus explicit learning, and (iii) associations between performance in graphotactic learning and literacy; followed by another factor explored across the thesis: (iv) age effects. I then turn to discuss the (v) theoretical, methodological, and practical implications of these results, and finally, (vi) future directions.

7.1 Evidence of graphotactic learning across the experiments

Previous group learning effects have been small (e.g., children were 3% above chance (50%) level in Samara et al., 2019) and this may have been a reason for the lack of significant correlations with literacy attainment (although this link has not been previously explored using Bayes factor analyses). Therefore, a starting point in the thesis (**Chapter 3**) was to adapt the methodology used in Samara and colleagues' studies (Samara & Caravolas, 2014; Samara et al., 2019) with the aim of finding stronger graphotactic (or phono-graphotactic) learning effects. We employed Bayes factor analyses to evaluate evidence for and, crucially, against hypotheses (as we also did for evaluating performance in our learning tasks). In **Experiment 1**, we took the visual stimuli in Samara and Caravolas (2014)—that is, pronounceable written English nonwords where the graphotactic cues were fully correlated with phonotactic cues—and supplemented these with audio stimuli so that the phonology was added overtly, through the pronunciation of a Southern English female speaker. In addition, the exposure task was modified to encourage participants to focus on the form of the entire "word" stimulus rather than any irrelevant features (such as color), and exposure was extended over two sessions (as in Samara et al., 2019) rather than a single session (as in Samara & Caravolas, 2014). A new task was devised for measuring sensitivity to pattern-embedding stimuli, to simulate more closely what children and adults do in naturalistic situations when producing spellings (I discuss the implications of this methodological contribution below, in Section 7.5.2). These adaptations, particularly the additional phonotactic cue, were expected to boost learning. In fact, this boost was rather small: While both children and adults showed sensitivity and generalization in both tasks, the group level performance increased by 4 and 9 percentage points for children and adults, respectively, compared to the same test in Samara and Caravolas (2014), when the phonology was not provided (but assumed covert); and only for adults there was substantial evidence that this difference was not due to chance. Moreover, the evidence for most of the correlations between performance at the graphotactic learning tests and literacy continued to be ambiguous, even though we had hoped to achieve conclusive evidence with the new paradigm (correlations are discussed further in Section 7.3 below). For this reason, we did not continue with the goal of searching for stronger effects, although the changes to exposure session and novel test were maintained for the rest of the thesis.

Although various cues were manipulated, one possibility is that the greater learning may have been due to the presence of the auditory input leading to cross-modal facilitation of the graphotactic patterns from equivalent and supporting phonotactic patterns (Robinson & Sloutsky, 2007). This led us to speculate whether participants in Samara and Caravolas (2014) covertly accessed the phonology or whether learning in that study was purely graphotactic without phonotactic support. This was suggested by the authors, on the basis that pronunciation was not encouraged or necessary in order to complete any of the tasks. However, it could not be determined if this was true based on evidence so far. A new approach was needed to address the key question of whether "purely" graphotactic learning can occur in children and adults.

One approach to removing cues from (covert) phonology was to use an artificial script. Previous studies have demonstrated that adults can learn visual patterns from nonlinguistic input presented sequentially (Fiser & Aslin, 2001, 2002), as well as from input mirroring the structure of written words (Chetail, 2017; Nigro et al., 2016). However, only one of these studies tested children (Nigro et al., 2016), and in this study, the stimuli they used were abstract shapes (same as in Fiser & Aslin, 2001) rather than unfamiliar orthography or orthography-like novel stimuli that have been used in artificial orthography studies with adults (Chetail, 2017; J. Taylor et al., 2011). This meant that inferences could be made regarding visual learning but the relevance to linguistic specific learning was less clear. In **Chapter 4** (**Experiment 2**), word-like stimuli were created by combining real English letters (consonants) with grapheme-like symbols (from J. Taylor et al., 2011) that are closer in features to the graphemes in written words (C. Vidal et al., 2017). The resulting semiartificial three-character strings retained the appearance of linguistic visual input but were unpronounceable, that is, the phonotactic confound was removed. Although it is possible to embed contextual constraints in these strings, since this is a first study of this type with children, this study manipulated the possible position of the symbols at the edges of words (as in Samara & Caravolas (2014) stimuli with positional constraints). By using real letters, the stimuli were less unfamiliar, and thus requiring no additional pre-training (as in Chetail, 2017) and, in addition, allowed for further manipulation: By adding English consonants not included at exposure, generalization was tested not only using recombinations of the graphemes used in training, but also of consonants familiar from English text exposure but not used in the experiment.

In line with Samara and Caravolas (2014) and Nigro et al. (2016), Experiment 2 demonstrated that, even when the phonology was not present, available or helpful, children and adults learned the positional graphotactic constraints embedded in semiartificial stimuli and they generalized these across other untrained graphemes as well. Our results go beyond previous research to show that developing spellers are sensitive to visual graphotactics before they are explicitly aware of the existence of patterns beyond letter-sound correspondences, in line with evidence from studies that show their sensitivity to patterns governing doublet use in natural language stimuli (Cassar & Treiman, 1997; Lehtonen & Bryant, 2005).

The group effects seen in Experiment 2 were stronger than those in Experiment 1 despite the use of unfamiliar stimuli and absence of the additional cue provided by phonotactics: e.g., children discriminated between legal and illegal items with 59%, and adults with 71% accuracy (with trained consonants). They were, however, similar to those found by Nigro et al. (2016): 60% for children. Notably, both Nigro et al. (2016) and Experiment 2 tested *positional* rather than *contextual* constraints. As Samara and Caravolas (2014) demonstrated, positional constraints are less complex and easier to learn, due to their unconditional nature (e.g., a character can or cannot appear at the beginning of a string). Thus, Experiment 2 leaves open the important question of whether children are also able to learn graphotactic patterns conditioned by *context*, when these have no phonotactic counterpart. We decided to move to a new paradigm to test this, believing that learning of contextual constraints with stimuli formed with unfamiliar symbols, even if in semiartificial strings, would require longer training (as has been done in spoken artificial or semiartificial language: Hudson Kam & Newport, 2009; Samara et al., 2017) which was not feasible in the time constraints of this thesis.

In **Chapter 5**, the graphotactic sensitivity to the more complex context-based conditions was assessed, but rather than with unfamiliar symbols, stimuli were created with the use of alphabet letters, as in previous studies. However, the purely graphotactic rules were created by using single and double English consonants at the end of CVC/C strings, to form homophones (e.g., *dd* and *d* map onto the same sound), thus retaining the familiarity of letters while removing the phonological confounds. The pattern complexity was also manipulated to produce one relatively simple rule and a second, more complex one: In **Experiment 3** all consonants always doubled following one of two word-medial vowels and never following the other (a “simple” pattern, which is also easier to verbalize); whereas in Experiment 4 some consonants doubled with one vowel (but not with the other) while the opposite was true for the other consonants (“complex” pattern, where doubling per se was not the rule). The studies in this chapter provided clear evidence that both types of novel context-based graphotactic constraints can be learned by both children and adults, via implicit learning processes. This evidence shows that lab-based studies can capture the acquisition of knowledge of complex statistical patterns that cannot be verbalized (a point that I return to below, in Section 7.2, where I discuss the difference between implicit and explicit learning), akin to those which are abundant in natural languages (for English patterns see Kessler, 2009).

The studies in Chapter 4 and Chapter 5 tested learning of both unconditional, positional constraints embedded in semiartificial strings, and constraints conditioned by context in homophonic nonwords under incidental conditions, and are a clear demonstration of the role of statistical learning mechanisms in graphotactic learning, adding to the literature that has speculated about a role for this in literacy (e.g., Cassar & Treiman, 1997; Danjon & Pacton, 2009; Pacton et al., 2001). An important endeavor in the current thesis was to monitor and assess the real-life validity of the measures used. The well-controlled experimental conditions used so far, however, present a drawback in that they have provided participants with a single source of information to be processed, namely the graphotactic information (“form”). In natural languages, however, most encounters with written words occur in meaningful context. In **Chapter 6**, the generalizability of laboratory research was addressed by adapting the methodology used throughout the thesis, to assess adults’ ability to pick up graphotactic constraints implicitly (**Experiment 6**), under conditions where participants are learning the meaning of novel words via exposure to written nonwords along with associated unfamiliar pictures. To this end, we devised a game focused on learning the meaning of words where participants received trial by trial feedback. Since the new paradigm introduced an additional word-learning task, which was the focus of learning, the graphotactic patterns were learned in a more indirect way, with less attention to form. The written stimuli were identical to those in Experiment 4, that is, the more complex, contextual patterns. Due to the added difficulty, and the fact that this was a new methodology, this study was carried out only with adults. In addition to looking at learning of the graphotactic patterns with above-chance performance in each of the test tasks (fill-in-the-blanks and legality judgment), in this chapter we also looked at performance in the word learning task—that is, learning of the association between the orthographic form of nonwords and the associated pictures, to see if both could be learned in tandem. Performance in word learning was assessed from performance in the word-learning training task used in exposure and we also looked at improvement over time, as more trial-by-trial feedback was provided (across the six blocks spread out over two consecutive days). The key result was that adults were able to concurrently learn both word-meaning mappings and the graphotactic patterns and the two types of learning did not impede each other: Correlational analyses revealed that those who were good at learning the spellings were *not* worse at learning the meanings (evidence for the null for a negative correlation), suggesting that, as in natural languages, when reading for meaning, orthographic processing can proceed in parallel. We didn't find evidence that learning graphotactics in an indirect way (Experiment 6) affects performance seen when focus was fully on form (Experiment 4), in line with research suggesting that meaningful context may not play a role in learning orthographic patterns (Nation et al., 2007); however the evidence here was inconclusive, that is, no evidence for the null.

It is important to consider whether there are limitations in terms of aspects of the stimuli used across the experiments that could cast doubt on the fact that participants learned novel graphotactic rules on the basis of the exposure in the experiments. An important goal in artificial language experiments is to create novel artificial patterns and control for prior experience. As noted in the chapters, while we attempted to do this across the experiments, except where we used semiartificial symbols (Experiment 2), this could not always be achieved. In Experiments 3 to 7, for example, where target rules used doublets, there was a potential confound given that some consonants are more frequently doubled than others, and occur with certain vowel in single or double form with varying frequency in English. However, we designed the stimuli such that none of the consonants were (i) illegal as doublets, (ii) illegal as singlets or doublets at the end of a word, and (ii) did not co-occur with any of the two vowels in English. Another confound was that, inevitably, in all but Experiment 2, some stimuli were real words, as noted in each *methods* section. The approach taken to deal with all of these potential confounds was to use lists and counterbalance across participants which patterns are legal and illegal items, limiting the possibility that such effects could have accounted for the learning effects. This solution removes the need for a control condition to find if participants are indeed at chance at pre-test or whether they start the experiments with biases.

There has been a debate in the literature about whether the knowledge measured in statistical learning tasks is better described as rule or chunks, with some suggesting that statistical computations are employed prior to– and utilized to infer chunk formation (e.g., Adini, Bonneh, Komm, Deutsch, & Israeli, 2015), while others propose that relevant units are discovered using different strategies (e.g., Orbán, Fiser, Aslin, & Lengyel, 2008). Therefore, I now turn to consider this question with respect to the current data. The artificial grammar learning literature (Reber, 1967; see also Section 1.5 for more detail) has demonstrated that performance in implicit learning tasks is driven by more than one property of the stimulus, and these can be highly correlated (Perruchet, 2019). For example, abstract representation of "rules" in the current experiments could be formed from either sensitivity to chunk information (e.g., the letter *f* is associated with *o* to form a chunk *fo*) or transitional probabilities (e.g., *f* can begin a word if followed by *o*). Thus, learning of context-based patterns in this thesis may be described as knowledge of bigram constraints or restrictions on letter co-occurrences (see also Pacton, Fayol, & Perruchet, 2002). While the majority of the experiments in this thesis cannot differentiate these possibilities, one place where there is potentially relevant data is in the comparison between Experiment 3 and Experiment 4: Because the patterns in Experiment 3 are more complex (i.e., all consonants always doubled following one of two word-medial vowels) compared to those in Experiment 4 (i.e., some consonants doubled with one vowel, but not with the other), whereas the bigram/trigram chunks were equal in number (eight in each experiment), if performance was better in Experiment 3, this would mean that participants learned a qualitatively different type of rule which is at a general level about "doubles" (e.g., "vowel-x occurs before singlets and vowel-y occurs before doublets"), compared to a more specific one (Experiment 4), which couldn't have been learned as bigram chunk. The evidence for such difference, however, was inconclusive (except in fill-in-the-blanks test with adults), and therefore, any conclusions for literacy development remain tentative.

In sum, our key investigation in this thesis revealed that visual statistical learning processes underlie the learning and generalization of purely graphotactic *positional* as well as *context-based* spelling patterns in children as young as 7 years old and adults, controlling for the possibility that statistical learning accounts for children's sensitivity to spoken language patterns. We demonstrated this sensitivity even under meaningful conditions and success in learning graphotactics was not hindered by success in learning word meanings.

7.2 Learning under implicit versus explicit conditions.

A goal of the experiments in this thesis was to measure learning under incidental implicit learning conditions, similar to that occurring naturalistically from text exposure. In Experiments 1 to 4 and 6, we created conditions for this type of learning and the evidence from the questionnaires suggested that this was largely successful: No children in any implicit experiment were able to report the spelling rules, and adults could only do so with the simple patterns in Experiment 3 (although we showed that even with those participants removed, learning was above chance). However, an important question is the extent to which such questionnaires are able to capture explicit knowledge: Subjective reports of awareness are often unreliable (Batterink et al., 2015) and, while they reveal participants' level of awareness, defining and measuring the true level of awareness is a complex problem (Cleeremans, Destrebecqz, & Boyer, 1998). Furthermore, the use of offline tasks to measure implicit learning has been scrutinized since the post-exposure tasks require participants to remember what is learned when blind to the patterned nature of the stimuli, and then make explicit judgments on novel stimuli. With this in mind, learning was taken to be implicit to the extent that it occurred under incidental conditions that did not promote explicit "rule" searching strategies (see also, Cleeremans et al., 1998; Saffran et al., 1997). While it is not fully clear and cannot say with certainty that participants do not have explicit knowledge of the manipulated patterns, at least for adults, we have evidence that performance was higher when they could verbalize the patterns that is, in Experiment 3. Other artificial language experiments also found similar qualitatively different behavior in children who could/could not report explicit patterns (Samara et al., 2017).

Many aspects of written language cannot be explained by rules that are easy to verbalize, or at least not rules easily taught to children at the start of formal education (Kessler, 2009). Nevertheless, much of the literacy acquisition takes place in the classroom. While the current thesis demonstrates that implicit learning processes may underpin spelling, it also speaks to the importance of addressing the question of how the two routes to learning (explicit instruction and implicit learning through exposure to print) contribute to proficiency in spelling and what their differential effectiveness is. Little work to date has directly investigated whether explicit instruction is more effective over and above implicit learning using the same paradigm and stimuli (Bosman et al., 2006; de Bree et al., 2018; Kemper et al., 2012). This thesis aimed to fill this gap and carried out a comparison between strictly comparable incidental and explicit conditions. In **Experiment 5 (in Chapter 5)** and **Experiment 7 (in Chapter 6)**, all the aspects of the equivalent implicit learning experiments (Experiment 3 and Experiment 6, respectively) were replicated and, in addition, explicit instructions as to the spelling patterns were provided prior to the print exposure. Experiment 5 tested the relative benefit of explicit instruction for relatively “simple” context-based patterns and Experiment 7 tested the same effect in a different (word learning task) for more complex context-based patterns. These studies revealed that being taught spelling patterns explicitly was advantageous (see Section 5.1.2 for an in-depth discussion) over and above incidental exposure. Further evidence for an advantage of explicit knowledge comes from analyses of awareness questionnaire in Experiment 3 with adults, where those who reported explicit awareness of the embedded patterns (even though they were not instructed about them) showed ceiling performance. It is notable that explicit instruction benefit was found even when the participants were engaged in the meaningful word-learning task in Experiment 7. Interestingly, the benefit was reversed for word learning itself, that is, there was weaker learning of the word picture associations in the explicit than the implicit condition. However, in both the implicit and explicit conditions there was evidence for the null, that is, for a negative correlation between the word learning and graphotactic tasks, indicating that being better at learning the graphotactic patterns did not make the participants worse at word learning. We saw this despite the fact that group performance at graphotactic learning was better than word learning in the explicit condition. One possibility for this difference is that the explicit instructions led to an additional attentional load due to searching for the patterns which had been disclosed at the start of the task (although active search was not encouraged), and this impeded the task of association of words with meanings. Returning to the learning of the graphotactics themselves, the fact that we saw stronger rather than weaker learning in the explicit condition is contrary to Reber's (e.g., Reber et al., 1991) notable argument, that deliberate effort to learn complex linguistic patterns impedes learning of these patterns. However, our rules are arguably not as difficult to verbalize and explain as those found in natural languages. Further experiments could explore whether different results are found if more complex naturalistic rules are used. Another way to explore this further would be to run the same experiment, using our more complex rules, with children, who would likely find the explanation of the rule harder. While the benefit of explicit learning of graphotactic patterns was demonstrated for short term retention, that is, in tests that followed immediately after exposure, this benefit might not hold for long term retention, and this could also be explored in an experiment using our stimuli and paradigm and delayed graphotactic learning tests. Although our studies show that explicit learning is stronger than implicit learning, they do not tell us which is more important for literacy, and this is addressed in the next section.

7.3 Associations between learning and literacy

The final goal in this thesis was a correlational one, to establish links between literacy acquisition and both statistical and explicit learning. As noted in Chapter 3 (Section 3.1.1), this investigation has been a topic of debate in the literature, with issues regarding the reliability of measures used and generally small effects. The hypothesized positive correlations were never observed in our studies measuring learning under incidental conditions. However, this lack of correlation is not surprising, given that the effect sizes across the experiments in this thesis were small, despite attempts to increase the amount of learning in Experiment 1. In the implicit condition across this thesis, children's performance ranged between 54% and 60%, and adults' between 53% and 67%. Some of the relevant correlations in each implicit experiment were inconclusive (i.e., are neither evidence for H_1 or for H_0), and, when the data was conclusive, it was against H_1 (i.e., variations in learning skill were unrelated to variations in reading and spelling ability). Interestingly, however, under explicit conditions, the evidence supported the hypothesis: There was a positive association between graphotactic learning when the patterns were explicitly revealed at the start of Experiment 5, and literacy performance.

For implicit learning, should we take the null findings to suggest that statistical learning skill is unrelated to literacy performance? While this is a possibility, they may instead suggest that correlation is not a sensitive way of establishing associations with the type of measures we employed throughout this thesis. As noted in the section discussing the issues with measures and correlations (Section 1.7.2), reliability, among other statistical properties, are essential. In the current thesis, however, the counterbalancing between lists and items meant that the items were not designed to measure sensitivity to the same type of knowledge. Such experimental control to maximize a tasks' internal consistency (as in Gebauer & Mackintosh, 2007) would come at the cost of our ability to measure generalizations in well-controlled conditions. It may be that, when correlations are found in the literature, they come from studies that can measure one type of knowledge, such as the serial reaction time task studies such as Howard, Howard, Japikse, and Eden (2006). Therefore, as discussed in Section 5.5.3, the null findings reported with correlations must be interpreted with caution.

Such mixed results and conclusions leave an important challenge for the field: What adaptations can be done to see acceptable levels of reliability in implicit learning studies, to capture individual differences in developing populations? I have discussed that some solutions (such as longer training duration, more complex patterns and unfamiliar stimuli, among a few) are not feasible with children, who have limited attention span and capacity to understand complex instructions. Recent calls for devising online tasks (e.g., reaction-time based tasks and neural measures such as event-related potentials, ERPs) to measure the course of perceptual detection of regularities are welcome (Krishnan & Watkins, 2019; Siegelman et al., 2017), to complement the offline recall tasks. They are more indirect (and possibly more sensitive) measures of tacit knowledge (Batterink, Cheng, & Paller, 2016; Siegelman, Bogaerts, Kronenfeld, & Frost, 2018). A recent attempt by Kuppuraj et al. (2018), however, did not find correlations between their measures and verbal short-term memory, despite high (0.67) task reliability. Nevertheless, such attempts are necessary in the area of literacy acquisition.

Despite the limitations and inconclusive evidence for implicit learning, the positive associations revealed in the explicit experimental condition strengthen the view that explicit learning contributes to learning to read and spell (see Section 5.5.3 for further discussion). It is, however, possible, that the children who were skilled to capture the explicit rules were also skilled to capture the graphotactic patterns and knowledge of rules had no direct effect of reading and spelling efficiency per se.

7.4 Age effects

Although age was not a central question in this thesis, most of the experiments were carried out with both children and adults and thus, our data can potentially provide evidence on how different age groups learn, given matched text exposure. Such comparison cannot be easily made by looking at adults' and children's reading and spelling performance using their natural language, since their age is confounded with text experience. By and large, in the two studies (Experiment 1 and Experiment 2) where age group performance was compared, adults were superior learners than children, a finding that goes against Reber's (1993) claim that implicit learning is age-invariant and contradicts previous statistical learning studies showing equivalent levels of performance between different age populations (e.g., Saffran et al., 1997). For naturalistic spelling, there is evidence that the level of complexity of the spelling patterns may affect the extent to which age effects are shown (Cassar & Treiman, 1997). It is possible, however, that in naturalistic context, the reason for an age effect for complex patterns is that those take longer to learn and adults have had more text exposure. In contrast, in controlled experimental conditions such as those in this thesis, the text exposure is matched across age groups. In the current thesis, both children and adults were exposed to unconditional, positional constraints as well as context-conditioned constraints, and with different types of materials, and the age effect was found in all cases. Furthermore, both children and adults found the positional constraints learning easier to learn (59% (children) and 71% (adults) group performance in Experiment 2) than contextual constraints (54% (children) and 56% (adults) group performance in Experiment 4). It was suggested earlier in the thesis that this demonstration is consistent with previous developmental studies on the acquisition of language-wide spelling-to-sound conditional patterns (e.g., Treiman & Kessler, 2006). This thesis, therefore, suggests that, while adults and children show similar patterns of learning, adults learn faster from matched text. This is in line with evidence from language learning experiments that used matched input for spoken language (second language or artificial language) and classroom based second language studies which generally found that older learners learn more in the short term, as revealed in Huang's (2016) review of 42 studies.

7.5 Implications

7.5.1 Theoretical Implications

As explained in Chapter 1 (Section 1.4), early theories of spelling development suggested that the processes involved in learning to spell follow in stage-like manner and progression to one stage depends on the mastery of the previous one (Frith, 1985; Gentry, 1982). The stages are described as distinct, that is, qualitatively different. These theories have been interpreted to suggest that knowledge of orthographic conventions emerges only in the very last, orthographic stage (e.g., Pollo et al., 2007). However, accounts of children's sensitivity to untaught positional restrictions (e.g., *ck* does not occur in syllable-initial positions in English) and orthographic constraints on allowable doublets (e.g., *hh* never occurs in English), have been shown from the very beginning of literacy. Accumulating evidence has challenged early theories' views and highlighted that children incorporate multiple sources of information (e.g., orthographic and morphological) when applying their knowledge in writing (e.g., Cunningham et al., 2001; Deacon & Kirby, 2004). It has been hypothesized that children succeed in applying complex orthographic knowledge into their own spelling via implicit learning (Pacton et al., 2001, 2013; Pollo et al., 2007) and this was recently demonstrated by Samara and colleagues (Samara & Caravolas, 2014; Samara et al., 2019) and the current thesis. This work shows that 7–year-olds have statistical learning abilities that allow them to avoid orthographical errors in the absence of explicit instruction, thus playing an important role in literacy development. However, prior to this thesis, little previous research had investigated whether and how do children learn patterns that do not reflect spoken language patterns (i.e., are purely visual with no sound based counterpart; e.g., *dd* does not begin written words, /d/ does; e.g., *ddoll). As explained in Chapter 1, Section 1.1, while many inconsistencies in the alphabetic system can be explained by phoneme-grapheme correspondences, not all have a sound-based explanation, and such approach cannot explain the variation in systems such as abjad, abugidas, morphosyllabaries or syllabaries (Daniels &

Share, 2018; Share, 2018), where the phonology is not relevant. The main contribution of this thesis is the demonstration of developing spellers' sensitivity to spelling patterns that do not have any phonological explanation. In light of the criticism regarding anglocentricities in literacy research (Share, 2008), it is important to acknowledge that all studies in this thesis were carried out with English speaking participants and used Latin alphabet to create the stimuli (except Experiment 2). We justify this by highlighting the huge variation in written language and the importance of a thorough consideration of the features of one particular system when aiming to understand the skills involved in learning a written language. However, the cognitive processes involved in learning our visually presented graphotactic patterns are assumed to be similar across languages and cultures (e.g., Samuelsson et al., 2005), and thus, it will be important in future work to compare the results from the studies in this thesis with comparable studies carried out in different scripts outside of the Latin alphabet such as the logographic system (where a written character represents a whole word or phrase). In Mandarin, for example, radicals are composed of specific positional consistencies (e.g., semantic radicals are usually on the left while phonetic radicals are on the right) that are not explicitly taught to children in schools, yet Chinese children have reportedly been able to pick them up nevertheless (Tong, Tong, & McBride, 2017).

7.5.2 Methodological and practical implications

Beyond the key investigations discussed so far, a further contribution of the current thesis was the methodological development and validation of a production fill-in-the-blanks task as a new measure of artificial pattern sensitivity. Unlike the legality judgment task (which we also employed for consistency with previous work), production performance simulates more closely what children and adults do in naturalistic situations. Because of the added difficulty in producing novel structures compared to those acquired from experience with participants' own language, our task was designed to provide additional support by making most of the individual components of a target word available and requiring the participants to select the one that would create a familiar word from two options. This should ensure that performance does not fall to floor effects, while still simulating processes involved in the production rather than recognition of words (see also Kohnen, Nickels, & Castles, 2009). In all experiments (except

Experiment 2, where fill-in-the-blanks could not be included due to methodological issues), above chance and similar performance (if slightly higher in the fill-in-the-blanks) was recorded in both test tasks. This, together with a positive correlation seen between the newly devised (fill-in-the-blanks) and previously established and widely used (legality judgment) task, at least for adults, confirmed that we have used a measure to tap into the same knowledge. This newly validated task also provided an additional indication of the robustness of our results.

In addition, the current thesis presents reliable and objective data due of the use of open science practices such as pre-registering specific hypotheses and methods, and the use of Bayes factor analyses. By pre-specifying some of our predictions meant that any results that were contrary to the prediction could be discussed and interpreted. All other results were therefore treated as exploratory and could not be accredited to the theory in a post-hoc manner. Second, with respect to the Bayes factor analyses, the thesis addressed a significant problem in the literature, namely the interpretation of no effect on the basis of a null from traditional p-values. By including correlational Bayes factors, we clearly demonstrate that nonsignificant correlations do not necessarily constitute evidence for the null, as often interpreted in the literature. For example, a recent study by Qi, Sanchez Araujo, Georgan, Gabrieli, and Arciuli (2019) reported significant associations between auditory but not visual statistical learning ability and reading fluency and took this as evidence that “hearing is more important than seeing” for literacy development. In line with Dienes (2014), we caution against these statements on the basis of frequentist results alone. Bayes factors, allow researchers to distinguish between noisy findings that are typical of work with children, and actual evidence for the null. When evidence for the null is found (as was the case in this thesis in several instances), this is theoretically relevant; and when ambiguous evidence is found, this may indicate that the paradigm may be measuring a certain amount of noise. Researchers in literacy have not made wide use of Bayes factors, probably because of the perceived difficulty of informing the priors when comparable past data is not available. Having employed a range of methods of modelling H_1 throughout this thesis, I demonstrated that it is possible to use Bayes factor analyses for a range of comparisons and tests. Because the choice of H_1 was either pre-registered or clearly justified, the possibility of experimenter bias was eliminated, especially the possibility of selecting an H_1 after collecting and analyzing the data. The use of robustness regions in this thesis also enhanced the transparency of the approach by allowing readers to evaluate how the choice of the H_1 affects the inferences that were drawn. This thesis can form the basis of further research in the area of statistical learning of spelling patterns and encourage the use of Bayes factors, particularly when looking for correlations with literacy attainment, where the evidence for the null has been consistently interpreted as no relationship.

7.6 Limitations and future directions

In the individual chapters in this thesis (as well as earlier in this chapter), limitations of specific designs and potential areas for future research have been identified. In this section, I bring all these points together and present broader implications for future work.

Following directly from this thesis, one area for future work is extending the semi-artificial experiment (Experiment 2) to assess contextual learning with children. In light of Nigro et al. (2016) finding that dyslexic children could not generalize the rules, it is also important to carry out the experiments with both positional and contextual constraints with dyslexic children, and use Bayes factors to evaluate the evidence for the null.

As pointed in Section 7.2 above, the benefit of explicit instruction may have been evident due to the relatively simple rules that were taught and verbalized. Therefore, future work could compare children's performance on implicit and explicit learning when the rule is the more difficult one (Experiment 4). It is important to establish whether we can still see the benefit of explicit instruction when teaching rules that are harder to understand, or, as indicated by Reber et al. (1991), we might see a reverse benefit. To the point of the benefit of explicit instruction, I also noted that the pattern of results may have been due to the demonstration of learning through short-term retention. I suggested that, one way of ensuring the validity of our tests and manipulations, is to incorporate delayed testing and assess the long-term retention of the learning effects seen following such brief exposure. C. F. Taylor and Houghton (2005) for example, demonstrated that artificial phonotactic learning induced under strict laboratory settings fades much quicker than similarly induced "learning" (i.e., reinforcement) of real phonotactic constraints. Such investigation could inform future study designs.

A limitation noted in Experiment 6 (Section 6.2.3) was that each target word that was associated with a picture in the word learning task, was accompanied by a foil in one trial (i.e., one picture and two word options), and therefore the total exposure to each word was doubled. Given that current evidence for a difference between learning in a meaningful condition (Experiment 6) and a situation where focus was solely on form (Experiment 4), a future study could address the issue of unequal number of exposures to the target words by using two pictures and one word rather than one picture and two words to select from when learning the meanings. Furthermore, it is important to also carry out this experiment with children.

I now turn to broader future directions that follow from this thesis, with particular consideration for the difficulties in adapting lab-based experiments with children. Firstly, as noted in Section 1.7.2, recent scrutiny of offline measures of statistical learning (Siegelman et al., 2017) point towards the need to devise online tasks to complement the offline recall tasks.

An interesting challenge for future lab-based work is how to move away from describing learning as purely implicit or explicit. It is more likely that these mechanisms interact in complex ways over the course of becoming literate (Karmiloff-Smith, 1991; Steffler, 2001). In particular, it would be interesting to explore whether learners register associations explicitly and integrate them implicitly or vice versa (Yang & Li, 2012). As pointed in Section 7.2 above, the use of awareness questionnaire helped identify, albeit with some level of uncertainty, that performance was higher when participants could verbalize the patterns. It is important, therefore, to establish a way of measuring explicit awareness in a reliable way. Future work also needs to explore more systematically, what specifically makes explicit instruction so beneficial in experiments (and real-life conditions). Does it suffice to bring the pattern embedding nature of the stimuli to participants' attention or does the rule need to be taught? Would the benefits have been eliminated if the rule was not exemplified by further exposure? These are questions with important implications for classroom instruction.

As noted earlier in this chapter, the ecological validity of our measures, particularly in regards to the type of materials used, was of particular concern and has been a longstanding issue in statistical learning research (Pacton et al., 2001;

Pelucchi et al., 2009; C. F. Taylor & Houghton, 2005). The stimuli used throughout this thesis, as well as those often used in similar research, lack in complexity in various ways including that they are identical (or very similar) in length, embed deterministic rather than probabilistic constraints (e.g., "*t* never follows *o*" rather than "*t* follows *o* 20% of the time"), and only one type of statistical information was tested in each experiment. Natural orthographies, on the other hand, are replete with statistical cues that learners detect simultaneously and which assist them in linguistic-related tasks. Therefore, despite the impressive speed with which children and adults demonstrated learning of graphotactics under implicit conditions, the challenges faced in lab-based experiments are not likely to be comparable to real-life learning. These limitations have also been highlighted by Siegelman et al. (2017) and may contribute to the mixed results found in studies looking for a relationship between statistical learning and literacy (discussed above, in Section 7.3). Although the results in this thesis indicate the importance of both text exposure and explicit instruction, to fully establish their relevance for literacy, our studies need to scale up to looking at more naturalistic complex spelling rules. Such research is needed in order to test applying the techniques established here to the learning of actual spelling rules (using a pre– post– test design, where knowledge of the rules in question is first tested, and then assess how this knowledge improves following exposure). The principles developed in this way could inform programs that aim to teach actual English rules and could be used to supplement classroom teaching.

While the findings in this thesis primarily concern typical populations, they are also relevant for poor spellers, including children diagnosed with dyslexia. Future experiments should investigate whether the advantage for explicit over incidental graphotactic learning also holds in these populations, as suggested by previous educational research (Wanzek et al., 2006). Graphotactics are an interesting case as they have been shown to affect nonword spellings produced by adults ranging from poor to good spelling ability (Treiman & Boland, 2017). This is in contrast to work documenting statistical learning deficits in dyslexic children (Ise, Arnoldi, Bartling, & Schulte-Körne, 2012; Pavlidou, Kelly, & Williams, 2010; Vicari et al., 2015; see, however, Nigro et al., 2016). We believe that our work combining insights from the literatures on real graphotactic sensitivity and the statistical learning of artificial languages has great potential to shed light on such conflicting findings.

7.7 Conclusion

In conclusion, this thesis demonstrated that visual statistical learning processes support the learning of graphotactic patterns that do not correlate with phonology. It demonstrates this for different types of constraints in different types of materials, and critically, provides the first direct evidence for this in relative beginner spellers tested using context-based graphotactic patterns. The demonstration of graphotactic sensitivity in 6–7-year-olds is in contrast to literacy models which predict that young children are not sensitive to letter patterns, morphological information, and other advanced sources of knowledge until they enter the final (correct or fully alphabetic) stage of spelling development (Frith, 1985; Gentry, 1982). The thesis also demonstrated that, under explicit rule instructions, learning improves substantially, and that only under explicit learning conditions were correlations found with measures of literacy. An implication of these findings for spelling instruction is that, although early incidental exposure to print may play a role and should be encouraged, explicit teaching, at least of the patterns studied here, may be altogether more effective.

References

Adini, Y., Bonneh, Y. S., Komm, S., Deutsch, L., & Israeli, D. (2015). The time course and characteristics of procedural learning in schizophrenia patients and healthy individuals. *Frontiers in Human Neuroscience*, *9*, 1–16. https://doi.org/10.3389/fnhum.2015.00475

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture*, *10*(1), 3–28. https://doi.org/10.1075/gest.10.1.02ali

Althoff, R. R., & Cohen, N. J. (1999). Eye-Movement-Based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(4), 997–1010. https://doi.org/10.1037/0278-7393.25.4.997

Anwyl-irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst : An online behavioral experiment builder. *Behavior Research Methods*. https://doi.org/https://doi.org/10.3758/s13428-019-01237-x

Arciuli, J. (2017). The multi-component anature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2016.0058

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children : the role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473. https://doi.org/10.1111/j.1467-7687.2009.00937.x

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Arciuli, J., von Koss Torkildsen, J., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, *5*(JUL), 1–8. https://doi.org/10.3389/fpsyg.2014.00747

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research : past , present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372: 20160047. https://doi.org/10.1098/rstb.2016.0047

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal* *of Memory and Language*, *92*, 265–280. https://doi.org/10.1016/j.jml.2016.07.004

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170–176. https://doi.org/10.1177/0963721412436806.Statistical

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

Atick, J. (1992). Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems*, *3*(2), 213–251. https://doi.org/10.1088/0954-898x/3/2/009

Baguley, T., & Kaye, D. (2009). Book review: "Understanding psychology as a science: an introduction to scientific and statistical inference" by Z. Dienes. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 695–698.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001.Random

Barry, C. (1994). Spelling routes (or roots or rutes). *Handbook of Spelling: Theory, Process and Intervention*, 27–49.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Batterink, L. J., Cheng, L. Y., & Paller, K. A. (2016). Neural measures reveal implicit learning during language processing. *Journal of Cognitive Neuroscience*, *28*(10), 1636–1649. https://doi.org/10.1162/jocn_a_00985

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Impplicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

Bendor-Samuel, J. (1996). African languages. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 689–691). Oxford University Press.

Bhattacharya, A., & Ehri, L. C. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities*, *37*(4), 331–348. https://doi.org/10.1177/00222194040370040501

Bissex, G. L. (1980). *Gnys at wrk: A child learns to write and read*. Cambridge, Massachusetts: Harvard University Press.

Bolinger, D. L. (1946). Visual morphemes. *Language*, *22*(4). https://doi.org/10.2307/409923

Bosman, A. M. . T., & de Groot, A. M. (1991). De ontwikkeling van woordbeelden bij beginnende lezers en spellers [The development of orthographic images in beginning readers and spellers]. *Pedagogische Studiën*, *68*(5), 199–215.

Bosman, A. M. T., & de Groot, A. M. (1992). Differential effectiveness of reading and non-reading tasks in learning to spell. In *Literacy without fronteers* (pp. 279–289).

Bosman, A. M. T., van Hell, J. G., Verhoeven, L., Hell, J. G. Van, & Verhoeven, L. (2006). Learning the spelling of strange words in dutch benefits from regularized reading. *Journal of Educational Psychology*, *98*(4), 879–890. https://doi.org/10.1037/0022-0663.98.4.879

Bosman, A. M. T., & Van Orden, G. C. (1997). Why spelling is more difficult than reading. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 173–194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bradley, L., & Bryant, P. E. (1979). Independence of reading and spelling in backward and normal readers. *Developmental Medicine & Child Neurology*, *21*, 504–514.

Brewer, W. F. (1974). There is no convincing evidence for operant or classical conditioning in adult humans. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes*. Lawrence Erlbaum Associates.

Brooks, D. N., Rosch, E., & Lloyd, B. B. (1978). Nonanalytic concept formation and memory for instances. *Cognition and Categorization*.

Brown, H., Weighall, A., Henderson, L. M., & Gaskell, G. M. (2012). Enhanced recognition and recall of new words in 7- and 12-year-olds following a period of offline consolidation. *Journal of Experimental Child Psychology*, *112*(1), 56–72. https://doi.org/https://doi.org/10.1016/j.jecp.2011.11.010

Bruck, M., & Waters, G. (1990). Effects of reading skill on component spelling skills. *Applied Psycholinguistics*, *11*(4), 425–437. https://doi.org/10.1017/S0142716400009668

Bryant, P., & Bradley, L. (1980). Why children sometimes write words which they do not read. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 355–370). London: Academic Press.

Burns, J. M., & Richgels, D. J. (1989). An investigation of task requirements associated with the invented spellings of 4-year-olds with above average intelligence. *Journal of Reading Behavior*, *21*(1), 1–14. https://doi.org/10.1080/10862968909547655

Burt, J., & Butterworth, P. (1996). Spelling in adults: Orthographic transparency, learning new letter strings and reading accuracy. *European Journal of Cognitive Psychology*, *8*(1), 3–44. https://doi.org/10.1080/095414496383194

Burt, J., & Fury, M. (2000). Spelling in adults: The role of reading skills and experience. *Reading and Writing*, *13*, 1–30. https://doi.org/10.1023/A:1008071802996

Butyniec-Thomas, J., & Woloshyn, V. E. (1997). The effects of explicit-strategy and whole-language instruction on students' spelling ability. *The Journal of Experimental Education*, *65*(4), 293–302. https://doi.org/10.1080/00220973.1997.10806605

Byrne, B. (1992). Procedure for reading: rationale, hypotheses, and data. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading Aquisition* (pp. 10–51). Routledge.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T., Johannesson, M., Kirchler, M., … Wu, H. (2019). Evaluating the Replicability of Social Science Experiments in Nature and Science. *Nature and Science*, 1–26.

Caravolas, M. (2004). Spelling development in alphabetic writing systems: A cross-linguistic perspective. *European Psychologist*, Vol. 9, pp. 3–14. https://doi.org/10.1027/1016-9040.9.1.3

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language*, *45*(4), 751–774. https://doi.org/10.1006/jmla.2000.2785

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology*, *92*(4), 307–321. https://doi.org/10.1016/j.jecp.2005.08.001

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, *24*(8), 1398–1407. https://doi.org/10.1177/0956797612473122

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*, *23*(5), 386–402. https://doi.org/10.1080/10888438.2019.1580284

Cardoso-Martins, C. (2005). Beginning reading acquisition in Brazilian Portuguese. In R. Malatesha Joshi & P. G. Aaron (Eds.), *Handbook og orthography and literacy* (pp. 171–187). New York: Routledge.

Cardoso-Martins, C., Corrêa, M. F., Lemos, L. S., & Napoleão, R. F. (2006). Is there a syllabic stage in spelling development? Evidence from Portuguese-speaking children. *Journal of Educational Psychology*, *98*(3), 628–641. https://doi.org/10.1037/0022-0663.98.3.628

Carrillo, S., & Alegría, J. (2014). The development of children' s sensitivity to bigram frequencies when spelling in Spanish, a transparent writing system. *Reading and Writing*, *27*, 571–590. https://doi.org/10.1007/s11145-013-9459-y

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, *89*(4), 631–644. https://doi.org/10.1037/0022-0663.89.4.631

Cataldo, S., & Ellis, N. (1988). Interactions in the development of spelling, reading and phonological skills. *Journal of Research in Reading*, *11*(2), 86–109. https://doi.org/10.1111/j.1467-9817.1988.tb00153.x

Catania, A. C. (1988). The behavior of organisms as work in progress. *Journal of the Experimental Analysis of Behavior*, *50*(2), 277–281. https://doi.org/10.1901/jeab.1988.50-277

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*, B69–B77. https://doi.org/10.1016/S0

Chambers, K. E., Onishi, K. H., & Fisher, C. (2010). A vowel is a vowel: Generalising newly-learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(3), 821–828. https://doi.org/10.1037/a0018991.A

Chambers, K. E., Onishi, K. H., & Fisher, C. (2011). Representations for phonotactic learning in infancy. *Language Learning and Development*, *7*(4), 287–308. https://doi.org/10.1080/15475441.2011.580447

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103–120. https://doi.org/10.1016/j.cognition.2017.02.015

Chetail, F., Balota, D., Treiman, R., & Content, A. (2015). What can megastudies tell us about the orthographic structure of English words? *Quarterly Journal of Experimental Psychology*, *68*(8), 1519–1540. https://doi.org/10.1080/17470218.2014.963628

Chomsky, N. (1957). Logical structures in language. *American Documentation*, *8*(4), 284–300. https://doi.org/10.1002/asi.5090080406

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*, *35*(1), 26–58.

Chomsky, N. (1979). The logical structure of linguistic theory. *Synthese*, *40*(2), 317–352.

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Christiansen, M. H., & Chater, N. (2015). The language faculty that wasn't: A usage-based account of natural language recursion. *Frontiers in Psychology*, Vol. 6. https://doi.org/10.3389/fpsyg.2015.01182

Christianson, K., & Devlin, J. T. (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society: August 7-10, 1997, Stanford University*.

Church, B. A., & Fisher, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language*, *39*(4), 523–542. https://doi.org/10.1006/jmla.1998.2601

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, Vol. 2, pp. 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to internal lexicon. In S. Dominic (Ed.), *Attention and performance IV*: *Vol. VI*. Hillsdale, NJ: Erlbaum.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Coulmas, F., Ehlich, K., & Winter, W. (Eds.). (1983). *Writing in focus*. Berlin: Moulton Publishers.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

Critchley, M. (1975). Developmental dyslexia: Its history, nature and prospects. In D. Duane & M. B. Rawson (Eds.), *Proceedings of the world congress on dyslexia on reading, perception and language (vol. xii, pp. 272)*.

Crystal, D. (1995). Speaking of Writing and Writing of Speaking. *Longman Language Review*, *1*, 5–8.

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing*, *14*(5–6), 549–568. https://doi.org/10.1023/a:1011100226798

Cunningham, A. E., Perry, K. E., Stanovich, K. E., & Share, D. L. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, *82*(3), 185–199. https://doi.org/10.1016/S0022-0965(02)00008-5

Daniels, P. T., & Share, D. L. (2018). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading*, *22*(1), 101–116. https://doi.org/10.1080/10888438.2017.1379082

Danjon, J., & Pacton, S. (2009). Children’s learning about properties of double letters: The case of French. *The 16th European Society for Cognitive Psychology Conference (ESCOP)*, 2–5.

de Bree, E., Geelhoed, J., & van den Boer, M. (2018). Overruled !: Implicit cues rather than an orthographic rule determine Dutch children’ s vowel spelling. *Learning and Instruction*, *56*, 30–41. https://doi.org/10.1016/j.learninstruc.2018.03.006

de Bree, E., van der Ven, S., & van der Maas, H. (2017). The Voice of Holland: Allograph production in written Dutch past tense inflection. *Language Learning and Development*, *13*(3), 215–240. https://doi.org/10.1080/15475441.2016.1217777 de Jong, P. F., & Share, D. L. (2007). Orthographic learning during oral and silent reading. *Scientific Studies of Readingfic*, *11*(1), 55–71. https://doi.org/10.1080/10888430709336634

Deacon, S. H. (2012). Sounds, letters and meanings: The independent influences of phonological, morphological and orthographic skills on early word reading accuracy. *Journal of Research in Reading*, *35*(4), 456–475. https://doi.org/10.1111/j.1467-9817.2010.01496.x

Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just “more phonological”? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics*, *25*(2), 223–238. https://doi.org/10.1017/S0142716404001110

Deacon, S. H., Leblanc, D., & Sabourin, C. (2011). When cues collide: Children’s sensitivity to letter- and meaning-patterns in spelling words in English. *Journal of Child Language*, *38*(4), 809–827. https://doi.org/10.1017/S0305000910000322

DeFrancis, J. (1989). *Visible speech: the diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

DeHaene, S. (2009). *Reading in the Brain: The New Science of How We Read*. Penguin Books.

DeHaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, Vol. 15. https://doi.org/10.1016/j.tics.2011.04.003

Dell, G. S. (2007). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349. https://doi.org/10.1080/01690969008407066

Dell, G. S., Reed, K. D., Adams, D. R., & 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. https://doi.org/10.1037/0278-7393.26.6.1355

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(July), 1–17. https://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In M. Overgaard (Ed.), *Behavioural Methods in Consciousness Research* (pp. 199–220). Oxford: Oxford University Press.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dienes, Z. (2020). *Obtaining evidence for no effect*. Retrieved from https://psyarxiv.com/yc7s5/

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit nnowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 875–887.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment : How conscious and how abstract? *Journal of Experimental Psychology: General*, *113*(4), 541–555.

Ehri, L. C. (1980). The development of orthographic images. In *Cognitive processes in spelling*.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to Spell: Research, Theory, and Practice Across Languages* (pp. 237–266). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C. (2008). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 135–154). https://doi.org/10.1002/9780470757642.ch8

Ehri, L. C., & Roberts, K. T. (1979). Do beginners learn printed words better in contexts or in isolation? *Child Development*, *50*(3), 675–685. https://doi.org/https://doi.org/10.2307/1128932

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing: An Interdisciplinary Journal*, *7*(3), 295–326. https://doi.org/https://doi.org/10.1007/BF03162082

Endress, A. D., & Mehler, J. (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and*

Performance, *36*(1), 235–250. https://doi.org/10.1037/a0017164
Esper, E. A. (1925). *Language monographs* (G. M. Bolling, A. Espinosa, & E. Sapir, Eds.). Philadelphia: Linguistic Society of America.
Ethnologue. (2020). Ethnologue: Languages of the world. Retrieved from https://www.ethnologue.com/
Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321–335. https://doi.org/10.1044/1092-4388(2009/07-0189)
Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Heinemann.
Fiser, J., & Aslin, R. N. (2001). Unsupervised statsistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392
Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458
Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521–537. https://doi.org/10.1037/0096-3445.134.4.521
Fisher, C., & Gleitman, L. R. (2002). Language acquisition. In H. Pashler (Ed.), *Steven's handbook of experimental psychology* (3rd ed.). Wiley.
Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, *12*(1), 13–18. https://doi.org/10.1111/1467-8721.01213
Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, *22*(1), 179–228. https://doi.org/10.1023/B:NALA.0000005557.78535.3c
Frith, U. (1979). Reading by eye and writing by ear. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of Visible Language. Nato Conference Series, vol 13*. https://doi.org/10.1007/978-1-4684-0994-9_23
Frith, U. (1980). *Cognitive processes in spelling*. New York: Academic Press.
Frith, U. (1985). Beneath the surface of developmental dyslexia. In *Journal of Chemical Information and Modeling* (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1). https://doi.org/10.1037//0278-7393.20.1.116

Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2). https://doi.org/10.1037//0278-7393.21.2.398

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*(1), 71–99. https://doi.org/10.1037/0033-2909.123.1.71

Frost, R. (2008). Orthographic systems and skilled word recognition processes in reading. In *The Science of Reading: A Handbook*. https://doi.org/10.1002/9780470757642.ch15

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*(5), 263–279. https://doi.org/10.1017/S0140525X11001841

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality vs. modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010.Domain

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243–1252. https://doi.org/10.1177/0956797612472207

Gambell, T., & Yang, C. (2005). Word segmentation: Quick but not dirty. *Unpublished Manuscript*, 1–36. Retrieved from http://www.ling.upenn.edu/~ycharles/papers/quick.pdf

Gaskins, I. W., Ehri, L. C., Cress, C., O'Hara, C., & Donnelly, K. (1997). Analyzing words and making discoveries about the alphabetic system: Activities for beginning readers. *Language Arts*, *74*(3), 172–184.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(1), 34–54. https://doi.org/10.1037/0278-7393.33.1.34

Gentry, R. J. (1982). An analysis of developmental analysis in GNYS AT WRK spelling. *The Reading Teacher*, *36*(2), 192–200.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*, 249–268. https://doi.org/10.1017/S0305000904006786

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281. https://doi.org/10.1037/0096-3445.113.2.256

Ghetie, I. (1978). *Istoria limbii romane literare: privire sintetica*. Editura Stiintifica si Enciclopedica.

Gilbert, L. C. (1934). Effect of reading on spelling in the ninth grade. *The School Review*, *42*(3), 952–977. https://doi.org/10.1086/439618

Gilbert, L. C. (1935). A study of the effect of reading on spelling. *Journal of Educational Research*, *28*(8). https://doi.org/10.1080/00220671.1935.10880525

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135. https://doi.org/10.1016/s0010-0277(99)00003-7

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*(5), 567–580. https://doi.org/10.1111/j.1467-7687.2004.00381.x

Goswami, U. (2012). Universals of reading: Developmental evidence for linguistic plausibility. *Behavioral and Brain Sciences*, *35*, 287–288. https://doi.org/doi:10.1017/S0140525X12000052

Goswami, U., & Bryant, P. (2016). Phonological skills and learning to read. In *Phonological Skills and Learning to Read*. https://doi.org/10.4324/9781315785110

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: an unnatural act. *Bulletin of the Orton Society*, *30*(1). https://doi.org/10.1007/BF02653717

Gough, P. B., Juel, C., & Griffith, P. L. (1992). Reading, spelling, and the orthographic cipher. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition*. Routledge.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x.Can

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers , readers , and writers ? A meta-analytic review. *Reading and Writing*, *27*, 1703–1743. https://doi.org/10.1007/s11145-014-9517-0

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, *25*(1–2). https://doi.org/10.1016/0010-0277(87)90007-2

Halliday, M. A. K. (1977). Aims and perspectives in linguistics. *ALAA Occasional Papers Number 1*. Mt. Gravatt, Brisbane: Applied Linguistics Association of Australia.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Harris, Z. S. (1955). From phoneme to morpheme. *Language*, *31*(2), 190–222. https://doi.org/10.2307/411036

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate:Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64.

Hayes, H., Treiman, R., & Kessler, B. (2006). Children use vowels to help them spell consonants. *Journal of Experimental Child Psychology*, *94*(1), 27–42. https://doi.org/10.1016/j.jecp.2005.11.001

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, G. M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674–687. https://doi.org/https://doi.org/10.1111/j.1467-7687.2012.01172.x

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Hollich, G., Jusczyk, P. W., & Luce, P. A. (2001). *Infants' memory for similar sounding words: phonetic false memories*. Minneapolis, MN: Paper presented to the Society for Research on Child Development.

Holmes, V. M., & Castles, A. E. (2001). Unexpectedly poor spelling in university students. *Scientific Studies of Reading*, *5*(4), 319–350. https://doi.org/10.1207/s1532799xssr0504_02

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, (48), 1393–1409.

https://doi.org/10.3758/s13428-015-0647-3

Howard, J. H., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia*, *44*(7), 1131–1144. https://doi.org/10.1016/j.neuropsychologia.2005.10.015

Huang, B. H. (2016). *A synthesis of empirical research on the linguistic outcomes of early foreign language instruction*. *13*(3), 257–273. https://doi.org/10.1080/14790718.2015.1066792

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66. https://doi.org/10.1016/j.cogpsych.2009.01.001.Getting

Hwayed, M. H. (2010). General phonological and morphological justifications of homophones. *Journal of Research Diyala Humanity*, (45), 712–730.

Invernizzi, M. A., Abouzeid, M. P., & Bloodgood, J. W. (1997). Integrated word study: Spelling, grammar, and meaning in the language arts classroom. *Language Arts*, *74*(3), 185–192. Retrieved from https://www.jstor.org/stable/41482857

Ise, E., Arnoldi, C. J., Bartling, J., & Schulte-Körne, G. (2012). Implicit learning in children with spelling disability : evidence from artificial grammar learning. *J Neural Transm*, (119), 999–1010. https://doi.org/10.1007/s00702-012-0830-y

Ise, E., Arnoldi, C. J., & Schulte-Körne, G. (2014). Development of orthographic knowledge in German-speaking children: A 2-year longitudinal study. *Journal of Research in Reading*, *37*(3), 233–249. https://doi.org/10.1111/j.1467-9817.2012.01535.x

Jeffreys, H. (1961). *Theory of Probability*. London: Oxford University Press.

Johnson, E. K., & 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. https://doi.org/10.1006/jmla.2000.2755

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339–345. https://doi.org/10.1111/j.1467-7687.2009.00886.x

Joseph, H. S. S. L., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238–248. https://doi.org/10.1016/j.cognition.2014.06.015

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*(3), 402–420. https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645. https://doi.org/10.1006/jmla.1994.1030

Kachergis, G., Yu, C., & Shiffrin, R. M. (2010). Cross-situational statistical learning: Implicit or intentional? *Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)*, (1069–7977).

Kapatsinski, V., & Johnston, L. H. (2010). Investigating phonotactics, lexical analogy, and sound symbolism using xenolinguistics: A novel word-picture matching paradigm. In S. Ohlson & R. Catrambone (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society, 2010-2015*. Austin, Tx: The Cognitice Science Society.

Karmiloff-Smith, A. (1991). Beyond modularity: Innate constraints and developmental change. In S. Carey & R. Gelman (Eds.), *The Epigenesis of Mind* (pp. 171–196). Lawrence Erlbaum Associates.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1). https://doi.org/10.1037/0278-7393.9.1.157

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology*, *94*(C). https://doi.org/10.1016/S0166-4115(08)62789-2

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

Kemp, N., & Bryant, P. (2003). Do Beez Buzz? Rule-Based and Frequency-Based Knowledge in Learning to Spell Plural -s. *Child Development*, *74*(1), 63–74. https://doi.org/10.1111/1467-8624.00521

Kemper, M. J., Verhoeven, L., & Bosman, A. M. T. (2012). Implicit and explicit instruction of spelling rules. *Learning and Individual Differences*, *22*(6), 639–649. https://doi.org/10.1016/j.lindif.2012.06.008

Kessler, B. (2009). Statistical learning of conditional orthographic correspondences. *Writing Systems Research*, *1*(1), 19–34. https://doi.org/10.1093/wsr/wsp004

Kessler, B., Pollo, T. C., Treiman, R., & Cardosso-Martins, C. (2013). Frequency analyses of prephonological spellings as predictors of success in conventional spelling. *Journal of Learning Disabilities*, *46*(3), 252–259. https://doi.org/10.1177/0022219412449440

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, *37*, 295–311. https://doi.org/https://doi.org/10.1006/jmla.1997.2522

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, *44*(4), 592–617. https://doi.org/10.1006/jmla.2000.2745

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 247–304. https://doi.org/10.3758/s13428-011-0118-4

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184. https://doi.org/10.1037/a0025405

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy : evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5

Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources. *Australian Journal of Learning Difficulties*, *14*(1), 113–150. https://doi.org/10.1080/19404150902783450

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

Kreiner, D. S., & Gough, P. B. (1990). Two ideas about spelling: Rules and word-specific memory. *Journal of Memory and Language*, *29*(1), 103–118. https://doi.org/10.1016/0749-596X(90)90012-O

Krishnan, S., & Watkins, K. E. (2019). A challenge for the procedural deficit hypothesis: How should we measure sequential learning in childhood? *Developmental Science*, *22*(4), 9–10. https://doi.org/10.1111/desc.12815

Kuhl, P. K. (2011). Early language learning and literacy: Neuroscience implications for education. *Mind, Brain, and Education*, *5*(3). https://doi.org/10.1111/j.1751-228X.2011.01121.x

Kuppuraj, S., Duta, M., Thompson, P., & Bishop, D. (2018). *Online incidental statistical learning of audiovisual word sequences in adults : a registered report*. (February). https://doi.org/10.1098/rsos.171678

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research : A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lany, J., & Saffran, J. R. (2010). *From statistics to meaning: infants' acquisition of lexical categories*. *21*(2), 1–14. https://doi.org/10.1177/0956797609358570.From

Lass, R. (1994). Old English - Historical linguistic companion. In *Old English: A Historical Linguistic Companion*. Cambridge University Press.

Lehtonen, A., & Bryant, P. (2005). Doublet challenge: form comes before function in children ' s understanding of their orthography. *Developmental Science*, *8*(3), 211–217.

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language*, *115*, 104152. https://doi.org/10.1016/j.jml.2020.104152

Lété, B., Peereman, R., & Fayol, M. (2008). Consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study. *Journal of Memory and Language*, *58*(4). https://doi.org/10.1016/j.jml.2008.01.001

Levy, Y. (2012). No reason to expect "reading universals." *Behavioral and Brain Sciences*, *35*, 263–329. https://doi.org/doi:10.1017/S0140525X12000088

Ludwig, O. (1983). Writing systems and written language. In F. Coulmas & K. Ehnlich (Eds.), *Writing in Focus* (pp. 31–42). Mouton Publishers.

Malmquist, E. (1958). Factors related to reading disabilities in the first grade of the elementary school. *Educational Research*, *1*(1), 69–72. https://doi.org/10.1080/0013188580010106

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science*, *12*(6), 1020–1025.

https://doi.org/10.1111/j.1467-7687.2009.00838.x

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(1), 77–80.

Marinelli, C. V., Romani, C., Burani, C., & Zoccolotti, P. (2015). Spelling acquisition in english and Italian: A cross-linguistic study. *Frontiers in Psychology*, *6*, 1843. https://doi.org/10.3389/fpsyg.2015.01843

Marinus, E., Kohnen, S., & McArthur, G. (2013). Australian comparison data for the Test of Word Reading Efficiency (TOWRE). *Australian Journal of Learning Difficulties*, *18*(2), 199–212. https://doi.org/10.1080/19404158.2013.852981

Mattingly, I. G. (1972). Reading, the linguistic process and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eyes: the relationships between speech and reading* (pp. 133–147). Cambridge: MA, MIT Press.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. In *Cognition* (Vol. 78). https://doi.org/10.1016/S0010-0277(00)00109-8

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494. https://doi.org/10.1006/cogp.1999.0721

McBride, C. A. (2016). Is Chinese special? Four aspects of Chinese literacy acquisition that might distinguish learning Chinese from learning alphabetic orthographies. *Educational Psychology Review*, *28*, 523–549. https://doi.org/10.1007/s10648-015-9318-2

McCarthy, J. (1979). On stress and syllabification. *Linguistic Inquiry*, *10*(3), 443–465. https://doi.org/10.2307/4178121

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(23/30), 746–748.

Meulemans, T., & Van Der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(4), 1007–1028. https://doi.org/10.1037/0278-

7393.23.4.1007
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychology Review*, *63*(1), 81–97. https://doi.org/10.1037/h0043158
Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x
Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*(SEP), 1–9. https://doi.org/10.3389/fpsyg.2010.00031
Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: Cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1081–1091. https://doi.org/10.1037/a0023700
Mommers, M. J. C. (1987). An investigation into the relation between word recognition skills, reading comprehension and spelling skills in the first two years of primary school. *Journal of Research in Reading*, *10*(2), 122–143. https://doi.org/10.1111/j.1467-9817.1987.tb00290.x
Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*(1), 71–84. https://doi.org/10.1016/j.jecp.2006.06.004
Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2015). Implicit learning of written regularities and its relation to literacy acquisition in a shallow orthography. *Journal of Psycholinguistic Research*, *44*(5), 571–585. https://doi.org/10.1007/s10936-014-9303-9
Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2016). Implicit learning of non-linguistic and linguistic regularities in children with dyslexia. *Annals of Dyslexia*, *66*(2), 202–218. https://doi.org/10.1007/s11881-015-0116-9
Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192
Nunes, T., & Bryant, P. (2006). Improving literacy by teaching morphemes. In T.

Nunes & P. Bryant (Eds.), *Improving Literacy by Teaching Morphemes*. https://doi.org/10.4324/9780203969557

Nunes, T., Bryant, P., & Olsson, J. (2003). Learning morphological and phonological spelling rules: An intervention study. *Scientific Studies of Reading*, *7*(3), 289–307.

Nunn, A. (1998). *Dutch orthography. A systematic investigation of the spelling of Dutch words*. The Hague, the Netherlands: Holland Academic Graphics.

Oakhill, J., Cain, K., & Elbro, C. (2014). *Understanding and teaching reading comprehension*. https://doi.org/https://doi.org/10.4324/9781315756042

Olson, D. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review*, *47*(3). https://doi.org/10.17763/haer.47.3.8840364413869005

Olson, D. (1994). *The world on paper: the conceptual and cognitive implications of writing and reading*. Cambridge: The Press Syndicate of the University of Cambridge.

Ongchoco, J. D. K., Uddenberg, S., & Chun, M. M. (2016). Statistical learning of movement. *Psychonomic Bulletin & Review*, *23*, 1913–1919. https://doi.org/10.3758/s13423-016-1046-1

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, *83*, 13–23. https://doi.org/10.1016/S0010-0277(01)00165-2

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2745–2750. https://doi.org/10.1073/pnas.0708424105

Ormrod, J. E. (1986). Learning to spell while reading: A follow-up study. *Perceptual and Motor Skills*, *63*(2). https://doi.org/10.2466/pms.1986.63.2.652

Pacton, S., Borchardt, G., Treiman, R., Lété, B., & Fayol, M. (2014). Learning to spell from reading: General knowledge about spelling patterns influences memory for specific words. *Quarterly Journal of Experimental Psychology*, *67*(5), 1019–1036. https://doi.org/10.1080/17470218.2013.846392

Pacton, S., Fayol, M., & Perruchet, P. (2002). The acquisition of untaught orthographic regularities in French. In E. C. Vernhoven & P. Reitsma (Eds.), *Precursors of Functional Literacy* (pp. 121–136).

https://doi.org/10.1075/swll.11.11pac
Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development*, *76*(2), 324–339. https://doi.org/10.1111/j.1467-8624.2005.00848_a.x
Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, *130*(3), 401–426.
Pacton, S., Sobaco, A., Fayol, M., & Treiman, R. (2013). How does graphotactic knowledge influence children's learning of new spellings? *Frontiers in Psychology*, *4*(OCT). https://doi.org/10.3389/fpsyg.2013.00701
Pavlidou, E. V, Kelly, M. L., & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning? *Dyslexia*, *161*, 143–161. https://doi.org/10.1002/dys
Peirce, J., Gray, J. R., Simpson, S., Macaskill, M., Höchenberger, R., Sogo, H., … Peirce, J. (2019). PsychoPy2 : Experiments in behavior made easy. *Behavior Research Methods*, *3*(51), 195–203. https://doi.org/10.3758/s13428-018-01193-y
Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3). https://doi.org/10.1111/j.1467-8624.2009.01290.x
Peña, M., Bonatti, L. L., Marina, N., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–608. https://doi.org/10.1126/science.1072901
Pereaa, M., & Carreiras, M. (2012). Perceptual uncertainty is a property of the cognitive system. *Behavioral and Brain Sciences*, Vol. 35. https://doi.org/10.1017/S0140525X12000118
Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.
Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading*, *7*(1), 3–24. https://doi.org/10.1207/s1532799xssr0701_02
Perfetti, C. A. (2011). Phonology is critical in reading: But a phonological deficit is not the only source of low reading skill. In Susan A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining Individual Differences in Reading: Theory and Evidence* (1st Editio). https://doi.org/10.4324/9780203817964
Perfetti, C. A., & Harris, L. N. (2013). Universal reading processes are modulated by language and writing system. *Language Learning and Development*, *9*,

296–316. https://doi.org/10.1080/15475441.2013.813828

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Chapter 13: Reading in English and Chinese: Evidence for a "Universal" phonological principle. *Advances in Psychology*, *94*, 227–248. https://doi.org/10.1016/S0166-4115(08)62798-3

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning : one phenomenon , two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Pinker, S. (1994). *Pinker_1994*. London: Penguin Group.

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics*, *60*(6), 941–951. https://doi.org/10.3758/BF03211930

Pollo, T. C., Kessler, B., & Treiman, R. (2009). Statistical patterns in children's early writing. *Journal of Experimental Child Psychology*, *104*(4), 410–426. https://doi.org/10.1016/j.jecp.2009.07.003

Pollo, T. C., Treiman, R., & Kessler, B. (2007). Three perspectives on spelling development. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-Word Reading: Behavioral and Biological Perspectives* (pp. 175–189). https://doi.org/10.4324/9780203810064

Pollo, T. C., Treiman, R., & Kessler, B. (2008). Preschoolers use partial letter names to select spellings : Evidence from Portuguese. *Applied Psycholinguistics*, *29*, 1–18. https://doi.org/10.1017/S0142716407080095

Powell, M. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *NA Report NA2009/06*.

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D. E., & Arciuli, J. (2019). Hearing matters more than seeing : A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading*, *23*(1), 101–115. https://doi.org/10.1080/10888438.2018.1485680

R, C. T. (2014). R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. URL Http://Www.R-Project.Org/.*

Rastle, K. (2019). EPS mid-career prize lecture 2017: Writing systems, reading, and language. *Quarterly Journal of Experimental Psychology*, *72*(4), 677–

692. https://doi.org/10.1177/1747021819829696

Rastle, K., Lally, C., Davis, M. H., & Taylor, J. S. H. (2021). The dramatic impact of explicit instruction on learning to read in a new writing system. *Psychological Science*. https://doi.org/https://doi.org/10.1177/0956797620968790

Ravid, D. D. (2012). *Spelling morphology. The psycholinguistics of Hebrew spelling*. https://doi.org/10.1007/978-1-4419-0588-8 Springer

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3). https://doi.org/10.1037//0033-2909.124.3.372

Read, C. (1975). Children's categorization of speech sounds in English. In *NCTE Research Report No. 17*.

Read, C. (1986). *Children's Creative Spelling*. https://doi.org/https://doi.org/10.4324/9780429488924

Read, C., & Treiman, R. (2013). Children's invented spelling: what we have learned in forty years. In M. Piatelli-Palmarini & R. C. Berwick (Eds.), *Rich Languages From Poor Inputs* (pp. 195–209). https://doi.org/10.1093/acprof:oso/9780199590339.003.0013

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

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. https://doi.org/10.1037/0278-7393.2.1.88

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious* (xii). https://doi.org/10.1093/acprof:oso/9780195106589.001.0001

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 492–502.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual ifferences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 888–896. https://doi.org/10.1037/0278-7393.17.5.888

Robinson, C. W., & Sloutsky, V. M. (2007). Visual statistical learning: getting some help from the auditory modality. *Proceedings of the 29th Annual Cognitive Science Society.*, 611–616. https://doi.org/10.1039/b209297b

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in Grades 4, 6, and 8. *Journal of Experimental Child Psychology*, *102*(1), 96–113. https://doi.org/10.1016/j.jecp.2008.01.004

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rouder, J. N. (2014). Optional stopping : No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*, 301–308. https://doi.org/10.3758/s13423-014-0595-4

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior*, *10*(1), 57–62. https://doi.org/10.1016/S0022-5371(71)80094-4

Saenger. (1997). *Space between words: The origins of silent reading*. California: Stanford University Press.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*, 172–196. https://doi.org/10.1006/jmla.2001.2839

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*(January), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997).

Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*(3), 484–494. https://doi.org/10.1037/0012-1649.39.3.484

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273–284. https://doi.org/10.1207/S15327078IN0402_07

Samara, A., & Caravolas, M. (2014). Statistical learning of novel graphotactic constraints in children and adults. *Journal of Experimental Child Psychology*, *121*(1), 137–155. https://doi.org/10.1016/j.jecp.2013.11.009

Samara, A., Singh, D., & Wonnacott, E. (2019). Statistical learning and spelling: Evidence from an incidental learning experiment with children. *Cognition*, *182*(March 2018), 25–30. https://doi.org/10.1016/j.cognition.2018.09.005

Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, *94*, 85–114. https://doi.org/10.1016/j.cogpsych.2017.02.004

Samuelsson, S., Byrne, B., Quain, P., Wadsworth, S., Corley, R., DeFries, J. C., … Olson, R. (2005). Environmental and genetic influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology*, *97*(4), 705–722. https://doi.org/10.1037/0022-0663.97.4.705

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. In *Brain mapping: An encyclopedic reference* (pp. 501–506). https://doi.org/http://dx.doi.org/10.1016/B978-0-12-397025-1.00276-1

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to reading ability, and if so, why? *Scientific Studies of Reading*, *23*(1), 64–76. https://doi.org/10.1080/10888438.2018.1482304

Seidenberg, M. S. (2011). Reading in different writing systems. In P. McCardle, J. Ren, & O. Tzeng (Eds.), *Dyslexia Across languages: Orthography and the Gene-Brain-Behavious Link* (pp. 146–168). Paul Brooke Publishing.

Seidenberg, M. S. (2012). Writing systems: Not optimal, but good enough. *Behavioral and Brain Sciences*, *35*(5), 305–307. https://doi.org/10.1017/S0140525X12000337

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*, 553–554. https://doi.org/10.1126/science.1078094

Seidenberg, M. S., Petersen, A., MacDonald, M. C., & Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(1), 48–62. https://doi.org/10.1037/0278-7393.22.1.48

Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development*, *5*(3), 191–202. https://doi.org/10.1080/15475440902754326

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174. https://doi.org/10.1348/000712603321661859

Seymour, P. H. K., & Elder, L. (1986). Beginning reading without phonology. *Cognitive Neuropsychology*, *3*(1), 1–36. https://doi.org/10.1080/02643298608252668

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Shanahan, T. (1984). Nature of the reading-writing relation: An exploratory multivariate analysis. *Journal of Educational Psychology*, *76*(3), 466–477. https://doi.org/https://doi.org/10.1037/0022-0663.76.3.466

Shankweiler, D. (1994). A review of Treiman, R. (1994). Beginning to spell. *Language and Speech*, 145–147.

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In *Advances in Child Development and Behavior* (Vol. 36, pp. 31–82). https://doi.org/10.1016/S0065-2407(08)00002-5

Share, D. L. (2011). On the role of phonology in reading acquisition: The self-teaching hypothesis. In S.A. Brady, D. Braze, & C. A. Fowler (Eds.), *New directions in communication disorders research. Explaining individual differences in reading: Theory and evidence* (pp. 45–68). https://doi.org/10.4324/9780203817964

Share, D. L. (2012). Frost and fogs, or sunny skies? Orthography, reading, and misplaced optimalism. *Behavioral and Brain Sciencesral and Bra*, *35*, 263–329. https://doi.org/doi:10.1017/S0140525X12000271

Share, D. L. (2018). Foundations for a universal model of learning to read. In A. Bar-On & D. D. Ravid (Eds.), *Handbook of Communication Disorders* (pp. 437–460). Berlin: De Gruyter Mouton.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining "learning" in statistical searning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42*(3), 692–727. https://doi.org/10.1111/cogs.12556

Skinner, B. F. (1953). Some contributions of an experimental analysis of behavior to psychology as a whole. *American Psychologist*, *8*(2), 69–78. https://doi.org/10.1037/h0054118

Smith, K. L., & Pitt, M. A. (1999). Phonological and morphological influences in the syllabification of spoken words. *Journal of Memory and Language*, *41*(2), 199–222. https://doi.org/10.1006/jmla.1999.2646

Sobaco, A., Treiman, R., Peereman, R., Borchardt, G., & Pacton, S. (2015). The influence of graphotactic knowledge on adults' learning of spelling. *Memory and Cognition*, *43*(4), 593–604. https://doi.org/10.3758/s13421-014-0494-y

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*(3). https://doi.org/10.1006/jmla.2000.2783

Spence, C., Senkowski, D., & Roder, B. (2009). Crossmodal processing. *Experimental Brain Research*, *198*, 107–111. https://doi.org/10.1007/s00221-009-1973-4

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing*, *28*(4), 467–490. https://doi.org/10.1007/s11145-014-9533-0

Stadler, M. A., & Frensch, P. A. (Eds.). (1998). *Handbook of implicit learning*. Sage Publications, Inc.

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, *86*(2), 221–234. https://doi.org/10.1037//0022-0663.86.2.221

Steffler, D. J. (2001). Implicit cognition and spelling development. *Developmental Review*, *21*(2), 168–204. https://doi.org/10.1006/drev.2000.0517

Steffler, D. J. (2004). An investigation of Grade 5 children's knowledge of the doubling rule in spelling. *Journal of Research in Reading*, *27*(3), 248–264. https://doi.org/10.1111/j.1467-9817.2004.00230.x

Stein, B. E., & Meredith, A. M. (1993). *Cognitive Neuroscience. The merging of the senses*. The MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, Vol. 9, pp. 255–266. https://doi.org/10.1038/nrn2331

Stemberger, J. P. (1983). The nature of /r/ and /l/ in English: evidence from speech errors. *Journal of Phonetics*, *11*(2), 139–147. https://doi.org/10.1016/s0095-4470(19)30812-5

Stemberger, J. P. (1985). An interactive activation model of language production. In A. . Ellis (Ed.), *Progress in the Psychology of Language* (Vol. 1, p. 143186). Hillsdale, NJ: Laurence Elbaum Associates.

Stuart, M., & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition*, *30*(2), 139–181. https://doi.org/10.1016/0010-0277(88)90038-8

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, Vol. 8, pp. 432–443. https://doi.org/10.1111/j.1467-7687.2005.00432.x

Taylor, C. F., & Houghton, G. (2005). Learning artificial phonotactic constraints: Time course, durability, and relationship to natural constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1398–1416. https://doi.org/10.1037/0278-7393.31.6.1398

Taylor, J., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(1), 60–76. https://doi.org/10.1037/a0020126

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide : Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. https://doi.org/10.1037/0012-

1649.39.4.706

Thorpe, S. (2017). The intricacies of momophones. *Word Ways*, *50*(4), 303.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(5), 1131–1142. https://doi.org/10.1037/0278-7393.30.5.1131

Tong, X., Tong, X., & McBride, C. (2017). Radical sensitivity is the key to understanding Chinese character acquisition in children. *Reading and Writing*, *30*(6), 1251–1265. https://doi.org/10.1007/s11145-017-9722-8

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE-2 Examiner's Manual* (Second Edi). Austin: Pro-Ed.

Treiman, R. (1993). *Beginning to spell: A study of first grade children*. Oxford: Oxford University Press.

Treiman, R. (1998). Why spelling?: The benefits of incorporating spelling into beginning reading instruction. *Word Recognition in Beginning Literacy*, 289–313.

Treiman, R. (2017). Learning to spell words: Findings, Theories, and Issues. *Scientific Studies of Reading*. https://doi.org/10.1080/10888438.2017.1296449

Treiman, R. (2018). Teaching and learning spelling. *Child Development Perspectives*, *12*(4), 235–239. https://doi.org/10.1111/cdep.12292

Treiman, R., & Boland, K. (2017). Graphotactics and spelling: Evidence from consonant doubling. *Journal of Memory and Language*, *92*, 254–264. https://doi.org/10.1016/j.jml.2016.07.001

Treiman, R., & Kessler, B. (1995). In defense of an onset-rime syllable structure for English. *Language and Speech*, *38*(2), 127–142. https://doi.org/10.1177/002383099503800201

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology*, *98*(3), 642–652. https://doi.org/10.1037/0022-0663.98.3.642

Treiman, R., & Kessler, B. (2013). Learning to use an alphabetic writing wystem. *Language and Learning Development*, *9*(4), 317–330. https://doi.org/10.1080/15475441.2013.812016

Treiman, R., & Kessler, B. (2014). How children learn to write words. In *How Children Learn to Write Words*.

https://doi.org/10.1093/acprof:oso/9780199907977.001.0001

Treiman, R., & Kessler, B. (2015). Choosing between alternative spellings of sounds: The role of context. *Journal of Experimental Psychology: Learning Memory and Cognition*, *Advance on*. https://doi.org/http://dx.doi.org/10.1037/xlm0000225

Treiman, R., & Kessler, B. (2019). Development of context-sensitive pronunciation in reading : The case of ‹ c › and ‹ g ›. *Journal of Experimental Child Psychology*, *182*, 114–125. https://doi.org/10.1016/j.jecp.2019.02.001

Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language*, *47*, 448–468. https://doi.org/https://doi.org/10.1016/S0749-596X(02)00010-4

Treiman, R., Kessler, B., Boland, K., Clocksin, H., & Chen, Z. (2017). Statistical learning and spelling: Older prephonological spellers produce more wordlike spellings than younger prephonological spellers. *Child Development*, *89*(4), e431–e443. https://doi.org/10.1111/cdev.12893

Treiman, R., Kessler, B., & Bourassa, D. (2001). Children's own names influence their spelling. *Applied Psycholinguistics*, *22*(4), 555–570. https://doi.org/10.1017/S0142716401004040

Treiman, R., Pollo, T. C., Cardoso-Martins, C., & Kessler, B. (2013). Do young children spell words syllabically? Evidence from learners of Brazilian Portuguese. *Journal of Experimental Child Psychology*, *116*(4), 873–890. https://doi.org/10.1016/j.jecp.2013.08.002

Treiman, R., & Wolter, S. (2018). Phonological and graphotactic influences on spellers' decisions about consonant doubling. *Memory and Cognition*, *46*(4), 614–624. https://doi.org/https://doi.org/10.3758/s13421-018-0793-9 Phonological

Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. *Journal of Memory and Language*, *29*(1), 66–85. https://doi.org/10.1016/0749-596X(90)90010-W

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Valle-Arroyo, F. (1990). Spelling errors in Spanish. *Reading and Writing*, *2*(1), 83–98. https://doi.org/10.1007/BF00383375

Van Leerdam, M., Bosman, A. M. T., & Van Orden, G. C. (1998). The ecology of spelling instruction: Effective training in first grade. In P. Reitsma & L. Verhoeven (Eds.), *Problems and Interventions in Literacy Development. Neuropsychology and Cognition, vol 15* (pp. 304–320). https://doi.org/doi.org/10.1007/978-94-017-2772-3_19

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1269–1291. https://doi.org/10.1037//0096-1523.20.6.1269

Varnhagen, C. K., Boechler, P. M., & Steffler, D. J. (1999). Phonological and orthographic influences on children's vowel spelling. *Scientific Studies of Reading*, *3*(4), 363–379. https://doi.org/10.1207/s1532799xssr0304

Varnhagen, C. K., Mccallum, M., & Burstow, M. (1997). Is children's spelling naturally stage-like? *Reading and Writing: An Interdisciplinary Journal*, *9*, 451–481. https://doi.org/https://doi.org/10.1007/978-94-017-3054-9_8

Venezky, R. L. (1970). Principles for the design of practical writing systems. *Anthropological Linguistics*, *12*(7), 256–270.

Vicari, S., Finzi, A., Menghini, D., Marotta, L., Baldi, S., & Petrosini, L. (2015). Do children with developmental dyslexia have an implicit learning deficit? *J Neurol Neurosurg Psychiatry*, *76*, 1392–1397. https://doi.org/10.1136/jnnp.2004.061093

Vidal, C., Content, A., & Chetail, F. (2017). BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behavior Research Methods*, *49*(6), 2093–2112. https://doi.org/10.3758/s13428-016-0844-8

Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology*, (31), 1–7. https://doi.org/10.1016/j.cub.2020.12.017

Vujović, M., Ramscar, M., & Wonnacott, E. (n.d.). Language learning as uncertainty reduction: The role of prediction error in linguistic generalization and item-learning. *Under Review*.

Wang, H. C., Castles, A., Nickels, L., & Nation, K. (2011). Context effects on orthographic learning of regular and irregular words. *Journal of Experimental Child Psychology*, *109*(1), 39–57. https://doi.org/10.1016/j.jecp.2010.11.005

Wanzek, J., Vaughn, S., Wexler, J., Swanson, E. A., Edmonds, M., & Kim, A.

(2006). A synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with LD. *Journal of Learning Disabilities*, *39*(6), 528–543. https://doi.org/10.1177/00222194060390060501

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(2), 387–398. https://doi.org/10.1037/0278-7393.32.2.387

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2017). The procedural learning deficit hypothesis of language learning disorders : we see some problems. *Developmental Science*, 1–13. https://doi.org/10.1111/desc.12552

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The seriol model and selective literature review. *Psychonomic Bulletin and Review*, *8*(2). https://doi.org/10.3758/BF03196158

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test (WRAT4).*

Wimmer, H., & Hummer, P. (1990). How German-speaking first graders read and spell: Doubts on the importance of the logographic stage. *Applied Psycholinguistics*, *11*(4), 349–368. https://doi.org/10.1017/S0142716400009620

Wonnacott, E. (2011). Balancing generalization and lexical conservatism : An artificial language study with child learners. *Journal of Memory and Language*, *65*(1), 1–14. https://doi.org/10.1016/j.jml.2011.03.001

Wonnacott, E., Brown, H., & 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*, 36–48. https://doi.org/10.1016/j.jml.2017.01.005

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure : Distributional learning in a miniature language. *Cognitive Psychology*, *56*, 165–209. https://doi.org/10.1016/j.cogpsych.2007.04.002

Wright, D. M., & Ehri, L. C. (2007). Beginners remember orthography when they learn to read words: The case of doubled letters. *Applied Psycholinguistics*, *28*(1), 115–133. https://doi.org/10.1017/S0142716407070063

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference.

Psychological Review, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yang, J., & Li, P. (2012). Brain networks of explicit and implicit learning. *PLoS ONE*, *7*(8), e42993. https://doi.org/10.1371/journal.pone.0042993

Yates, M., Locker, L., & Simpson, G. B. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Memory and Cognition*, *31*(6), 856–866. https://doi.org/10.3758/BF03196440

Yoncheva, Y. N., Blau, V. C., Maurer, U., & McCandliss, B. D. (2010). Attentional focus during learning impacts N170 ERP responses to an artificial script. *Developmental Neuropsychology*, *35*(4), 423–445. https://doi.org/10.1080/87565641.2010.480918

Zhang, L., & Treiman, R. (2015). Writing dinosaur large and mosquito small: Prephonological spellers' use of semantic information. *Scientific Studies of Reading*, *19*(6), 434–445. https://doi.org/10.1080/10888438.2015.1072820

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, Vol. 131, pp. 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 547–559. https://doi.org/10.1037/0096-1523.27.3.547

Appendix A

Modelling the H_1, Method C

A main effect with two levels, a_1 and a_2, is given as follows:

$$a = (a_1 - \log(\text{chance})) - (a_2 - \log(\text{chance}))$$

Applying our assumption for a plausible maximum difference to the above calculation therefore gives:

$$a = (a_1 - \log(0.5)) - (a2 - \log(0.5))$$

$$= (a_1 - 0) - (0 - 0)$$

$$= a1$$

The grand mean is given as follows:

$$\bar{e} = \frac{(a1 - \log(0.5)) + (a2 - \log(0.5))}{2}$$

$$2\bar{e} = a1$$

$$= a$$

Therefore, the main effect of *a* is equal to twice the grand mean (ē). We therefore set *x* (the estimate for *a*) to half this value, since the maximum for the half normal is equal to 2SD. Since we use centred coding in the mixed-effects models, twice the grand-mean corresponds to twice the intercept estimate. Therefore, x is equal to the beta coefficient for the intercept from the mixed-model.

Appendix B

Links to preregistrations, data analyses scripts and data files

@ **Preregistrations**

a. Experiment 3: https://osf.io/mn254
b. Experiment 4: https://osf.io/kz26g
c. Experiment 5: https://osf.io/m76ck

@ **Data Analyses Scripts**

a. Samara, Singh, and Wonnacott (2019): https://rpubs.com/DSingh/Body_Rime
b. Experiment 1: https://rpubs.com/DSingh/Phono_Grapho
c. Experiment 2: https://rpubs.com/DSingh/SemiartificialPositional
d. Experiments 3, 4, and 5: https://osf.io/tsxf3/ and https://rpubs.com/DSingh/Graphotactics
e. Experiments 6 and 7: https://rpubs.com/DSingh/WordLearning

@ **Data**

All data files can be found at: https://osf.io/gkjvd/?view_only=40316c03d01b4af9863f51d8cee1c6ad, and individual data folders are:

- Experiment 1: children: https://osf.io/8fv9d/; adults https://osf.io/r9yzf/
- Experiment 2: children: https://osf.io/c2hqk/; adults: https://osf.io/b3x5n/
- Experiment 3: children: https://osf.io/6fngu/; adults: https://osf.io/4r6vs/
- Experiment 4: children: https://osf.io/3vt4d/; Adults: same file as Experiment 3: https://osf.io/4r6vs/
- Experiment 5: children: https://osf.io/9ae6u/
- Experiment 6 and Experiment 7, adults: https://osf.io/pes39/

Appendix C

Stimuli for Experiment 1

	list1	list2	list3	list4
exposure	dot	don	det	den
exposure	dop	dos	dep	des
exposure	mon	mot	men	met
exposure	mos	mop	mes	mep
exposure	lot	lon	len	let
exposure	lop	los	les	lep
exposure	fon	fot	fet	fen
exposure	fos	fop	fep	fes
exposure	tem	ted	tod	tom
exposure	tef	tel	tol	tof
exposure	ned	nem	nom	nod
exposure	nel	nef	nof	nol
exposure	pem	ped	pom	pod
exposure	pef	pel	pof	pol
exposure	sed	sem	sod	som
exposure	sel	sef	sol	sof
legal_unseen	don	dot	den	det
legal_unseen	dos	dop	des	dep
legal_unseen	mot	mon	met	men
legal_unseen	mop	mos	mep	mes
legal_unseen	lon	lot	let	len
legal_unseen	los	lop	lep	les
legal_unseen	fot	fon	fen	fet
legal_unseen	fop	fos	fes	fep
legal_unseen	ted	tem	tom	tod
legal_unseen	tel	tef	tof	tol
legal_unseen	nem	ned	nod	nom
legal_unseen	nef	nel	nol	nof
legal_unseen	ped	pem	pod	pom
legal_unseen	pel	pef	pol	pof
legal_unseen	sem	sed	som	sod
legal_unseen	sef	sel	sof	sol
illegal	det	den	dot	don
illegal	dep	des	dop	dos
illegal	men	met	mon	mot
illegal	mes	mep	mos	mop
illegal	len	let	lot	lon
illegal	les	lep	lop	los
illegal	fet	fen	fon	fot
illegal	fep	fes	fos	fop
illegal	tod	tom	tem	ted
illegal	tol	tof	tef	tel
illegal	nom	nod	ned	nem
illegal	nof	nol	nel	nef
illegal	pom	pod	pem	ped
illegal	pof	pol	pef	pel
illegal	sod	som	sed	sem
illegal	sol	sof	sel	sef

Appendix D

Stimuli for Experiment 2, Version 1

	list1	list2	list3	list4
exposure	∖⋂cフ	∖⋂c⋖	フc∖⋂	⋖c∖⋂
exposure	∖⋂g⋖	∖⋂gフ	⋖g∖⋂	フg∖⋂
exposure	∖⋂vフ	∖⋂v⋖	フv∖⋂	⋖v∖⋂
exposure	∖⋂s⋖	∖⋂sフ	⋖s∖⋂	フs∖⋂
exposure	⦄c⋖	⦄cフ	⋖c⦄	フc⦄
exposure	⦄gフ	⦄g⋖	フg⦄	⋖g⦄
exposure	⦄v⋖	⦄vフ	⋖v⦄	フv⦄
exposure	⦄sフ	⦄s⋖	フs⦄	⋖s⦄
legal_unseen	∖⋂c⋖	∖⋂cフ	⋖c∖⋂	フc∖⋂
legal_unseen	∖⋂gフ	∖⋂g⋖	フg∖⋂	⋖g∖⋂
legal_unseen	∖⋂v⋖	∖⋂vフ	⋖v∖⋂	フv∖⋂
legal_unseen	∖⋂sフ	∖⋂s⋖	フs∖⋂	⋖s∖⋂
legal_unseen	⦄cフ	⦄c⋖	フc⦄	⋖c⦄
legal_unseen	⦄g⋖	⦄gフ	⋖g⦄	フg⦄
legal_unseen	⦄vフ	⦄v⋖	フv⦄	⋖v⦄
legal_unseen	⦄s⋖	⦄sフ	⋖s⦄	フs⦄
illegal	フc∖⋂	⋖c∖⋂	∖⋂cフ	∖⋂c⋖
illegal	⋖g∖⋂	フg∖⋂	∖⋂g⋖	∖⋂gフ
illegal	フv∖⋂	⋖v∖⋂	∖⋂vフ	∖⋂v⋖
illegal	⋖s∖⋂	フs∖⋂	∖⋂s⋖	∖⋂sフ
illegal	⋖c⦄	フc⦄	⦄c⋖	⦄cフ
illegal	フg⦄	⋖g⦄	⦄gフ	⦄g⋖
illegal	⋖v⦄	フv⦄	⦄v⋖	⦄vフ
illegal	フs⦄	⋖s⦄	⦄sフ	⦄s⋖
legal_new	∖⋂r⋖	∖⋂r⋖	⋖r∖⋂	⋖r∖⋂
legal_new	∖⋂tフ	∖⋂tフ	フt∖⋂	フt∖⋂
legal_new	⦄k⋖	⦄k⋖	⋖k⦄	⋖k⦄
legal_new	⦄dフ	⦄dフ	フd⦄	フd⦄
illegal_new	⋖r∖⋂	⋖r∖⋂	∖⋂r⋖	∖⋂r⋖
illegal_new	フt∖⋂	フt∖⋂	∖⋂tフ	∖⋂tフ
illegal_new	⋖k⦄	⋖k⦄	⦄k⋖	⦄k⋖
illegal_new	フd⦄	フd⦄	⦄dフ	⦄dフ

Appendix E

Awareness Questionnaire

Were you aware of any patterns/rules that exist in Zorib's orthography?

Please Select...

Please state how you made your choices when filling in the blanks to make words in Zorib's language?

Please state how you made your choices when deciding which words can exist in Zorib's language and which cannot.

Thank you for participating in this experiment. You are helping us find out how children and adults learn to spell when they are not explicitly aware of the patterns and rules that exist in a language.

Next

Appendix F

Stimuli for Experiment 3 and Experiment 5

	list1	list2	list3	list4
exposure	deff	dess	def	des
exposure	dell	dett	del	det
exposure	duf	dus	duff	duss
exposure	dul	dut	dull	dutt
exposure	gess	geff	ges	gef
exposure	gett	gell	get	gel
exposure	gus	guf	guss	guff
exposure	gut	gul	gutt	gull
exposure	meff	mess	mef	mes
exposure	mell	mett	mel	met
exposure	muf	mus	muff	muss
exposure	mul	mut	mull	mutt
exposure	ress	reff	res	ref
exposure	rett	rell	ret	rel
exposure	rus	ruf	russ	ruff
exposure	rut	rul	rutt	rull
legal_unseen	dess	deff	des	def
legal_unseen	dett	dell	det	del
legal_unseen	dus	duf	duss	duff
legal_unseen	dut	dul	dutt	dull
legal_unseen	geff	gess	gef	ges
legal_unseen	gell	gett	gel	get
legal_unseen	guf	gus	guff	guss
legal_unseen	gul	gut	gull	gutt
legal_unseen	mess	meff	mes	mef
legal_unseen	mett	mell	met	mel
legal_unseen	mus	muf	muss	muff
legal_unseen	mut	mul	mutt	mull
legal_unseen	reff	ress	ref	res
legal_unseen	rell	rett	rel	ret
legal_unseen	ruf	rus	ruff	russ
legal_unseen	rul	rut	rull	rutt
illegal	des	def	dess	deff
illegal	det	del	dett	dell
illegal	duss	duff	dus	duf
illegal	dutt	dull	dut	dul
illegal	gef	ges	geff	gess
illegal	gel	get	gell	gett
illegal	guff	guss	guf	gus
illegal	gull	gutt	gul	gut
illegal	mes	mef	mess	meff
illegal	met	mel	mett	mell
illegal	muss	muff	mus	muf
illegal	mutt	mull	mut	mul
illegal	ref	res	reff	ress
illegal	rel	ret	rell	rett
illegal	ruff	russ	ruf	rus
illegal	rull	rutt	rul	rut

Appendix G

Stimuli for Experiment 4, Experiment 6, and Experiment 7

	list1	list2	list3	list4
exposure	des	deff	dess	def
exposure	det	dell	dett	del
exposure	duss	duf	dus	duff
exposure	dutt	dul	dut	dull
exposure	geff	ges	gef	gess
exposure	gell	get	gel	gett
exposure	guf	guss	guff	gus
exposure	gul	gutt	gull	gut
exposure	mes	meff	mess	mef
exposure	met	mell	mett	mel
exposure	muss	muf	mus	muff
exposure	mutt	mul	mut	mull
exposure	reff	res	ref	ress
exposure	rell	ret	rel	rett
exposure	ruf	russ	ruff	rus
exposure	rul	rutt	rull	rut
legal_unseen	deff	des	def	dess
legal_unseen	dell	det	del	dett
legal_unseen	duf	duss	duff	dus
legal_unseen	dul	dutt	dull	dut
legal_unseen	ges	geff	gess	gef
legal_unseen	get	gell	gett	gel
legal_unseen	guss	guf	gus	guff
legal_unseen	gutt	gul	gut	gull
legal_unseen	meff	mes	mef	mess
legal_unseen	mell	met	mel	mett
legal_unseen	muf	muss	muff	mus
legal_unseen	mul	mutt	mull	mut
legal_unseen	res	reff	ress	ref
legal_unseen	ret	rell	rett	rel
legal_unseen	russ	ruf	rus	ruff
legal_unseen	rutt	rul	rut	rull
illegal	def	dess	deff	des
illegal	del	dett	dell	det
illegal	duff	dus	duf	duss
illegal	dull	dut	dul	dutt
illegal	gess	gef	ges	geff
illegal	gett	gel	get	gell
illegal	gus	guff	guss	guf
illegal	gut	gull	gutt	gul
illegal	mef	mess	meff	mes
illegal	mel	mett	mell	met
illegal	muff	mus	muf	muss
illegal	mull	mut	mul	mutt
illegal	ress	ref	res	reff
illegal	rett	rel	ret	rell
illegal	rus	ruff	russ	ruf
illegal	rut	rull	rutt	rul

Appendix H

Kuder-Richardson Formula 20 (KR-20) Split-half Reliability for Fill-in-the-blanks and Legality Judgment for Children and Adults' by list.

Experiment 1

Age Group	Procedure	List	$\rho_{(KR-20)}$
Children			
	Fill-in-the-blanks		
		List 1	0.65
		List 2	-0.13
		List 3	-0.23
		List 4	0.20
	Legality Judgment		
		List 1	-0.40
		List 2	-0.31
		List 3	-0.53
		List 4	-0.52
Adults			
	Fill-in-the-blanks		
		List 1	0.15
		List 2	0.57
		List 3	-1.41
		List 4	0.67
	Legality Judgment		
		List 1	-0.36
		List 2	0.51
		List 3	0.87
		List 4	0.69

Experiments 3, 4, and 5

Age Group	Experiment	List	$\rho_{(KR-20)}$
Children			
	Experiment 3		
		List 1	0.15
		List 2	-0.21
		List 3	-0.92
		List 4	0.09
	Experiment 4		
		List 1	-0.94
		List 2	0.45
		List 3	0.14
		List 4	-0.09
	Experiment 5		
		List 1	0.75
		List 2	0.61
		List 3	0.89
		List 4	0.85
Adults			
	Experiment 3		
		List 1	0.88
		List 2	0.73
		List 3	0.84
		List 4	0.88
	Experiment 4		
		List 1	-0.80
		List 2	-1.19
		List 3	-0.15
		List 4	-0.85

Appendix I

Nonword-object allocation for stimuli in Experiments 6 and 7