LANGUAGE, COGNITION AND NEUROSCIENCE
2026, VOL. 41, NO. 1, 110–135
https://doi.org/10.1080/23273798.2025.2576908

REGULAR ARTICLE

OPEN ACCESS

Cognitive abilities underlying the earliest stages of second language acquisition: an artificial language study

Panagiotis Kenanidis[a], Miquel Llompart[a,b], Diana Pili-Moss[c] and Ewa Dąbrowska[a,d]

[a]Chair of Language and Cognition, Department of English and American Studies, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; [b]Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain; [c]Institute of English Studies, Faculty of Education, Leuphana Universität Lüneburg, Lüneburg, Germany; [d]Department of English Language and Linguistics, University of Birmingham, Birmingham, UK

ABSTRACT

A central issue in second language (L2) acquisition concerns how explicit learning (EL) and implicit statistical learning (ISL) aptitudes contribute during the earliest stages of learning, and whether the effect of EL precedes that of ISL, as traditionally assumed, or can instead follow it. This paper explores these possibilities by tracking the contributions of these two aptitudes, as well as sustained attention (SA) to vocabulary and grammar learning across five sessions of exposure to an artificial language. Results indicated that vocabulary and grammar learning were modulated by EL and SA, with ISL additionally accounting for variance in vocabulary learning. Crucially, contrary to the standard view, for both grammar and vocabulary, the ISL effects were most pronounced early on, whereas the EL effects increased over time. These results underscore the dynamic interplay between the two aptitudes in early L2 acquisition and highlight the time-varying nature of their contributions.

ARTICLE HISTORY
Received 27 February 2024
Accepted 1 October 2025

KEYWORDS
Second language acquisition; artificial language; implicit statistical leaning; explicit learning; individual differences

Introduction

It is widely accepted that second language (L2) learning outcomes are subject to large interindividual variation (Dörnyei, 2006; Kidd et al., 2018). The study of individual differences offers a powerful tool for gaining insights into the mechanisms underlying language development and for testing the predictions of different models of L2 acquisition. In the pursuit of identifying the cognitive abilities and qualities that make someone adept at learning a new language, researchers have shown long-standing theoretical and empirical attention to the roles of aptitudes for explicit learning and implicit statistical learning[1] (Andringa & Rebuschat, 2015; Granena, 2013; Li & DeKeyser, 2021). Contemporary cognitive models of L2 acquisition posit that these factors may exert their influence differentially across domains of L2 learning and stages of development (DeKeyser, 2015; Ellis & Wulff, 2020; Ullman, 2020). Within this framework, a central theoretical issue concerns whether explicit aptitude is primarily involved in initial L2 learning, with implicit aptitude contributing in more advanced phases, as traditionally assumed (e.g. Li & DeKeyser, 2021), or whether implicit processes can exert an influence from the outset, setting the stage for subsequent explicit learning (see Godfroid, 2022). Yet, so far, surprisingly few studies have examined how individual variation in these abilities affects learning outcomes over time, let alone considered the temporal dynamics of their contributions.

To examine how individual difference variables impact L2 learning and processing, a growing body of studies has used artificial language learning paradigms (Ettlinger et al., 2016). Such paradigms afford fine-grained control over the properties of the linguistic input and exposure conditions, while also limiting the influence of extraneous factors such as prior language exposure and experience. However, to date, the vast majority of studies have either tested L2 learning outcomes after only a single session of exposure to the artificial language or have focused on a particular subset of cognitive abilities, specifically declarative and procedural memory (e.g. Morgan-Short et al., 2014; Pili-Moss et al., 2020; Walker et al., 2020), leaving the temporal interplay of EL and ISL aptitudes largely unexplored.

CONTACT Panagiotis Kenanidis panos.kenanidis@fau.de Chair of Language and Cognition, Department of English and American Studies, Friedrich Alexander University Erlangen-Nuremberg, Bismarckstraße 6, 91054 Erlangen, Germany

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23273798.2025.2576908.

Using an artificial language paradigm, the present study set out to expand the current literature on cognitive individual differences in adult L2 learning by investigating the extent to which implicit statistical learning and explicit learning aptitudes underpin L2 vocabulary and grammar learning outcomes at various time points of exposure. Furthermore, the role of sustained attention was also considered. Although closely linked to L2 acquisition, the predictive power of this ability has been insufficiently examined in artificial language learning contexts. Importantly, while sustained attention is essential for explicit learning, the extent to which implicit statistical learning relies on it remains debated (Thiessen et al., 2013; Williams, 2009). Thus, by examining the individual and joint contributions of these cognitive abilities to learning success, this study seeks to contribute to a better understanding of their timing and extent of engagement during the earliest stages of L2 learning.

Implicit statistical and explicit learning in L2 learning

Implicit statistical learning (henceforth ISL) is broadly defined as the ability to detect and learn complex regularities in the environment and is characterised as a process that happens incidentally, without conscious intention and instructions to learn and without awareness of what is being learned (Christiansen, 2019; Frost et al., 2019; Granena, 2020). ISL proceeds rapidly and automatically when learners are exposed to input but leads to a gradual accumulation of knowledge through repeated exposure (Andringa & Rebuschat, 2015; Ellis, 2005). It is typically thought to occur in incidental exposure contexts where participants are not informed about the learning target and the presence of a subsequent test phase (Granena & Yilmaz, 2018; Williams, 2009) and can result in the acquisition of implicit knowledge (Rebuschat, 2013). On the other hand, explicit learning (EL) is generally conceptualised as a conscious, attention-driven operation at the point of learning, which is essential for the initial conscious registration of new information (Ellis, 2005; Hulstijn, 2005). In contrast to ISL, it is commonly observed under intentional exposure conditions in which learners engage in deliberate hypothesis testing and memorisation of patterns or rules and contributes primarily to the emergence of explicit knowledge, which is often signified by its verbalizability (Rebuschat & Williams, 2012; Williams, 2009).

In the context of adult second language (L2) acquisition, a number of scholars have acknowledged the significant role of EL mechanisms. This is because L2 acquisition is customarily the outcome of a conscious decision and commonly takes place in instructional settings, which promotes explicit learning (Granena & Yilmaz, 2018). Moreover, abilities related to ISL, such as procedural memory, may be less readily available in adults, as they are thought to be affected by maturational constraints (DeKeyser, 2022; Paradis, 2009) and previous experiences (e.g. L1-tuned attention; Ellis, 2006; MacWhinney, 2005). Nevertheless, rather than functioning in isolation, the involvement of both EL and ISL abilities is required for successfully acquiring a language later in life (DeKeyser, 2003; Ellis, 2022). Indeed, there is evidence that the two operate in parallel. The dynamic nature of the interaction between EL and ISL and language acquisition can be found in the fact that they exert their influence on L2 learning outcomes in a complementary way. For example, EL appears to be more effective at promoting learning of simple grammatical structures, while ISL facilitates learning of more difficult ones (Li & DeKeyser, 2021; Suzuki et al., 2023; though see Antoniou et al., 2016; for an opposite pattern of results in incidental contexts). Furthermore, it has been argued that EL and ISL processes have time-varying effects, with the former being involved in initial L2 learning (Li, 2016) and the latter becoming more important in advanced stages (Li & DeKeyser, 2021). Importantly, however, this claim has yet to be tested in longitudinal studies.

The idea that EL and ISL are differentially engaged at various stages of learning appears to align with proposals from studies exploring the roles of declarative and procedural memory in language acquisition. This alignment stems from the close overlap between ISL and procedural memory, and EL and declarative memory, respectively. These two memory systems are at the centre of attention of several influential theoretical models of L2 learning (Skill Acquisition Theory: DeKeyser, 2015; Paradis' neurolinguistic theory of bilingualism: Paradis, 2009; Declarative/Procedural model: Ullman, 2020), which notwithstanding some differences, all converge in predicting that at early stages of L2 learning, adults rely on their declarative memory abilities, and that with increased exposure to the L2, grammatical, but not necessarily lexical, knowledge may gradually depend more on procedural memory. However, although the distinction between EL and ISL is often equated with that between declarative and procedural memory, the terms are not isomorphic. The former distinction is based on the presence or absence of conscious awareness at the point of learning, whereas the latter refers to the brain structures underlying the two memory systems (Reber, 2008; Ullman, 2020). Additionally, it has been argued that both declarative and procedural memory may be actively involved in ISL (Batterink et al., 2019; Reber, 2013), while declarative memory appears to be the only system that underlies EL (Ullman, 2016, 2020).

Hence, given the lack of a perfect overlap between the two memory systems and the two learning processes, and since declarative memory may be implicated in some aspects of ISL, the relationship between ISL and EL could be more synergistic or reciprocal, rather than sequential, particularly during the earlier stages of exposure. One possibility is that initial EL or noticing may be required for promoting further ISL (Ellis, 2005; Schmidt, 1990). A second possibility is that, even before noticing takes place, early, often short-lived, ISL processes could trigger subsequent explicit processing (Godfroid, 2022; MacWhinney, 2022), potentially through detection of unexpected events during learning or shifts in attentional focus (Haider & Frensch, 2005). Therefore, the main aim of this study was to contribute to the line of research on ISL and EL by testing the dynamic interactions between them, and how they predict language outcomes at different time points during the very earliest stages of the L2 learning process. Understanding their independent contributions and interactions over the time course of L2 acquisition holds great promise for advancing the knowledge about their role in L2 acquisition (Batterink et al., 2019), as well as for informing models of L2 acquisition.

Implicit statistical and explicit learning as predictors of L2 learning

ISL and EL are widely acknowledged as key individual factors influencing L2 acquisition and processing. Individuals with better ISL abilities are found to be better at segmenting word boundaries, extracting chunks and (grammatical) regularities from linguistic input, and learning new form-meaning mappings, particularly in uninstructed contexts (e.g. Brooks & Kempe, 2013; Misyak & Christiansen, 2012; Speciale et al., 2004; see Armstrong et al., 2017; Siegelman, 2020, for reviews, and Ren et al., 2023, for a meta-analysis). Conversely, learners with strong EL skills are more adept at consciously reflecting on the new linguistic content (typically, upon receiving instructions to do so), with the objective of inferring the grammatical functions of linguistic elements, deriving rules, and extracting arbitrary associations (e.g. sounds and meanings) from the input and retaining them in memory (Li, 2016; Skehan, 2002).

Still, capturing the contribution of individual differences in these abilities to language acquisition has been far from straightforward. This applies particularly to ISL. Initially, ISL was thought to be a relatively stable cognitive capacity across individuals (Reber, 1993); however, this assertion was later revised (Reber & Allen, 2000), with subsequent studies affirming that significant variability exists among individuals (Kalra et al., 2019; Kaufman et al., 2010). Complicating matters further, so far, there seems to be little evidence that ISL functions as a unitary construct operating similarly across different domains and modalities (see Frost et al., 2019; Siegelman, 2020 for discussion); that is being good at detecting embedded statistical regularities in aural input does not necessarily imply equal success when materials are presented in a different modality or when different types of responses are required (e.g. motor/non-motor, verbal/non-verbal; Perruchet, 2021). As a result, tasks used to assess ISL often fail to correlate with each other, yielding inconsistent findings (Buffington et al., 2021; Godfroid & Kim, 2021).

Measuring aptitude for EL, on the other hand, poses fewer challenges. Study participation per se can prompt participants to maintain a high level of alertness and employ deliberate hypothesis testing and, given that EL requires minimal exposure, its effects can be readily observed even in experiments of short duration (DeKeyser, 2003). Though also multi-componential in nature (Skehan, 2002), positive correlations between explicit tasks are typically reported (Buffington et al., 2021; Granena, 2019) – as well as with other conscious cognitive abilities, such as working memory (Bo et al., 2009; Frensch & Miner, 1994). Finally, a challenge common to both EL and ISL is that efforts to map the relationship between these abilities and language outcomes may be confounded by the prior knowledge learners bring into the learning task. For instance, differences in previous exposure to specific grammatical structures and L2 proficiency, often difficult to control, may affect the extent to which learners will rely on ISL or EL mechanisms (see Godfroid & Kim, 2021).

One approach to addressing this last issue is through using artificial language paradigms, which allow researchers to fully control both the type and amount of input participants receive, minimising the influence of previous experience (Morgan-Short, 2020). Most pertinent to our aim of examining the time-varying effects of ISL and EL on L2 acquisition, earlier studies employing such paradigms have sought to link individual differences in declarative memory and procedural memory to variation in L2 grammar across different stages of exposure. Despite utilising a range of language measures (e.g. grammaticality judgment tests, picture selection tasks) and various implicit sequence or probabilistic learning tasks assessing procedural memory, results from this line of research find positive associations between declarative memory and L2 grammar learning at earlier stages (Hamrick, 2015; Morgan-Short et al., 2014; Pili-Moss, 2022; Pili-Moss et al., 2020;

Walker et al., 2020) and, in some cases, a stronger relationship between procedural memory and language outcomes after sufficient exposure/practice (Hamrick, 2015; Morgan-Short et al., 2014; cf. Pili-Moss et al., 2020; for reviews, see Hamrick et al., 2018; Morgan-Short et al., 2022). Notably, these studies examined grammar learning in uninstructed conditions. Although participants received training in vocabulary, no metalinguistic explanation about the language was offered, in an attempt to approximate more natural/immersion-like settings and to limit over-reliance on explicit, declarative memory. Such a design seems to also provide a promising framework for investigating the dynamic effects of ISL and EL on the L2 acquisition process.

In contrast to grammar, fewer artificial language studies have tested the associations between individual differences in ISL and EL, or procedural and declarative memory, and L2 vocabulary across multiple sessions. An exception is Walker et al. (2020), who probed the roles of procedural and declarative memory in novel lexical learning at two time points: immediately after exposure and following a 24-hour delay. In this study, participants were not explicitly trained on the language; instead, they had to extract and acquire novel words through cross-situational statistics. The findings revealed a nuanced relationship between the two memory abilities and learning outcomes. On day 1, procedural memory was predictive of success with nouns and adjectives, while declarative memory was associated with verb learning. By day 2, however, declarative memory was also linked to improved accuracy in learning nouns and adjectives. The authors interpreted these findings as indicative of a shift in explicit attention from the well-acquired verbs on day 1 to the other lexical categories. These results not only underscore the strong role of explicit, conscious processes in novel vocabulary acquisition (Ellis, 2005; Ullman, 2020) but also provide evidence for a more synergistic relationship between explicit and implicit processes. Further research is, therefore, warranted to determine the role of ISL and EL in the early stages of L2 lexical learning.

Sustained attention in L2 learning

Sustained attention is defined as the ability to maintain alertness and focus on a particular target over an extended period of time (Langner & Eickhoff, 2013; Mirsky et al., 1991). It is regarded as a lower-order attentional function that contributes to other higher level cognitive constructs, such as working memory and EL. Sustaining attention to a task is thought to be crucial for effective learning (Wilson & Korn, 2007). However, conscious stimulus processing over time is effortful and, consequently, may lead individuals to experiencing momentary fluctuations and lapses of attention, causing them to briefly disengage from the task at hand. Such fluctuations are reflected in a decline in task performance and can be manifested as slower reaction or reading times or increased errors (Esterman & Rothlein, 2019; Smallwood & Schooler, 2006).

Previous research has demonstrated that learners' ability to sustain attention over time correlates positively with performance in language production and comprehension tasks (Hubbard & Federmeier, 2021; Jongman et al., 2015). To date, however, the largest body of research on the role of sustained attention in language acquisition has focused either on children with developmental language disorder or on comparing monolinguals and bilinguals with the aim of measuring the impact of bilingualism on cognition. There has been much less research on how sustained attention contributes to the early stages of L2 acquisition in adults and, to the best of our knowledge, its contribution has not been systematically investigated within the context of artificial language learning. There is some evidence that allocating attention to novel linguistic items and constructions results in greater vocabulary and grammar learning (Godfroid et al., 2013; Indrarathne & Kormos, 2017). Nevertheless, the strength of the effect of sustained attention on the retention and learning of different aspects of an L2 during the initial stages of exposure is still unclear. An important aspect of examining sustained attention is that it also allows for exploring its potential associations with EL and ISL and ascertaining whether individual differences in this ability might influence their effect on language learning outcomes. While EL seems to rely heavily on attentional resources, the extent to which the more automatic ISL requires attention is a still a matter of ongoing debate (Kaufman et al., 2010; Thiessen et al., 2013; West et al., 2021; Williams, 2009).

The current study

Although the study of individual differences in L2 acquisition has received extensive attention (Li et al., 2022), earlier work has largely focused on a particular set of variables (mainly working, declarative, and procedural memory), leaving other abilities less explored. Additionally, a large body of research is based on natural languages, which despite their high ecological validity, may introduce some confounding variables (e.g. input quantity and level of proficiency), thus making it difficult to establish clear causal relationships between L2 achievement and different cognitive mechanisms. Such confounds can be mitigated by using artificial language paradigms. However, to date, the length of most experiments is often confined to one or two sessions, providing just a limited number of data points to trace learning and thereby constraining the understanding of how individual difference variables are involved at different stages of learning.

To address these gaps, the current study employs a new fully artificial language paradigm/design in an attempt to explore how learners' individual differences in aptitudes for ISL and EL, and sustained attention influence the early development of L2 vocabulary and grammar, how these abilities interact, and whether their effects change as a function of exposure time (Li & DeKeyser, 2021). Although, as mentioned earlier, both ISL and EL are comprised of multiple components, not all of those components were targeted here. ISL was measured through the visual statistical learning (VSL) task developed by Siegelman et al. (2017). This task was selected for two main reasons: (a) it is characterised by good test-retest reliability and internal consistency, and (b) it correlates well with the auditory statistical learning task, which measures ISL in the auditory modality (Godfroid & Kim, 2021; Siegelman et al., 2018), thereby accounting, to some extent, for the componentiality of ISL.

On the other hand, EL was measured through a sequence learning task, which mirrored the VSL task, with key differences: participants were informed about the learning targets, provided with online visual cues indicating sequence boundaries, and were given explicit instructions to memorise the sequences. Doing so allowed us to compare the impact of ISL and EL abilities on language learning by testing them using very similar tasks, instead of employing tasks that tap into different modalities (e.g. serial reaction time, henceforth SRT, task and (verbal) paired associates, respectively). However, it should be noted that rather than directly assessing the absence or presence of awareness at the point of learning, these tasks primarily capture aptitudes for learning under conditions that encourage either implicit (or, at least, incidental) detection and extraction of patterns or guided, conscious memorisation, without isolating awareness as a defining feature. Finally, practical reasons contributed to the selection of these tasks. Since data collection took place online, we opted for tasks that do not involve a motor component and, hence, do not rely on reaction times, like the SRT and alternating SRT do, ensuring that differences in participants' network quality would not affect the results.

The data presented here were collected as part of a larger online study investigating incidental learning of novel grammatical structures using an artificial language paradigm in L1 English and L1 German speakers (see Kenanidis et al., 2023 for a discussion on the role of L1 experience in novel grammar learning). The target language, Kepidalo, uses case marking and has flexible word order, making inflectional morphology a critical cue for interpreting transitive sentences. Participants completed five experimental sessions during which they received vocabulary training while being incidentally exposed to the novel grammar and were assessed on their knowledge of the novel grammar. Following previous artificial language studies (e.g. Amato & MacDonald, 2010; Morgan-Short et al., 2014; Pili-Moss, 2022), vocabulary training was initially provided, incorporating corrective feedback in order to "bootstrap" the learning of grammar. Hence, vocabulary learning occurred under more intentional conditions, with feedback, while grammar learning took place under more incidental conditions, without direct instruction or awareness of the rules.

The study was guided by the following questions:

RQ1: Which cognitive abilities (EL, ISL, sustained attention) support early L2 learning?

RQ2: Do the effects of EL and ISL on vocabulary learning under intentional exposure conditions remain stable or change during the earliest stages of L2 learning?

RQ3: Do the effects of EL and ISL on grammar learning under incidental exposure conditions remain stable or change during the earliest stages of L2 learning?

Regarding our first research question, based on earlier findings, positive relationships between all three predictors and artificial language learning outcomes were expected. Our predictions for the next two research questions were more tentative. Overall, since aptitude for EL is posited to play a prominent role in L2 acquisition, we expected to find a strong effect of EL on both vocabulary and grammar learning throughout the task, but particularly at the earliest stages of exposure, where more hypothesis testing is likely needed. Concerning ISL, two opposite hypotheses were considered. If the effect of ISL follows a pattern similar to that found for procedural memory, then its influence should grow stronger over time. Alternatively, if the facilitative role of ISL on L2 learning becomes more evident in the face of increased cognitive complexity/difficulty, here referring to contextual difficulty arising from the lack of instructions and familiarity with the artificial language rather than grammatical complexity (DeKeyser, 2016; Housen & Simoens, 2016), then ISL will serve as a better predictor at early stages of exposure.

Method

The current study is an extension of Kenanidis et al. (2023), presenting original analyses on the role of individual differences predictors in shaping the learning outcomes observed in that earlier work. We proceed by providing an overview of the participants in the original study, the artificial language paradigm that learners were exposed to and the tasks that were administered. All artificial language learning tasks are described with reference to the original study.

Participants

Forty-one native speakers of English (mean age = 22.02 years, SD = 4.17) took part in the study. They were native speakers of English that were raised in monolingual families in the UK, and their level of education ranged from 12 to 21 years (mean = 15.9, SD = 1.8). Participants were recruited online through Prolific and through advertisements on social networks (Facebook and Twitter) and were paid 60.70 GBP for their participation. Data from these latter participants were collected via a link sent by email. Informed consent was obtained online from all participants and the study was conducted in accordance with the Declaration of Helsinki.

Ten additional participants completed more than one session but their submissions were rejected for one of the following reasons: (1) they were not UK residents as indicated by their computer-recorded local timestamps (*N* = 3), (2) they did not complete all five sessions of the study (*N* = 6) and (3) they were not actively engaged with the artificial language learning tasks (as evidenced by extremely short reaction times during the training blocks; *N* = 1).[2]

Materials

Artificial language paradigm

In order to create a meaningful context for learning the novel language, participants were invited to complete an online computer-based game, during which they were tasked with saving the earth from an alien attack. Participants had to learn a new language and travel to an imaginary galaxy where they would complete a set of cognitive challenges (the individual differences tasks) to collect four weapons that would allow them to accomplish their goal. We decided to gamify the experiment since the addition of gamification elements, such as meaningful contexts, has been shown to improve learning achievement and motivation (Su & Cheng, 2015) and to lead to higher participant retention rates (Krause et al., 2015).

The target artificial language employed in this study, Kepidalo, consists of 14 disyllabic pseudowords (Appendix A): 8 nouns, half of which end in *-i*, and half of which in *-a*, 4 verbs, and 2 adjectives that were marked with different suffixes depending on the class of the noun to which they referred (*-i* or *-a*).[3]

Kepidalo displayed variable word order, with Subject-Object-Verb (SOV) being the canonical word order and Object-Subject-Verb the non-canonical, as shown in (1). To distinguish between the subject and the object, the accusative marker *-o* was affixed to the noun that was the object of the action. Adjectives were optional, appeared post-nominally and were also case marked to agree with the nouns they modified. Thus, Kepidalo sentences could be between three and five words in length. All artificial language stimuli were generated using a speech synthesiser (Google Cloud Text-to-Speech) and were presented to participants auditorily.

(1)	a.	Noun_{NOM} – $(\text{Adj}_{\text{NOM}})$ – Noun_{ACC} – $(\text{Adj}_{\text{ACC}})$ – Verb
	b.	Noun_{ACC} – $(\text{Adj}_{\text{ACC}})$ – Noun_{NOM} – $(\text{Adj}_{\text{NOM}})$ – Verb

Each sentence referred to an animated scene displaying two aliens performing an action. There were eight aliens (corresponding to each noun), which could be either dark red or light green in colour (corresponding to each adjective) and could appear carrying out one of the following four actions: approaching, catapulting, chasing, or jumping over (corresponding to each verb). The word-referent pairings were the same across all participants. All animated scenes were initially generated in PowerPoint and were subsequently converted to GIF files. For a detailed description of the construction of the auditory and visual stimuli employed in the study, readers are referred to Kenanidis et al. (2023).

General procedure and design

The study consisted of five experimental sessions that were scheduled between 1 and 2 days apart. The sessions varied in terms of duration, with the first one being the longest (mean = 83.5 min) and the last one the shortest (mean = 35.6 min; Sessions 2, 3 and 4 lasted approximately 50 min each). All tests were created and administered online using the Gorilla experiment builder (Anwyl-Irvine et al., 2020).

An overview of the study design is provided in Table 1. Before beginning the study, participants completed a brief questionnaire that included questions about their demographics and linguistic background. Participants were then informed about the aim of the game but were not given any metalinguistic information about the grammar of the artificial language or about the presence of a test phase. Note that while grammar learning was incidental, this was not the case for the learning of the vocabulary. In fact, in the pre-training, participants

Table 1. Summary of study design.

Session 1	Session 2	Session 3	Session 4	Session 5
Pre-training Lexical Training	Lexical Training	Lexical Training	Lexical Training	Grammaticality Judgment Task
Grammatical Comprehension Test	Grammatical Comprehension Test	Grammatical Comprehension Test	Grammatical Comprehension Test	
Implicit Statistical Learning Task	Sustained Attention Task	Explicit Learning Task		

were explicitly informed that their task was to learn the names of the aliens (nouns) and, similarly, before the Lexical Training, they were told that though they "learned enough words", their task is to "learn some more Kepidalo" by paying attention to the sentences and the feedback they received. These instructions were aimed at raising learners' awareness about their learning target and drawing their attention towards the semantic aspects of the language. This manipulation was consistent with previous artificial language learning studies in which the main goal of the study (i.e. syntax learning) was deliberately disguised (e.g. Grey et al., 2014; Maie & DeKeyser, 2020; Rogers et al., 2016).

Finally, grammatical knowledge was assessed through two distinct tasks: a picture selection task and a grammaticality judgment task (GJT). The picture selection task aimed to enhance the comparability of the grammatical data with that obtained for vocabulary, as an identical task was employed in both instances (including feedback in the vocabulary task). This approach also facilitates direct comparisons with recent artificial language studies that utilised similar tasks (e.g. Rebuschat et al., 2021; Walker et al., 2020). In contrast, the GJT provided a more in-depth examination of participants' understanding of grammatical structures by focusing on specific aspects of grammar (word order versus case marking, and nominative versus accusative case marking). This task was intentionally designed to be more challenging for participants, as each sentence was presented only once and did not provide any visual cues for support (see descriptions below for more information).

Pre-training. Participants were trained on the eight nouns of the artificial language. Noun learning was assessed by means of a 4 alternative-forced-choice (4AFC) task. In each trial, four aliens of the same colour were presented, one on each quadrant of the screen, and a novel word referring to one of these aliens was played. Participants had to select the appropriate referent for the novel word presented and were given explicit visual and audio feedback on their responses. All nouns were presented in the nominative form (e.g. *alg-i*). During the task, each alien served as a target an equal number of times. Participants proceeded to the next task irrespective of their final score. Data from this task were not used in the present study. For a more detailed description of the task see Kenanidis et al. (2023).

Lexical training. Immediately after the pre-training and at the beginning of each of the following 3 sessions, participants were trained on the vocabulary of the artificial language via a two-alternative forced-choice task (2AFC). Each trial involved the auditory presentation of a sentence which was accompanied by two animated scenes (Figure 1A). If participants wanted, they could listen to each sentence a second time, and the videos repeated automatically until the participants responded. Participants were asked to select, as quickly and accurately as possible, the scene that matched the sentence. The two scenes differed in only one aspect, thus creating minimal pairs. That is, the target and distractor scenes were identical expect for one of the aliens (noun test trials), the action performed (verb test trials), or the colour of one of the aliens (adjective test trials). Therefore, knowledge of the lexical item referring to the aspect that was different in the two scenes was necessary for participants to correctly identify the target scene. Auditory and visual feedback was provided after each trial. The location of the target and distractor scenes was counterbalanced.

The task was divided into 3 training blocks of 90 trials each. The canonical SOV word order occurred in 200 sentences and the non-canonical OSV word order in 70 sentences. Noun, verb and adjective test trials appeared equally frequently (90 trials per lexical category) and were randomly intermixed within the 3 blocks. A different pseudorandomised order of presentation of the same 270 sentences was used for each session, but the sequence in which the sentences were presented was the same for all participants.

Grammatical comprehension test. To determine whether and when participants had learned the functional role of case marking in the artificial language, and by extension the presence of two word orders, a separate 2AFC task was administered. The procedure followed was identical to the lexical training task except that no feedback on the correctness of the response was provided. Participants heard a Kepidalo sentence and saw two animated scenes in which the same two

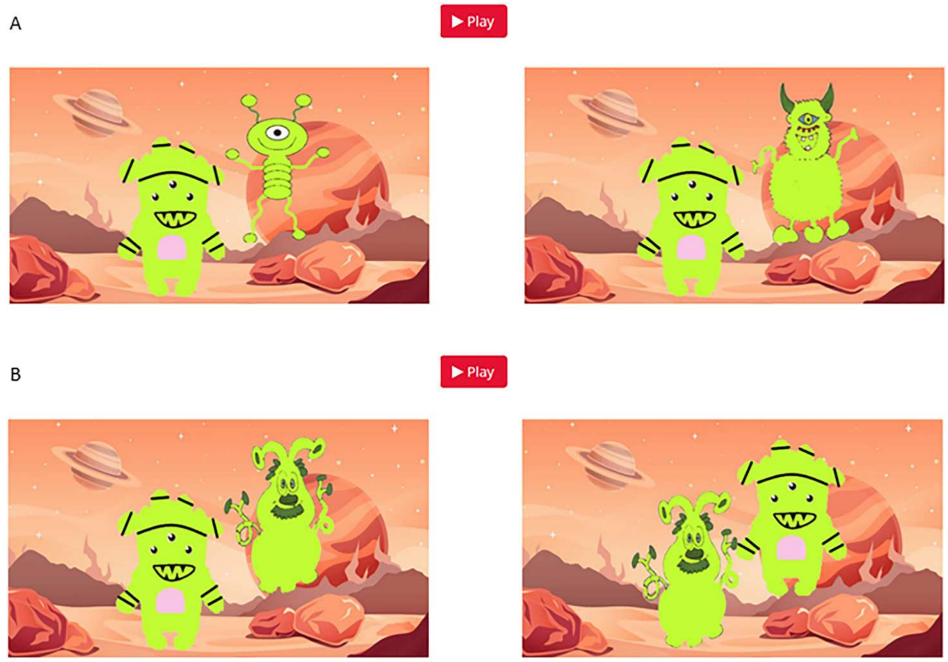

Figure 1. Screenshots of trials from (A) the Lexical Training task (Left scene: the (green) prad is jumping over the (green) velg, Right scene: the (green) torg is chasing the (green) velg) and (B) the Grammatical Comprehension test (Left scene: the (green) ird is jumping over the (green) velg, Right scene: the (green) velg is chasing the (green) ird).

aliens appeared performing the same actions, but with the agent and patient roles reversed (Figure 1B). Correct performance, therefore, required identification of the case markers. Learners were exposed to 90 previously unheard sentences, 70 of which had SOV and 20 had the OSV word order. The test sentences were the same in all sessions but were presented in a different randomised order in each session. The position of the target video on the screen was counterbalanced across trials.

Grammaticality judgment task. In session 5, a Grammaticality Judgment Task (GJT) was administered to participants to further assess their knowledge of the grammar of the novel artificial language. Participants were exposed to 80 previously unheard sentences, half of which were grammatical and half of which were ungrammatical. There were five different types of violations (see Appendix B), for each of which 8 ungrammatical test items were presented. From those five violation types, two involved errors in word order, while the remaining three had case marking errors.

In each trial, a sentence was presented, and participants were asked to decide whether the sentence sounded correct or not. All sentences were played only once, and no feedback was provided. In this task, half of the grammatical sentences appeared with the SOV and half with the OSV word order. The order of sentence presentation was randomised but was the same for all subjects.

Implicit statistical learning. ISL was measured through the VSL task that was originally developed by Siegelman et al. (2017). This task displays improved psychometric properties compared to previous tasks measuring SL and shows high internal reliability and consistency making it suitable for studying individual differences. The task involved two distinct phases. First, participants were familiarised with a continuous stream of 16 complex shapes, which unbeknownst to them, consisted of 8 triplets (i.e. sequences of three shapes). Importantly, the triplets differed in that half of them were defined by high transitional probabilities (TPs = 1) within the shapes (triplets: 5-6-7, 8-9-10, 11-12-13 and 14-15-16) – meaning that, within the triplets, each shape always occurred after the same shape (e.g. 5-6, 6-7) – whereas the other half consisted of the same four shapes and had lower TPs (= .33; triplets: 1-2-3, 2-1-4, 3-4-2 and 4-3-1) – meaning that each shape could follow 3 possible shapes (e.g. shape 2 could occur after 1 in the triplet 1-2-3, after 4 in the triplet 3-4-2 or after a shape from a different triplet in the triplet 2-1-4). The exposure phase consisted of 24 presentations of each of the 8 triplets. Each shape appeared on the screen for 1000 ms. To ensure that participants were actively engaged in the task, a small modification to the original task was made. The serial presentation of the continuous sequence of stimuli was split into 12 blocks, within which each triplet appeared twice in a pseudo-random order, with the constraint that the same triplet never appeared twice consecutively. At the end of each block, participants were presented with an attention check asking them to press the space bar as quickly as possible. The intention was to exclude participants who failed to respond within 10 s to two attention check items. However, all subjects responded within the time window and were thus included in the final analysis.

Prior to the experiment, participants were informed that they would be presented with some shapes that would appear in the centre of the screen, one after the other, and their task was simply to attend to these shapes, as they would later be asked some questions about them. No mention was made of the fact that the shapes would appear in any particular order.

Immediately after the end of the familiarisation phase, we tested whether participants have noticed any patterns in the order in which the shapes appeared by asking them to indicate if the continuous stream of items they saw consisted of no fixed sequences or of sequences of 2, 3, 4 or 6 shapes (5AFC). Following that, participants' knowledge of the triplets was tested using forced-choice trials. There were 42 test items which were divided into 2 blocks. The first block consisted of 34 pattern recognition trials, including 22 2AFC and 12 4AFC trials, in which participants had to discriminate between familiar and foil sequences of triplets ($N = 24$) and pairs of shapes within the triplets ($N = 10$). The second test block contained 8 pattern completion trials. Participants were presented with a target pair or triplet, but with one of the shapes replaced with a question mark and had to choose which shape best completed the pattern from a possible selection of two or three. Four triplets and four pairs served as targets in this block.

Trials within each block were randomised. Following Siegelman et al. (2017), chance-level performance in this task was set at 39.7% accuracy (or 16.7/42) and individual chance level, above which participants showed learning, was defined as correct responses on 23 trials or more. One point was awarded for each correctly identified target, with a maximum of 42 points.

Sustained attention. Sustained attention was assessed using a modified version of the Continuous Performance Task-X (CPT-X) originally developed by Rosvold et al. (1956). In a target detection task, black-colored block letters were presented, one after the other, in the centre of the screen against a white background and participants were instructed to respond by pressing the space bar every time they saw the letter "X" (Go trials) and to withhold responses to non-target letters (No-Go trials). A low target frequency task (low Go, high No-Go) was used in this study since it is considered to be a more sensitive measure of sustained attention than tasks involving high frequency of Go stimuli or tasks that require transient attention (Ballard, 2001; Carter et al., 2013).

Visual stimuli consisted of six letters, the target letter "X" and the distractors "A", "E", "D", "W", "Y". The non-target letters were subdivided into letters of low visual similarity to the target "X" (i.e. "A", "E", "D") and letters that bear high-similarity to the target letter (i.e. "W", "Y"). Each stimulus was presented for 250 ms, followed by an inter-stimulus interval (the time between the offset of a stimulus until the onset of the next one) of 900 ms, during which a fixation dot (.) appeared on the screen to reduce any afterimage effects (MacLean et al., 2009). Stimuli were arranged in blocks of 6, with each letter appearing once in each block in a pseudo-randomised order that varied from block to block. Each block, therefore, contained 1 target and 5 nontarget letters. Targets were presented with a probability of 16.67%.

The task started with a short practice phase, consisting of 3 blocks with 3 target letters. Participants were instructed to pay equal attention to accuracy and speed of response. After completion of the practice phase, feedback about the number of correct identifications of the target stimuli was given on the screen. The practice phase was followed by the test phase during which 360 letters were presented with the target "X" appearing 60 times. At the end of the task, participants once again received feedback about their performance. The test phase took approximately 7 min to complete. D-prime scores were calculated as an index of participants' ability to detect and respond to the target stimuli (see Data analysis section).

Explicit learning. To measure EL, we developed a task which was very similar to the ISL task described before but was different in the instructions given to the participants and the manner in which the shapes were presented. In this task, participants were exposed to 24 new complex shapes (created at mathigon.org/tangram) that were organised into 8 triplets (Figure 2). In contrast to the original task, the end of each sequence of three shapes was marked by the presentation of a fixation cross, which appeared in the centre of the screen for 250 ms, and subjects were informed about the presence of sequences in the input and were explicitly instructed to pay attention and memorise them, as a test will follow. Each triplet was presented 7 times, divided over 4 blocks; 3 blocks of 2 repetitions of each triplet and 1 block in which each triplet appeared once. The number of repetitions was chosen after piloting the task with 6 repetitions of each triplet, which proved to be insufficient for significant learning to occur, and 8 repetitions, which led to ceiling effects. To

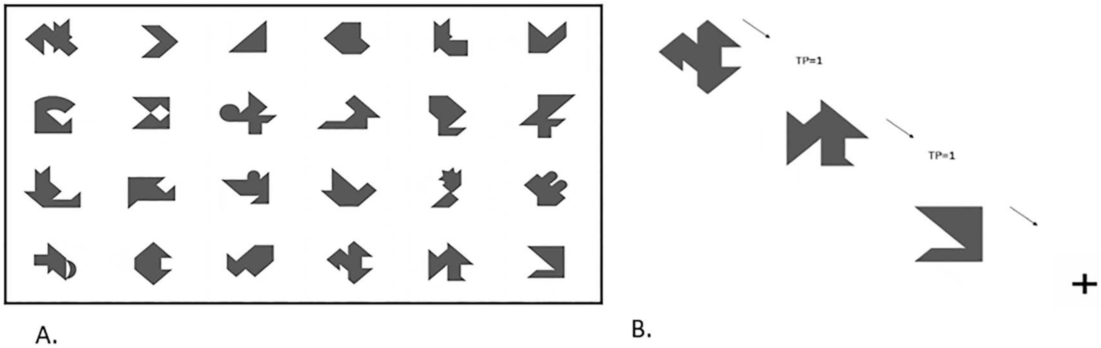

Figure 2. (A) The 24 shapes used in the explicit learning task. (B) Schematic representation of the Transitional Probability structure in the explicit learning task.

ensure participants remained attentive to the task, there was an attention check between blocks that was identical to the one in the ISL task.

To familiarise participants with the idea of triplets, the task began with a practice trial in which participants first saw three simple shapes (square, circle, and triangle) presented in a sequence, and were then presented with the three shapes displayed simultaneously in the same linear order and a foil consisting of three other shapes and asked to select the sequence they were shown earlier. The practice trial was followed by the familiarisation phase. At the end of the familiarisation phase, a forced choice task was used for assessing participants' learning of triplets. The task had the same format as the VSL task described earlier. Specifically, the test phase contained two parts. The pattern recognition part which was composed of 34 trials, during which participants were asked to choose which patterns of visual stimuli they were more familiar with, from either two ($N = 22$) or four ($N = 12$) alternatives. In the second part of the test phase, 8 pattern completion trials were presented, during which participants were required to complete the missing shape in a pattern. The correct answer was either one out of two (2AFC) or one out of three (3AFC) shapes. Once again, knowledge of triplets as well as pairs within the triplets was tested. Within each test block, the test items were presented in a pseudo-randomised order with the restriction that no two consecutive trials tested the same triplet. Trials were administered in the same order for all participants. Participants were awarded 1 point for every correct response; thus, possible scores ranged from 0 to 42.

The explicit learning task always followed the VSL task, as completing the explicit task first was likely to bias participants towards directing their attention to sequences of shapes within triplets and hence invalidate it as a measure of implicit learning.

Data analysis

Accuracy scores on the Lexical Training and the Grammatical Comprehension trials were calculated as the number of correct responses per session. For the GJT task, following signal detection theory (Macmillan & Creelman, 2004), participants' ability to discriminate between correct and incorrect sentences was measured using *d*-prime (d'). Specifically, four scores were obtained for each participant: hits (grammatical sentences judged as acceptable), misses (grammatical sentences judged as unacceptable), false alarms (ungrammatical sentences judged as acceptable) and correct rejections (ungrammatical sentences judged as unacceptable). From these scores, d' scores were calculated for each participant (Huang & Ferreira, 2020) using the "psycho" package (version 0.6.1; Makowski, 2018) in RStudio (version 4.3.1; R Core Team, 2022). Higher d' scores indicate greater discrimination sensitivity. The same procedure was used for the responses obtained from the Sustained Attention task. Here, higher *d*-prime scores reflected better ability to maintain sensitivity to targets (Corkum et al., 1995). For the ISL and the Explicit Learning scores were calculated as the sum of correct responses.

Data from one participant in the EL were missing due to the participant not properly exiting the web page. Additionally, results for three participants in the Sustained Attention task were excluded; either because of task malfunction (N = 1) or because they failed to understand the task instructions, as evidenced by their performance throughout the task[4] ($N = 2$). Missing data for these participants were imputed using a predictive mean matching method from the mice package (version 3.13.0; van Buuren & Groothuis-Oudshoorn, 2011) in R.

Statistical analyses. In order to determine which individual-differences measures predicted accuracy in the Lexical Training trials, the Grammar Comprehension

Test trials and the GJT, mixed-effects models were employed. Data from each task were submitted to separate generalised linear mixed-effects regression models with a logit linking function (Jaeger, 2008) with Accuracy coded as the binary outcome variable (correct = 1, incorrect = 0). The models were implemented using the glmer function of the lme4 package (version 1.1.25; Bates et al., 2015) in R. In the Lexical Training and Grammatical Comprehension Test models, the main effects included the variables Session, which was centred on zero (Session 1 = −1, Session 2 = −0.5, Session 3 = 0.5 and Session 4 = 1) and treated as a continuous variable, and Word Order (treatment coded with SOV as baseline). The numerical independent variables that were added to the models as fixed effects were ISL (VSL), EL (Explicit Learning), and Sustained Attention, all of which were scaled and zero centred. Additionally, the two-way interactions of interest that were included in the models were: Session * ISL, Session * EL, Session * Word Order, ISL * Attention, EL * Attention, and ISL * EL.

In the mixed-effects model for the GJT, the fixed effects and interactions that were tested were the same as in the two previous models, with two exceptions: (1) the Session variable and its interactions with other predictors were not included in the model, given that the GJT was administered only once during the study, and (2) the model included sentence Grammaticality and Error Type as predictors, both of which were effect-coded such that grammatical sentences and word order violations corresponded to 0.5, while ungrammatical sentences and case marking violations were coded as −0.5.

In all models, a two-step backward elimination approach, similar to that outlined in Zuur et al. (2009), was employed (see also Gries, 2021). Models were initially fit with all main predictors and two-way interactions of interest. Following Barr et al. (2013), the random-effect structure included varying intercepts for participants and test items and slopes for all predictors of interest (albeit knowing that this is likely to lead to model convergence issues). The model for the Lexical Training trials had random intercepts for test items nested within distractor type (verb, noun, adjective). The first step involved determining the best random-effects structure. The structure was simplified by iteratively removing the random slopes with the smallest SD until the model converged without singularity warnings. We then proceeded with finding the best fixed-effects structure. This was done by progressively dropping the interaction terms identified by drop1 as contributing less to the model fit. The drop1 function looks at all the predictors and determines which one is the most droppable by comparing the reduced models against the previous full model by means of a likelihood ratio test. The interactions were removed if this improved the fit of the model as indicated by the Akaike information criterion (AIC; Akaike, 1974) values. This procedure was repeated iteratively until no further improvement in model fit was observed. All significant interactions observed were further unpacked via post-hoc analyses using the emmeans package (version 1.8.1-1; Lenth, 2021). Finally, Spearman-Brown corrected reliability estimates for all artificial language (see Appendix C) and individual-differences measures were computed using the splithalf (version 0.8.2; Parsons, 2021) and splithalfr (for d’ scores) packages (version 2.2.0; Pronk, 2021) in R. Finally, R^2 for all models were computed using the MuMiN package (version1.47.5; Bartoń, 2022).

Data availability. This study was not preregistered. All the experimental stimuli and scripts that were created and used as part of this study are available on Gorilla Open Materials (https://app.gorilla.sc/openmaterials/486012). The data sets and scripts required to reproduce the analyses reported in this paper are available on the Open Science Framework at: https://osf.io/dvw5s/?view_only=59bb6e899bd94dbc9daf401a508757f2.

Results

Participants’ performance on all individual-differences measures is summarised in Table 2. Correlation matrices showing the relationship between the artificial language learning and individual differences measures are provided in Appendix D.

To get a first impression of how participants performed in the artificial language tasks, mean accuracy scores and mean reaction times on the Lexical Training and Grammatical Comprehension test were calculated for each session (Table 3). Given that both lexical and grammatical learning were measured by means of 2AFC tasks, chance-level performance was 135 (out of 270) correct responses in the Lexical Training task and 45 (out of 90) correct responses in the Grammatical Comprehension test blocks.

In the Lexical Training trials, participants achieved high levels of accuracy from as early as Session 1 (M = 186 correct trials). Even though performance in the Lexical Training task never reached ceiling levels, Table 3 suggests that participants showed clear improvements from session to session. Regarding word order, participants achieved similar accuracy for SOV (M = 79.2%, SD = 11.9%) and for OSV sentences (M = 77.9%, SD = 12.4%). By contrast, although participants performed above chance on the grammar trials from Session 1 onwards, they only seem to show a marginal

Table 2. Descriptive statistics for the individual-differences measures.

Individual differences measures	M (SD)	Median	IQR	SE	Range	Skew	Kurtosis	Reliability
Implicit Statistical Learning[a]	27.02 (7.06)	27.0	11	1.10	12–39	−0.17	−0.96	.83
Sustained Attention (d')[b]	4.63 (0.38)	4.67	0.30	0.06	3.60–5.34	−0.34	0.46	.88
Explicit learning[c]	31.12 (7.81)	32.0	12	1.22	16–42	−0.27	−1.24	.90

[a]Mean accuracy rea scores were calculated out of a maximum of 42.
[b]D-prime scores (higher scores indicate higher sensitivity to targets).
[c]Mean accuracy rea scores were calculated out of a maximum of 42.

Table 3. Mean accuracy and reaction times across sessions in the Lexical Training and Grammatical Comprehension test blocks.

	Lexical Training			Grammatical Comprehension test		
	M%	Accuracy M (SD)	Reaction Time M (SD)	M%	Accuracy M (SD)	Reaction Time M (SD)
Session 1	68.9	186.1 (32.7)	5217.0 (1782.2)	62.7	56.4 (11.1)	3686.0 (1428.6)
Session 2	79.4	214.3 (34.7)	4158.2 (1316.1)	64.4	58.0 (10.8)	2952.3 (854.4)
Session 3	82.4	222.6 (37.7)	3645.1 (1049.6)	64.4	58.0 (9.8)	2498.9 (794.0)
Session 4	84.9	229.1 (35.1)	3468.6 (1206.5)	67.0	60.3 (12.1)	2164.5 (800.2)
Overall	78.9	213.0 (38.5)	4122.2 (1516.1)	64.6	58.2 (11.0)	2825.4 (1148.4)

improvement throughout the remainder of the study. Furthermore, in the Grammatical Comprehension test, accuracies for SOV ($M = 74.7\%$, $SD = 19.9\%$) and OSV ($M = 29.4\%$, SD = 21.6%) sentences diverge substantially. With regard to the GJT, rather unsurprisingly, performance at the group level was better on grammatical sentences ($M = 90.6\%$, $SD = 12.5\%$) than on ungrammatical ones ($M = 37.4\%$, $SD = 14.4\%$).

We next turn to the main part of our analysis, where we outline the results of the generalised linear mixed-effects regression models for the Lexical Training, Grammatical Comprehension Test and GJT. The coefficients and significance levels of all fixed effects in the best-fitting model for the Lexical Training trials are reported in Table 4. The model showed that performance on SOV sentences improved over time, with a significant main effect of Session. Furthermore, the lack of Word Order effect suggested that there were no differences in performance across the two types of sentences. Regarding the effects of cognitive abilities, while ISL did not emerge as a significant predictor, EL and Sustained Attention did. This indicates that accuracy on the Lexical Training trials was modulated by performance on these measures, with higher scores in these tasks predicting higher accuracy on the Lexical Training trials. Additionally, the model revealed significant interactions between EL and Session and ISL and Session (see Appendix E for the full results). Interestingly, a contrasting pattern was found. The interaction between EL and Session showed that the effect of the EL ability increased as a function of Session, reaching significance from session 2 onward and becoming progressively stronger, as shown by follow-up simple slope analyses, which allowed us to obtain separates slope for each session. In contrast, the interaction between ISL and Session indicated that the effect of ISL was most

Table 4. Mixed-effects model fitted to the Lexical Training data.

Variable	$\hat{\beta}$	*SE*	*Z*	*p*	*OR (CI)*
(Intercept) = SOV	1.91	0.38	5.06	<.001	6.73 (3.21–14.08)
Session	0.78	0.14	5.76	<.001	2.18 (1.67–2.84)
Explicit Learning	0.50	0.16	3.02	.003	1.64 (1.19–2.27)
Implicit Statistical Learning	0.08	0.17	0.48	.629	1.08 (0.78–1.50)
Sustained Attention	0.29	0.09	3.33	<.001	1.33 (1.12–1.58)
Word Order = OSV	−0.09	0.07	−1.32	.187	0.92 (0.81–1.04)
Session: Explicit Learning	0.33	0.12	2.82	.005	1.39 (1.11–1.74)
Session: Implicit Statistical Learning	−0.26	0.12	−2.22	.027	0.77 (0.61–0.97)
Implicit Statistical Learning: Sustained Attention	0.31	0.10	3.06	.002	1.37 (1.12–1.67)
Random effects	Variance	SD			
Participant	0.92	0.96			
Item: Distractor Type	0.17	0.42			
Distractor Type	0.36	0.60			
Session | Participant	0.45	0.67			
Session | Item: Distractor Type	0.02	0.13			
Session | Distractor Type	0.02	0.14			
Marginal R^2	.14				
Conditional R^2	.44				

pronounced (and significant only) in the first session and became attenuated over time (Figure 3). Finally, there was a significant interaction between ISL and Sustained Attention, for which results from a follow-up simple slope analysis found that the effect of Sustained Attention became stronger as participants' scores in the ISL task increased but was nonsignificant for those who had low scores. The interaction between EL and Sustained Attention was not retained in the best-fitting model, in line with the model selection procedure described above.

Table 5 presents the results of the model for the Grammatical Comprehension test data. Overall, participants performed significantly above chance on SOV sentences, as indicated by the significant model intercept. Furthermore, the main effect of Session showed that their performance on this type of sentences improved significantly over time. On the other

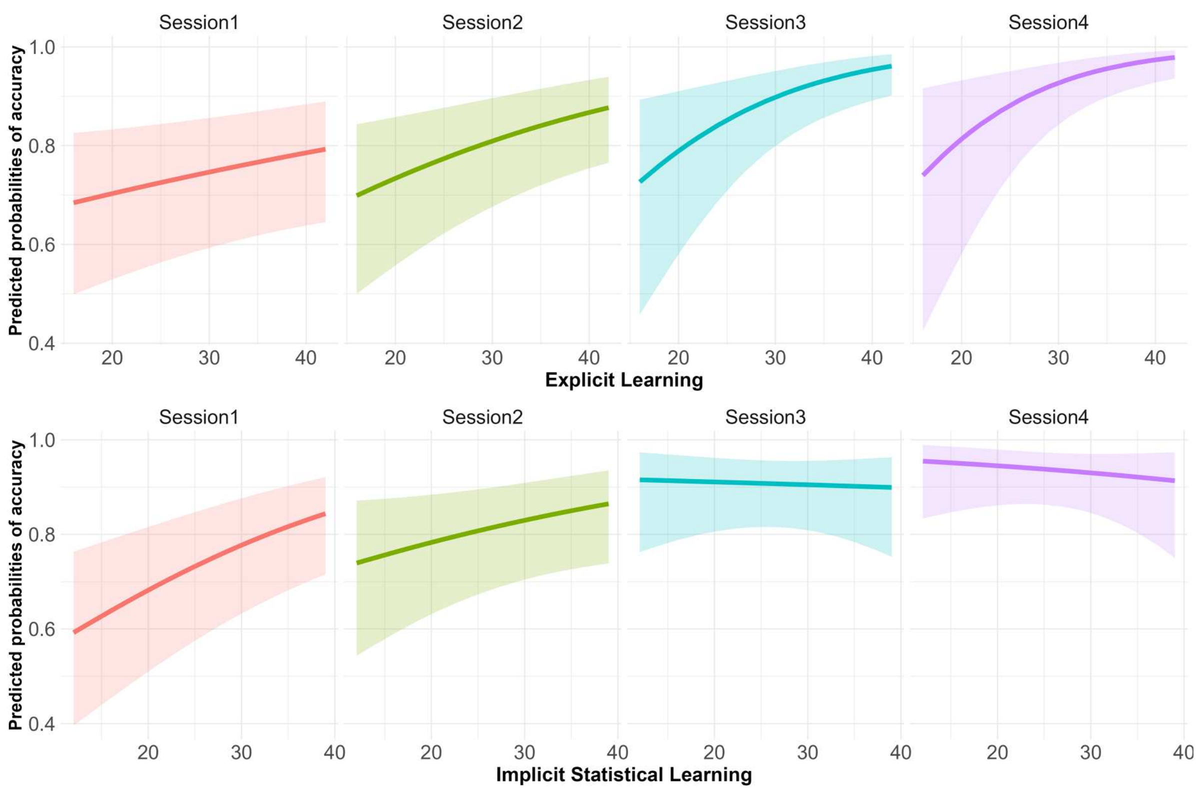

Figure 3. Predicted probabilities of an accurate response across sessions in the Lexical Training task, as a function of EL and ISL abilities. The plot was generated using the ggeffect function from the ggeffects package (Lüdecke, 2018) in R.

Table 5. Mixed-effects model fitted to the Grammatical Comprehension test data.

Variable	$\hat{\beta}$	*SE*	*z*	*p*	*OR (CI)*
(Intercept) = SOV	1.47	0.20	7.27	<.001	4.37 (2.94–6.50)
Session	0.21	0.08	2.61	.009	1.23 (1.05–1.44)
Explicit Learning	0.04	0.04	0.90	.369	1.04 (0.96–1.13)
Implicit Statistical Learning	−0.04	0.04	−0.91	.361	0.96 (0.88–1.05)
Sustained Attention	0.04	0.04	0.98	.328	1.04 (0.96–1.12)
Word Order = OSV	−2.68	0.40	−6.73	<.001	0.07 (0.03–0.15)
Session:Word Order	−0.26	0.09	−3.02	.003	0.77 (0.65–0.91)
Session: Explicit Learning	0.24	0.08	3.06	.002	1.27 (1.09–1.48)
Session: Implicit Statistical Learning	−0.10	0.08	−1.30	.193	0.90 (0.77–1.05)
Random effects	Variance	SD			
Participant	1.35	1.16			
Item	0.51	0.72			
Word Order | Participant	4.99	2.23			
Session | Participant	0.18	0.42			
Session | Item	0.05	0.23			
Marginal R^2	.20				
Conditional R^2	.50				

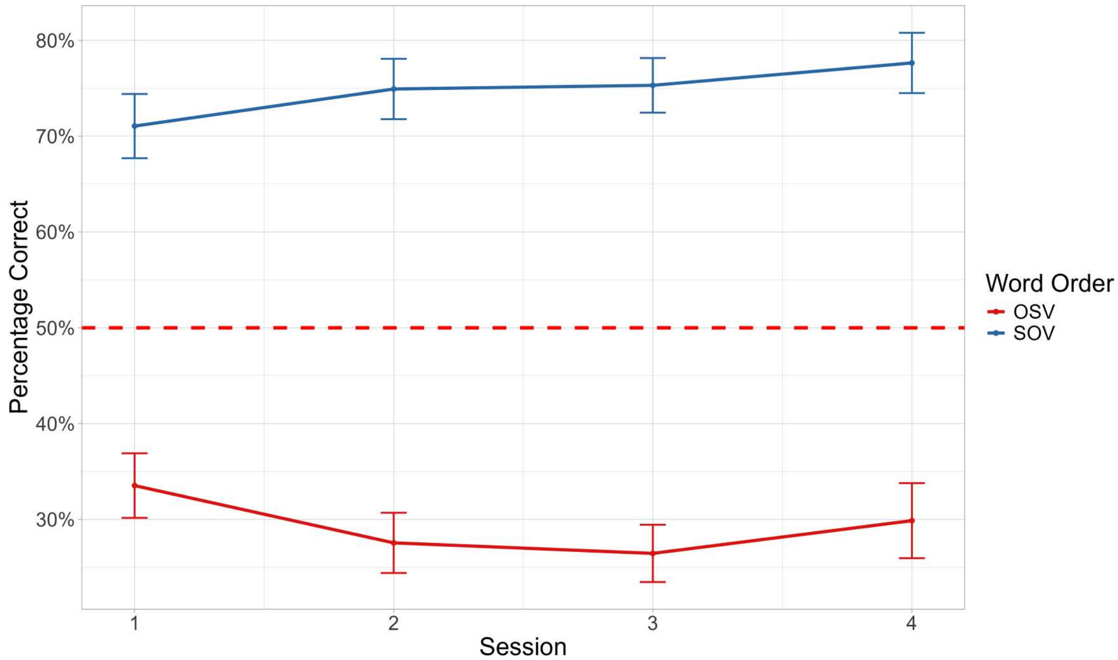

Figure 4. Performance on the two word orders across the four sessions in the Grammatical comprehension test. The red dashed line represents chance performance and error bars, calculated over by-subject means, represent standard errors of the mean.

hand, there was an effect of Word Order and an interaction between Session and Word Order, suggesting that participants were significantly less accurate on OSV than on SOV sentences, and that the effect of Word Order was significant only for the latter type of sentences. Post-hoc analyses revealed that this difference in accuracy remained significant throughout the study (Figure 4).

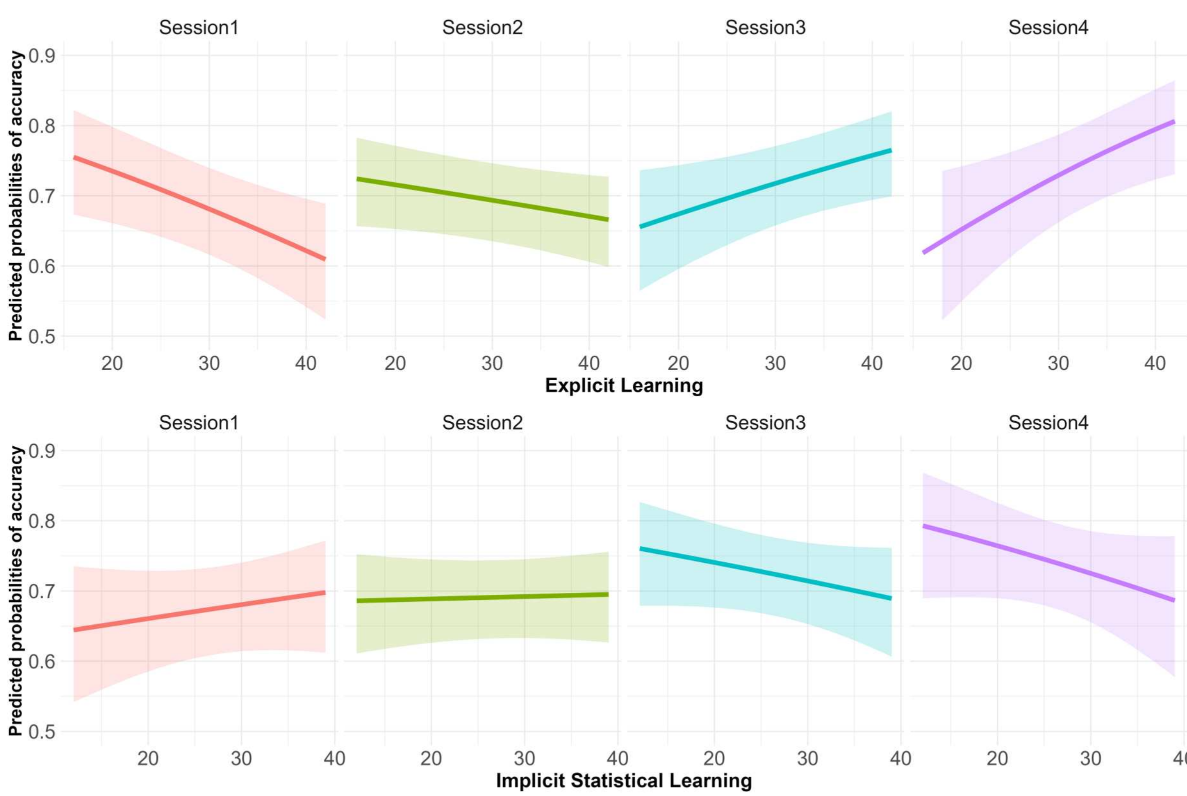

Figure 5. Predicted probabilities of an accurate response across sessions in the Grammatical Comprehension test, as a function of EL and ISL abilities. The plot was generated using the ggeffect function from the ggeffects package (Lüdecke, 2018) in R.

Table 6. Mixed-effects model fitted to the Grammaticality Judgment task data.

Variable	$\hat{\beta}$	*SE*	*Z*	*p*	*OR (CI)*
(Intercept)	1.87	0.29	6.40	<.001	6.51 (3.67–11.55)
Grammaticality	4.66	0.71	6.55	<.001	105.90 (26.24–427.43)
Word Order	0.23	0.50	0.46	.647	1.26 (0.47–3.37)
Explicit Learning	0.31	0.13	2.44	.015	1.36 (1.06–1.73)
Implicit Statistical Learning	−0.01	0.12	−0.01	.992	1.00 (0.79–1.26)
Sustained Attention	0.30	0.12	2.56	.010	1.35 (1.07–1.69)
Error	2.47	0.52	4.73	<.001	11.84 (4.25–33.00)
Explicit Learning: Sustained Attention	0.37	0.12	2.97	.003	1.45 (1.13–1.84)
Random effects	Variance	SD			
Participant	0.30	0.55			
Item	4.44	2.11			
Grammaticality	8.69	2.95			
Marginal R^2	.41				
Conditional R^2	.81				

The model of Grammatical Comprehension test also revealed a significant positive interaction between EL and Session, similarly to the Lexical Training model. Follow-up simple slope analyses (see Appendix E) indicated that the effect of EL was negative in the first two sessions, but positive in the latter two; that is, participants with higher scores in the EL task were more likely to show higher accuracy only as the study progressed (Figure 5). In contrast to the previous model, the interaction between ISL and Session was not significant in this case (Figure 5). However, it is worth noting that the direction of the interaction was the same as for the Lexical Training model, pointing towards a tendency for the effect of implicit learning ability to decrease as a function of Session.

Finally, the results of the model of the GJT administered only in the final session (Table 6) revealed a significant positive intercept suggesting that participants performed significantly above chance in this task. The model returned significant main effects of Grammaticality and Error Type, with participants showing significantly higher accuracy on the grammatical than on the ungrammatical sentences, and on word order than on case marking violations. In addition, significant positive

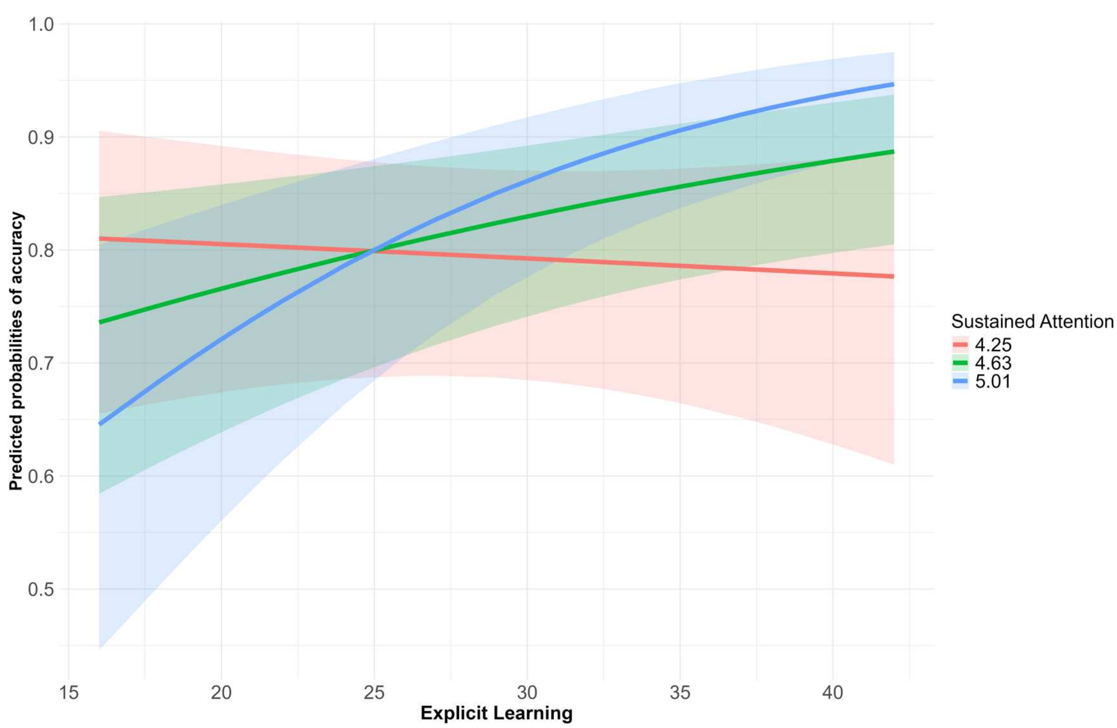

Figure 6. Predicted probabilities of an accurate response across sessions in the Grammaticality Judgment task, as a function of Explicit Learning and Sustained Attention abilities. The plot was generated using the ggpredict function from the ggeffects package (Lüdecke, 2018) in R.

effects of EL and Sustained Attention were found, indicating that participants with higher scores in the EL task and in the Sustained Attention task performed better in the GJT. There was also a significant positive interaction between EL and Sustained Attention, for which follow-up simple slope analyses revealed that the effect of EL on GJT accuracy became stronger as Sustained Attention scores increased, but was not significant and even negative for participants with low scores in this task (Figure 6).

Given the low accuracy rates for ungrammatical sentences in the GJT, and following a suggestion from an anonymous reviewer, an additional linear regression model on *d* prime scores was conducted to test whether participants performed above chance in distinguishing between grammatical and ungrammatical sentences. The results indicated that performance was significantly better than would be expected by chance ($\widehat{\beta} = 1.24$, $t = 14.1$, $p < .001$). The model also showed main effects of EL ($\widehat{\beta} = 0.23$, $t = 2.19$, $p = .028$) and sustained attention ($\widehat{\beta} = 0.21$, $t = 2.13$, $p = .04$), such that, once again, higher scores in these abilities were associated with higher sensitivity scores.

Discussion

In the present study, we examined the contribution of individual differences in three cognitive abilities that have been linked to adult L2 acquisition, namely aptitudes for ISL and EL, and sustained attention, to the initial stages of novel vocabulary and grammar learning, under intentional and incidental exposure conditions, respectively. By investigating how individual differences in ISL and EL aptitudes influence learning outcomes as a function of time, an additional aim of the study was to elucidate the time-course of actuation of these abilities in the earliest stages of the L2 learning process, thereby allowing us to test the earlier claim that they play distinct roles at different learning stages, with explicit aptitude being drawn upon more in initial learning and implicit aptitude more in advanced learning (e.g. Li & DeKeyser, 2021). To this end, using a micro-longitudinal design comprising five sessions within a span of 5–10 days, adult native speakers of English were exposed to Kepidalo, an artificial language which involved case marking and flexible word order, aspects of grammar that are not present in learners' L1.

Overall, we found that, while participants quickly learned the novel word-referent mappings and their accuracy rates on the Lexical Training task increased significantly throughout the study, they had only partial success in discovering and learning the grammatical structure underlying the artificial language. These learning difficulties appeared to persist despite extensive exposure to input but did not apply to the same extent to all aspects of grammar. Specifically, by the end of the study, learners possessed substantial knowledge of the verb-final character of the artificial language, and a clear learning effect was also observed for the canonical SOV order. In contrast, performance on sentences with the non-canonical OSV order remained below chance throughout the study. This difficulty can be directly related to the limited success in learning case marking, as clearly highlighted by the results of the GJT, causing learners to apply the canonical SOV pattern. Nevertheless, despite the lack of significant learning gains on the noncanonical OSV order, the presence of this structure in input is likely to have interfered with participants' accuracy on the canonical order, preventing it from reaching ceiling (for a more detailed discussion of this point, see Kenanidis et al., 2023). In sum, our findings add to the body of literature showing that learning case marking under incidental exposure conditions can be a challenging task for adult learners, particularly for those whose L1 is morphologically poor (Grey et al., 2014; Kenanidis et al., 2023; Rogers et al., 2016). With respect to the role of cognitive abilities examined here and their influence on the development of vocabulary and grammar learning, the results revealed several notable relationships, including significant interactions of EL and ISL aptitudes with Session. We elaborate on these findings below.

Cognitive predictors of individual differences in adult L2 learning

The first research question explored was which cognitive abilities support early L2 learning. With respect to vocabulary learning, the results showed significant independent contributions of EL and sustained attention, with the former becoming progressively more pronounced with time. In addition, a time-varying effect of ISL was also found, with its impact being positive in the very initial stages of exposure, during sessions 1 and 2, but negative in the final two sessions (see Figure 3).

The association between EL aptitude and participants' performance on the Lexical Training task does not come as a surprise, given that vocabulary learning took place under more intentional exposure conditions, which are thought to encourage the involvement of conscious learning strategies. This finding appears to be in line with theoretical accounts that posit a strong role for EL processes in vocabulary learning (Ellis, 2005; Ullman, 2016, 2020). ISL was also predictive of Lexical Training performance, but only in the earliest stages of exposure. This finding corroborates previous research showing an association between ISL and L2 outcomes (Brooks & Kempe, 2013; Chen et al., 2022; McDonough & Trofimovich, 2016) and extends the evidence for the role of this ability in novel vocabulary learning (e.g. Speciale et al., 2004).

Our results also revealed a positive link between vocabulary learning and sustained attention. This finding is consistent with previous research suggesting that the ability to maintain attention on the learning content for an extended period of time can lead to higher language learning gains (Godfroid et al., 2013; Indrarathne & Kormos, 2017). Failure to sustain attention to a task can be caused either by increased task complexity (Head & Helton, 2012) or monotony (Langner & Eickhoff, 2013). In the case of the Lexical Training task, both parameters might have contributed to the overall sustained attention effect, with the former being more important initially and the latter becoming stronger later on, as the task became more repetitive. However, task complexity appears to be the primary driver of the observed effect. Firstly, sustained attention correlated more strongly with vocabulary learning in earlier stages of training ($r = .23$ in Session 1, $r = .14$ in Session 4; for detailed correlation matrices see Appendix D). Secondly, its effects were modulated by ISL, such that they were stronger for learners who had higher scores in the test assessing ISL, which, as mentioned above, also played a stronger role during initial vocabulary learning. This pattern was further supported by adding the interaction between sustained attention and session to the final model of the Lexical Training task, which showed that the effect of sustained attention was stronger in the first session and decreased over time. However, the inclusion of this interaction did not improve the model's fit and was hence removed from the reported model (see OSF script for details).

In contrast to vocabulary, despite receiving extensive exposure to stimuli, participants' performance in the Grammatical Comprehension test blocks showed only small improvement after the first session (Table 3). With respect to the contributions of cognitive abilities, EL emerged as a predictor of learning outcomes, but only when moderated by session. This suggests that, with time, learners who are better at explicitly forming and testing various hypotheses became more likely to identify the grammatical structure of the language, achieving better learning outcomes. Conversely, higher ISL scores were more likely to lead to higher accuracy rates at early stages of learning; however, in spite of a tendency in that direction, the interaction between ISL and session was not significant.

Finally, our results showed that EL ability also played an important role in GJT performance. This effect of EL was found to be more pronounced for learners with stronger sustained attention ability. Note that our GJT was untimed, i.e. participants could take as much time as they needed. Untimed GJTs are thought to prompt learners to turn their attention to the form of the target sentence leading them to notice the presence of potential errors and consciously access their (explicit) grammatical knowledge (R. Ellis, 2005; Godfroid et al., 2015) to detect mismatches between their knowledge and the input (Ellis, 1994). Hence, learners with higher EL scores were better able to keep the input active while reflecting on their knowledge in order to decide whether the sentence was grammatical or not. This process, however, tended to be more effective for those participants who were most likely to stay motivated and engaged with the task, as indicated by their ability to sustain attention over time.

Taken together, our results suggest that ISL and EL aptitudes contribute differently for vocabulary and grammar. This was found to be the case under different exposure conditions (intentional for vocabulary and incidental for grammar), which is consistent with the idea that both processes can be engaged irrespective of the learning context (DeKeyser, 2015; Ellis, 2005; Leung & Williams, 2014). Importantly, as already briefly discussed above, the relative influence of ISL and EL varied at different time points of learning but followed the same temporal pattern across linguistic domains, an issue we turn to in the next section.

Differential time-dependent effects of implicit statistical learning and explicit learning on adult L2 learning

Research questions 2 and 3 asked how aptitudes for ISL and EL affect (intentional) vocabulary and (incidental) grammar learning and how potential effects might change at different stages of the learning process. Our findings revealed a somewhat surprising pattern of interactions between session and ISL and EL abilities; the effect of ISL on lexical development was most pronounced early on, but eventually decreased and even became negative, whilst its effect on grammatical comprehension, though never reaching significance, showed a similar trajectory. Meanwhile, the effect of EL on both vocabulary and grammar increased as learners received more exposure to the artificial language (see Figures 3 and 5).

These findings indicate that performance at the very earliest stages of exposure appears to have been guided primarily by ISL processes, with EL gradually taking over. This pattern aligns well with the more inductive nature of the ISL task in contrast to the more conscious, strategic one of the EL task – which, however, also remains largely inductive, as it involves consciously learning sequences from the input without prior instruction about the specific shape combinations. Specifically, the ISL task taps into participants' ability to infer patterns from a dynamically presented complex set of stimuli. Such an ability is particularly relevant at the early stages of learning, when participants are exposed to novel linguistic input and are still engaged in extracting information about various aspects of the language. While learners may direct attention to certain features of the input, ISL primarily facilitates the initial (unconscious) processing of statistical regularities. By the later sessions, however, once participants have extracted the necessary linguistic structures from the input, the role of ISL diminishes, allowing EL to take over. This shift suggests that learners no longer benefit from further exploration of the input's structure but instead engage in explicit hypothesis testing to refine their knowledge.

Drawing more direct associations between the observed ISL and EL effects and particular linguistic aspects, as mentioned in the introduction, the early effect of ISL on lexical learning may relate to its contribution in speech segmentation (Armstrong et al., 2017; Batterink & Paller, 2017), an aspect of language that is assumed to be learned entirely implicitly (Ullman, 2016, 2020; Williams & Rebuschat, 2023). Better ISL ability may have enabled learners to quickly parse the artificial language speech stream by exploiting cues such as stress (note that all nonwords had a trochaic stress; e.g. 'al-gi, 'mu-lek) and distributional information more effectively, allowing them to quickly discover the novel lexical items. However, segmentation can be achieved at a remarkably fast pace (Cunillera et al., 2009), thus, making it difficult to attribute the effect of ISL solely on speech segmentation and word extraction. In fact, ISL has also been assumed to support the storage (Batterink & Paller, 2017; Thiessen et al., 2013) and retrieval/recognition of the extracted chunks (e.g. words) (Batterink & Paller, 2017; Karuza et al., 2014). Therefore, the early effect of ISL may also reflect interindividual differences in the initial stages of word learning.

Indeed, word learning, at least for verbs and adjectives, could have been more implicit at the beginning of the study. In the Lexical Training task, learners were not given any instructions about what to focus on when listening to the sentences. Moreover, the task was rather complex in that participants had to learn multiple word-referent mappings while listening to sentences. It is thus likely that, instead of proposing and testing explicit hypotheses about potential word-referent mappings, learners gradually established such mappings by tracking co-occurrence statistics, as posited by associative models of cross-situational learning (McMurray et al., 2012; Yu & Smith, 2012).

A similar tendency was also detected in the Grammatical Comprehension task, indicating that having greater ISL abilities may initially be advantageous for learning the grammatical regularities of the novel language (Figure 6). However, in this case, the interaction between ISL and session did not reach significance, possibly reflecting the limited learning observed across sessions. There is ample evidence suggesting that learners are highly sensitive to distributional statistics in the linguistic input. Such information can be drawn upon to discern words that occur in similar distributional patterns, facilitating grammatical categorisation and thereby the learning of linguistic structure (Monaghan et al., 2005; Redington et al., 1998). In this case, nouns, which in the present study were pretrained, may operate as anchor points cueing grammatical categorisation. Over time, accurate recognition of the nouns can enable learners to infer that: (i) nouns reliably precede adjectives, (ii) the second noun reliably precedes verbs, which always occur in sentence final position, and consequently (iii) that the sentences have Noun (adjective)-Noun (adjective)-Verb word order. However, while this information may allow for faster processing of the artificial language sentences, it does not necessarily lead to accurate performance in the Grammatical Comprehension test blocks, as in order to respond correctly, learners had to attend to case marking. This would account for the reduced effect of ISL on grammar learning.

In contrast to ISL, we found a more robust relationship between grammatical comprehension and EL, but in interaction with session. As learners' vocabulary knowledge improves and after identifying the regularities underlying the input, conscious and strategic processing becomes more effective. Learners can draw from their early acquired knowledge to guide their subsequent learning either by formulating and testing specific hypotheses or by paying focused attention to certain aspects of the grammar, which is particularly important for learning forms of low perceptual salience, such as case markers (Ellis, 2006, 2016). These processes are more efficient for those with higher EL ability, as they are better at storing and maintaining linguistic information in their memory and at engaging in attentional processing.

Comparison to previous results

Given the scarcity of studies exploring the simultaneous effects of EL and ISL aptitude on language outcomes in a longitudinal fashion, results from studies employing measures of declarative and procedural memory emerge as an obvious candidate for comparison (Hamrick, 2015; Morgan-Short et al., 2014; Pili-Moss et al., 2020; Walker et al., 2020). Although, as mentioned in the introduction, there is not a one-to-one mapping between these terms, their close relationship, along with the similarities in the design of these studies and the present one (i.e. use of artificial languages and multiple sessions) and, to a certain degree, in the tasks employed to assess individual differences in the target cognitive abilities (e.g. procedural memory tasks are often operationalised as measures of ISL aptitude; Granena, 2013; Suzuki et al., 2023), allows for a meaningful comparison between them.

Notably, the gradual shift from initial reliance on ISL to EL aptitude observed here seemingly contradicts the documented tendency for individual differences in declarative memory to predict early learning outcomes and, in some cases, for variation in procedural memory to be associated with L2 learning at later stages (Hamrick, 2015; Morgan-Short et al., 2014; though see Pili-Moss et al., 2020; Pili-Moss, 2022, for evidence of an early effect of procedural memory on adult L2 production and L2 comprehension, respectively). These differences, however, could be ascribed to several factors. Firstly, participants in Morgan-Short et al. (2014) were exposed to the artificial language through different task-types than the ones used here. Specifically, in each of the four training sessions, participants completed 129 training trials, during which they simply watched short videos showing a game token or a move on a computerised board game, while listening to the corresponding phrase or sentence, and 240 practice trials where they were either listened to sentences and made the corresponding game moves or had to describe a move that was visually presented. This type of stimulus presentation might have simplified the learning process by eliminating, at least in a subset of the trials, the ambiguity associated with identifying the referents to which each sentence may refer, which is not the case in the 2AFC tasks. Secondly, and possibly as a corollary to the previous point, results from Morgan-Short et al. (2014) found that participants achieved far better learning outcomes compared to learners of Kepidalo, thus being in position to proceduralise their initial (explicit) grammatical knowledge, which accounts for the greater contribution of procedural memory to language at later stages of exposure. It may thus be because learners in the present study demonstrated only partial knowledge of the grammar that late ISL effects did not surface.

The different outcomes for the present and previous studies might also partly stem from the variation in the measures used to assess language learning outcomes and cognitive abilities. Regarding the former, in Morgan-Short et al. (2014) and in Hamrick (2015), the only two studies that reported a late effect of procedural memory, language learning was assessed through two GJTs, administered at two different time points. On the other hand, Pili-Moss et al. (2020), who reported further analyses of the Morgan-Short et al.'s (2014) data, focused on a more continuous learning measure by looking at participants' performance in the language practice trials and, for L2 comprehension, failed to find the late procedural memory effects reported in the original study (note that, in the follow-up study, this effect emerged only when a different outcome measure was used, the coefficient of variation of reaction times). One explanation the authors proposed to account for these findings was that, in contrast to GJT, during the language practice trials, participants were required to process both visual and auditory information, leading to stronger engagement of declarative memory throughout the study. Such an explanation also fits well with the present results, since the processing demands of our tasks likely engaged declarative memory more strongly, yielding pronounced EL effects.

Similarly, the cognitive tests used to measure learners' ISL and EL abilities in the current study differ from those employed in previous research. Although this might not be of great concern for measures of explicit aptitude (including declarative memory), given that they typically show high reliability (Bokander & Bylund, 2020) and positive intercorrelations (Buffington et al., 2021; Granena, 2019), it can be problematic for ISL measures (including procedural memory), since correlations between them appear to be low or sometimes even negative (Buffington et al., 2021; Godfroid & Kim, 2021). Therefore, caution should be exercised in the generalizability of the reported patterns of results, as it could be the case that these findings may be contingent upon the different linguistic and ISL measures utilised. Further carefully designed longitudinal investigation is clearly required in order to determine the exact nature of the link between ISL and L2 learning process.

Implications and limitations

The present study examined the independent and joint effects of ISL and EL aptitudes on the earliest phases of adult L2 learning, which served to crucially reveal a complex pattern as to how these abilities intertwine during the first stages of L2 development. On the one hand, a main finding to emerge from the present study is the strong influence of EL aptitude on novel language acquisition. This effect was observed consistently, regardless of learning conditions (incidental or intentional) and across different linguistic domains (grammar or vocabulary). The robust role of explicit aptitude aligns with previous research emphasising its importance in adult L2 learning (Abrahamsson & Hyltenstam, 2009; Li, 2016; Skehan, 2002), reinforcing the idea that adult learners rely heavily on explicit cognitive resources to process and internalise novel linguistic information.

On the other hand, the statistical evidence for the ISL effect was comparatively weaker, particularly on grammatical learning. Nevertheless, our argument regarding ISL's time-varying role does not rest solely on findings from a single task in isolation, but is supported by converging evidence from both the Grammatical Comprehension and Lexical Training tasks. Notably, the ISL effect followed a similar trajectory in both tasks, with its influence not only weaking but also changing direction across sessions (from positive in the first two sessions to negative in the last two; see Figures 3 and 5).

At least two additional findings can serve to corroborate this earlier emerging ISL effect. Kenanidis et al. (2023) found that among German native speakers learning the same artificial language employed here (without, however, completing the individual differences measures), the effect of explicit knowledge, assessed via a post-test metalinguistic awareness questionnaire, emerged as significant only after the second session. Although this does not necessarily imply that early learning was not guided by explicit processes, it indicates that such processes may gain prominence over time, allowing more implicit processes to operate in the initial stages, at least in settings similar to the present study. Further evidence comes from a cross-situational learning study by Walker et al. (2020), where learning of novel nouns and adjectives was predicted by a measure of implicit aptitude (procedural memory, SRT) during the first day of testing with explicit aptitude (declarative memory) becoming the key predictor on the second day. These findings collectively reinforce the notion that the ISL effect evolves dynamically over time.

Crucially, the absence of a significant ISL effect in grammatical comprehension does not contradict the broader observation that ISL did not facilitate learning at later sessions, despite theoretical claims suggesting that implicit aptitude should exert a stronger influence in more advanced rather than in initial stages of L2 acquisition (e.g. Li & DeKeyser, 2021). Instead, our results suggest that the role of ISL may not follow a straightforward trajectory of increasing importance over time. These findings suggest that ISL and EL aptitudes influence L2 acquisition in more intricate and variable ways across different stages, rather than aligning neatly with a simple early versus late-stage distinction. Such an interpretation seems to be more consistent with the multicomponential nature of ISL aptitude, which encompasses a range of abilities whose relevance may vary across tasks and stages of learning (Godfroid & Kim, 2021; Granena, 2020).

At this stage, it is also worth noting that our study has some limitations. Firstly, it should be acknowledged that the task employed in the present study to measure individual differences in learners' ISL ability, VSL (Siegelman et al., 2017), is not without its shortcomings. Specifically, earlier research has suggested that one potential issue with the use of judgement-based alternative-forced-choice measures may be that they jeopardise the implicitness of the learning process as participants are asked to employ meta-cognitive strategies, by reflecting on the acquired statistical regularities in order to answer each trial (Isbilen et al., 2020). However, there is at least some offline evidence that learning during the VSL task was indeed implicit. First, as explained in the method section, we asked participants whether they noticed any recurring patterns and if so, what the patterns would be. Nine out of 41 participants responded that the stream consisted of reoccurring triplets, which is right at chance level (5AFC). Furthermore, only 5 of these 9 participants showed signs of learning, with the remaining 4 performing below chance, indicating that there was no correlation between accuracy and explicit knowledge of the statistical regularities. While the reliability of such a snapshot measure is not without flaws (Shanks & St. John, 1994), participants' responses do provide some indication of the implicitness of the learning process. Finally, ISL was found to relate to performance in the artificial language task in a different way compared to EL, suggesting that the two tasks actually tapped into different abilities.

Another limitation lies in the fact that grammar learning was limited to word order because of learners' large difficulties in identifying and learning novel case marking. Although these learning outcomes are in accord with those observed in earlier studies, the overall small learning effect found here limits, to some extent, the scope of our analyses on the effects of ISL and EL on novel grammar learning (see script in the OSF for a series of post-hoc power analyses).

Future longitudinal studies will now be needed to corroborate and extend the findings reported here. These could be done by employing training regimes that involve even longer exposure phases or artificial languages with simpler grammatical structures (and/or smaller vocabulary) to allow participants more opportunities for developing more automatised knowledge. Alternatively, research could also target learners with a morphologically rich L1 background or even multilingual speakers (e.g. Grey et al., 2014; Kenanidis et al., 2023). More systematic investigation is also needed to examine the extent to which the ISL and EL effects reported herein can be replicated using different tasks, in order to account for the multi-componential nature of the two learning processes. While tasks assessing EL aptitude typically correlate positively, the task employed here included a sequential component, which is not present in other EL measures. For example, tasks like LLAMA_B or LLAMA_F (Meara, 2005) are more associative and inferencing measures, respectively, and hence are expected to show correlations of different strength with different aspects of language (LLAMA_B-vocabulary and LLAMA_F-grammar). With regards to assessing ISL, in addition to the VSL task employed here, studies in a controlled laboratory environment could also employ the widely used SRT or ASRT tasks, both of which include an added motor component, to assess whether and how their predictive value may vary across different stages of acquisition. Performance on these tasks has been found to positively predict grammar processing/accuracy, particularly in advanced learners, but not vocabulary learning (Granena, 2013; Suzuki & DeKeyser, 2015; though see Walker et al., 2020 for a positive relationship between SRT scores and vocabulary learning at early stages of exposure). Another possibility would be to employ more novel online ISL measures, such as the statistically-induced chunking recall task which has been shown to have high reliability (Isbilen et al., 2020) and to correlate positively with sensitivity to multiword chunks (Isbilen et al., 2022).

Finally, another potentially fruitful avenue for future research would be to couple the administration of the ISL and EL tasks with neuroimaging measures during the language learning tasks. Examining the effects of ISL and EL on different aspects of language learning while also tracking the dynamic development of neural representations of L2 over time (brain activity patterns) would allow to further disentangle the relationship/overlap between ISL-EL and declarative-procedural memory.

Conclusion

Extending previous work on individual differences in L2 learning, the present study explored the role of a number of cognitive abilities (aptitudes for ISL and EL, and sustained attention) in accounting for variation in adult learners' ability to pick up novel word-referent mappings under intentional exposure conditions and to incidentally detect and process grammatical structures that are not present in their L1 by means of an artificial language learning paradigm. Moreover, by employing a micro-longitudinal design with five separate sessions, the study sought to examine the dynamic nature of the relationships between ISL and EL abilities and the earliest stages of L2 acquisition. Our results showed that individual differences in ISL, EL and sustained attention predicted vocabulary learning, but only EL and, to a lesser extent, sustained attention were found to play a role in determining grammar learning outcomes. Importantly, the study provided evidence that ISL and EL abilities can contribute differently to learning outcomes at different stages of exposure: ISL was found to be more engaged during the initial stages, whereas the involvement of EL ability became increasingly more evident as learning progressed.

Overall, the consistent effect of EL on the learning of different aspects of language seems to complement previous theoretical and empirical work underscoring its facilitative role in the L2 learning process (e.g. Abrahamsson & Hyltenstam, 2009; DeKeyser, 2015; Ellis, 2005). In contrast, the observed time-course of the ISL and EL contributions appears to deviate from that reported in previous L2 studies in the declarative/procedural-model framework (Hamrick, 2015; Morgan-Short et al., 2014). While these discrepancies may be partially explained by differences in the training regimes and the particular measures used in the study to assess both the target cognitive abilities and the language learning outcomes, they may also tap into crucial differences in the engagement of EL and ISL aptitude as a function of the properties of the target language and the challenges it poses to learners. Despite the exploratory nature of the study, when considered alongside previous research, our findings suggest that the relationship between aptitude for EL and ISL and language learning outcomes may be more nuanced than the traditional dichotomy that associates them with the early and later stages of L2 acquisition, respectively.

Notes

1. Drawing from the similarities between the approaches of implicit learning and statistical learning (in terms of research focus, experimental paradigms, and shared pattern of findings) Conway and Christiansen (2006) and Perruchet and Pacton (2006) have proposed the use of the term "implicit statistical learning". Here, we use the terms *implicit statistical learning* (ISL) and *explicit learning* (EL) to refer to implicit statistical and explicit learning aptitude, respectively. Furthermore, the terms ISL and EL aptitude and ISL and EL ability are used interchangeably.

2. Participants' submissions were rejected if their mean reaction times in 2 of the 3 lexical training blocks of each session were shorter than the average length of the sentences (1967 ms).
3. The grammar of the artificial language used in this study, while not directly modelled after any specific natural language, exhibits similarities to the grammars of several natural languages, most notably Japanese, Korean, and Warlpiri.
4. In the sustained attention task, one participant responded with a button press to all, but one visual stimulus and the second participant had an excessively high rate of incorrect responses (0 hits in the practice trials and 25 misses, 84 false alarms during the test part). Notably, sensitivity analyses that were conducted by running models without the imputed data (Appendix F) for all language measures (i.e., Lexical Training, Grammatical Comprehension, and GJT) confirmed that all main effects and interactions, including the results of the post-hoc analyses, remained unchanged, with no effects emerging or disappearing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by an Alexander von Humboldt Professorship (ID-1195918) awarded to the fourth author.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, *116*(1), 143–148. https://doi.org/10.1016/j.cognition.2010.04.001

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning: An introduction. *Studies in Second Language Acquisition*, *37*(2), 185–196. https://doi.org/10.1017/S027226311500008X

Antoniou, M., Ettlinger, M., & Wong, P. C. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLoS One*, *11*(7), e0158812. https://doi.org/10.1371/journal.pone.0158812

Anwyl-Irvine, A. L., Massonié, J., Flitton, A., Kirkham, N. Z., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

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*(1711), 20160047. https://doi.org/10.1098/rstb.2016.0047

Ballard, J. C. (2001). Assessing attention: Comparison of response-inhibition and traditional continuous performance tests. *Journal of Clinical and Experimental Neuropsychology*, *23*(3), 331–350. https://doi.org/10.1076/jcen.23.3.331.1188

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), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2022). *MuMIn: Multi-model inference* (R package version 1.47.5). https://cran.r-project.org/web/packages/MuMIn/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*, *11*(3), 482–503. https://doi.org/10.1111/tops.12420

Bo, J., Borza, V., & Seidler, R. (2009). Age-related declines in visuospatial working memory correlate with deficits in explicit motor sequence learning. *Journal of Neurophysiology*, *102*(5), 2744–2754. https://doi.org/10.1152/jn.00393.2009

Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*(1), 11–47. https://doi.org/10.1111/lang.12368

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & Cognition*, *41*(2), 281–296. https://doi.org/10.3758/s13421-012-0262-9

Buffington, J., Demos, A. P., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, *43*(3), 635–662. https://doi.org/10.1017/S0272263121000127

Carter, L., Russell, P. N., & Helton, W. S. (2013). Target predictability, sustained attention, and response inhibition. *Brain and Cognition*, *82*(1), 35–42. https://doi.org/10.1016/j.bandc.2013.02.002

Chen, Y., Li, L., Wang, M., & Wang, R. (2022). Which cognitive factors predict L2 grammar learning: Cognitive control, statistical learning, working memory, or attention?. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.943988

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

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

Corkum, V., Byrne, J. M., & Ellsworth, C. (1995). Clinical assessment of sustained attention in preschoolers. *Child Neuropsychology*, *1*(1), 3–18. https://doi.org/10.1080/09297049508401338

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., Pujol, J., & Rodríguez-

Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*, *48*(3), 541–553. https://doi.org/10.1016/j.neuroimage.2009.06.069

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & H. M. Long (Eds.), *The handbook of second language acquisition* (pp. 312–348). Blackwell Publishers. https://doi.org/10.1002/9780470756492.ch11

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). Routledge. https://doi.org/10.4324/9780203628942-11

DeKeyser, R. M. (2016). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition*, *38*(2), 353–363. https://doi.org/10.1017/S0272263116000024

DeKeyser, R. M. (2022). Age effects in naturalistic and instructed second language acquisition. Two sides of one coin. In A. G. Benati & J. W. Scwieter (Eds.), *Second language acquisition theory: The legacy of professor Michael H. Long* (pp. 197–210). John Benjamins. https://doi.org/10.1075/bpa.14.10dek

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review*, *19*(1), 42–68. https://doi.org/10.1075/aila.19.05dor

Ellis, N. C. (1994). Implicit and explicit language learning: An overview. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). Academic Press.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305–352. https://doi.org/10.1017/S027226310505014X

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194. https://doi.org/10.1093/applin/aml015

Ellis, N. C. (2016). Salience, cognition, language complexity, and complex adaptive systems. *Studies in Second Language Acquisition*, *38*(2), 341–351. https://doi.org/10.1017/S027226311600005X

Ellis, N. C. (2022). Second language learning of morphology. *Journal of the European Second Language Association*, *6*(1), 34–59. https://doi.org/10.22599/jesla.85

Ellis, N. C., & Wulff, S. (2020). Usage-based approaches to L2 acquisition. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (pp. 63–82). Routledge. https://doi.org/10.4324/9780429503986-4

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174–180. https://doi.org/10.1016/j.copsyc.2019.03.005

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. (2016). The relationship between artificial and second language learning. *Cognitive Science*, *40*(4), 822–847. https://doi.org/10.1111/cogs.12257

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, *22*(1), 95–110. https://doi.org/10.3758/BF03202765

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

Godfroid, A. (2022). Hypotheses about the interface between explicit and implicit knowledge in second language acquisition. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 709–723). Routledge. https://doi.org/10.4324/9781003018872

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483–517. https://doi.org/10.1017/S0272263113000119

Godfroid, A., & Kim, K. M. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*(3), 606–634. https://doi.org/10.1017/S0272263121000085

Godfroid, A., Loewen, S., Jung, S., Park, J. H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*(2), 269–297. https://doi.org/10.1017/S0272263114000850

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*(4), 665–703. https://doi.org/10.1111/lang.12018

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, *41*(2), 313–336. https://doi.org/10.1017/S0272263118000256

Granena, G. (2020). *Implicit language aptitude*. Cambridge University Press. https://doi.org/10.1017/9781108625616

Granena, G., & Yilmaz, Y. (2018). Aptitude-treatment interaction in L2 learning: A research synthesis. *Studies in English Education*, *23*(4), 803–830. https://doi.org/10.22275/SEE.23.4.02

Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, *36*(4), 611–645. https://doi.org/10.1017/S0272263113000727

Gries, S. T. (2021). (Generalized linear) mixed-effects modeling: A learner corpus example. *Language Learning*, *71*(3), 757–798. https://doi.org/10.1111/lang.12448

Haider, H., & Frensch, P. A. (2005). The generation of conscious awareness in an incidental learning situation. *Psychological Research*, *69*(5-6), 399–411. https://doi.org/10.1007/s00426-004-0209-2

Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences*, *44*, 9–15. https://doi.org/10.1016/j.lindif.2015.10.003

Hamrick, P., Lum, J. A., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences*, *115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Head, J., & Helton, W. S. (2012). Natural scene stimuli and lapses of sustained attention. *Consciousness and Cognition*, *21*(4), 1617–1625. https://doi.org/10.1016/j.concog.2012.08.009

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition.

Studies in Second Language Acquisition, *38*(2), 163–175. https://doi.org/10.1017/S0272263116000176

Huang, Y., & Ferreira, F. (2020). The application of signal detection theory to acceptability judgments. *Frontiers in Psychology*, *11*, 73. https://doi.org/10.3389/fpsyg.2020.00073

Hubbard, R. J., & Federmeier, K. D. (2021). Dividing attention influences contextual facilitation and revision during language comprehension. *Brain Research*, *1764*, 147466. https://doi.org/10.1016/j.brainres.2021.147466

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*(2), 129–140. https://doi.org/10.1017/S0272263105050084

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition*, *39*(3), 401–430. https://doi.org/10.1017/S027226311600019X

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, 105123. https://doi.org/10.1016/j.cognition.2022.105123

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*(7), e12848. https://doi.org/10.1111/cogs.12848

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, *68*(4), 710–730. https://doi.org/10.1080/17470218.2014.964736

Kalra, P. B., Gabrieli, J. D., & Finn, A. S. (2019). Evidence of stable individual differences in implicit learning. *Cognition*, *190*, 199–211. https://doi.org/10.1016/j.cognition.2019.05.007

Karuza, E. A., Emberson, L. L., & Aslin, R. N. (2014). Combining fMRI and behavioral measures to examine the process of human learning. *Neurobiology of Learning and Memory*, *109*, 193–206. https://doi.org/10.1016/j.nlm.2013.09.012

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

Kenanidis, P., Dąbrowska, E., Llompart, M., & Pili-Moss, D. (2023). Can adults learn L2 grammar after prolonged exposure under incidental learning conditions? *PLoS One*, *18*(7), e0288989. https://doi.org/10.1371/journal.pone.0288989

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Krause, M., Mogalle, M., Pohl, H., & Williams, J. J. (2015). A playful game changer: Fostering student retention in online education with social gamification. *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, 95–102. https://doi.org/10.1145/2724660.2724665

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*(4), 870–900. https://doi.org/10.1037/a0030694

Lenth, R. V. (2021). *Emmeans: Estimated marginal means, aka least-squares means* (R package version 1.7.4-1). https://CRAN.R-project.org/package/emmeans

Leung, J. H., & Williams, J. N. (2014). Crosslinguistic differences in implicit language learning. *Studies in Second Language Acquisition*, *36*(4), 733–755. https://doi.org/10.1017/S0272263114000333

Li, S. (2016). The construct validity of language aptitude. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/10.1017/S027226311500042X

Li, S., & DeKeyser, R. M. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence: Introduction to the special issue. *Studies in Second Language Acquisition*, *43*(3), 473–497. https://doi.org/10.1017/S0272263121000024

Li, S., Hiver, P., & Papi, M. (2022). *The Routledge handbook of second language acquisition and individual differences*. Taylor & Francis. https://doi.org/10.4324/9781003270546

Lüdecke, D. (2018). Ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

MacLean, K. A., Aichele, S. R., Bridwell, D. A., Mangun, G. R., Wojciulik, E., & Saron, C. D. (2009). Interactions between endogenous and exogenous attention during vigilance. *Attention, Perception, & Psychophysics*, *71*(5), 1042–1058. https://doi.org/10.3758/APP.71.5.1042

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Psychology Press. https://doi.org/10.4324/9781410611147

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press. https://doi.org/10.1093/oso/9780195151770.003.0004

MacWhinney, B. (2022). Synthesis: The psycholinguistics of learning. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics*. Routledge. https://doi.org/10.4324/9781003018872

Maie, R., & DeKeyser, R. M. (2020). Conflicting evidence of explicit and implicit knowledge from objective and subjective measures. *Studies in Second Language Acquisition*, *42*(2), 359–382. https://doi.org/10.1017/S0272263119000615

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *The Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

McDonough, K. I. M., & Trofimovich, P. (2016). The role of statistical learning and working memory in L2 speakers' pattern learning. *The Modern Language Journal*, *100*(2), 428–445. https://doi.org/10.1111/modl.12331

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Meara, P. M. (2005). *LLAMA language aptitude tests: The manual*. Lognostics. http://www.lognostics.co.uk/tools/llama/llama_manual.pdf

Mirsky, A. F., Anthony, B. J., Duncan, C. C., Ahearn, M. B., & Kellam, S. G. (1991). Analysis of the elements of attention:

A neuropsychological approach. *Neuropsychology Review*, *2*(2), 109–145. https://doi.org/10.1007/BF01109051

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

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143–182. https://doi.org/10.1016/j.cognition.2004.09.001

Morgan-Short, K. (2020). Insights into the neural mechanisms of becoming bilingual: A brief synthesis of second language research with artificial linguistic systems. *Bilingualism: Language and Cognition*, *23*(1), 87–91. https://doi.org/10.1017/S1366728919000701

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, *17*(1), 56–72. https://doi.org/10.1017/S1366728912000715

Morgan-Short, K., Hamrick, P., & Ullman, M. T. (2022). Declarative and procedural memory as predictors of second language development. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 67–81). Routledge. https://doi.org/10.4324/9781003270546-6

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. John Benjamins. https://doi.org/10.1075/sibil.40

Parsons. (2021). Splithalf: Robust estimates of split half reliability. *Journal of Open Source Software*, *6*(60), 3041. https://doi.org/10.21105/joss.03041

Perruchet, P. (2021). Why is the componential construct of implicit language aptitude so difficult to capture?: A commentary on the special issue. *Studies in Second Language Acquisition*, *43*(3), 677–691. https://doi.org/10.1017/S027226312100019X

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

Pili-Moss, D. (2022). Long-term memory predictors of adult language learning at the interface between syntactic form and meaning. *PLoS One*, *17*(10), e0275061. https://doi.org/10.1371/journal.pone.0275061

Pili-Moss, D., Brill-Schuetz, K. A., Faretta-Stutenberg, M., & Morgan-Short, K. (2020). Contributions of declarative and procedural memory to accuracy and automatization during second language practice. *Bilingualism: Language and Cognition*, *23*(3), 639–651. https://doi.org/10.1017/S1366728919000543

Pronk, T. (2021). *splithalfr: Estimates split-half reliabilities for scoring algorithms of cognitive tasks and questionnaires* (Version 2.2.0) [Computer software]. https://github.com/tpronk/splithalfr/

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195106589.001.0001

Reber, A. S., & Allen, R. (2000). Individual differences in implicit learning: Implications for the evolution of consciousness. In R. G. Kunzendorf & B. Wallace (Eds.), *Individual differences in conscious experience* (pp. 227–247). John Benjamins. https://doi.org/10.1075/aicr.20.11reb

Reber, P. J. (2008). Cognitive neuroscience of declarative and nondeclarative memory. *Advances in Psychology*, *139*, 113–123. https://doi.org/10.1016/S0166-4115(08)10010-3

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

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595–626. https://doi.org/10.1111/lang.12010

Rebuschat, P., Monaghan, P., & Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition*, *206*, 104475. https://doi.org/10.1016/j.cognition.2020.104475

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*(4), 829–856. https://doi.org/10.1017/S0142716411000580

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*(4), 425–469. https://doi.org/10.1207/s15516709cog2204_2

Ren, J., Wang, M., & Arciuli, J. (2023). A meta-analysis on the correlations between statistical learning, language, and reading outcomes. *Developmental Psychology*, *59*(9), 1626–1644. https://doi.org/10.1037/dev0001577

Rogers, J., Révész, A., & Rebuschat, P. (2016). Implicit and explicit knowledge of inflectional morphology. *Applied Psycholinguistics*, *37*(4), 781–812. https://doi.org/10.1017/S0142716415000247

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., Jr., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*(5), 343–350. https://doi.org/10.1037/h0043220

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Shanks, David R, & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367–395. https://doi.org/10.1017/S0140525X00035032

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass*, *14*(3), e12365. https://doi.org/10.1111/lnc3.12365

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

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

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning*. John Benjamins. https://doi.org/10.1075/lllt.2.06ske

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*(2), 293–321. https://doi.org/10.1017/S0142716404001146

Su, C. H., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, *31*(3), 268–286. https://doi.org/10.1111/jcal.12088

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*(4), 860–895. https://doi.org/10.1111/lang.12138

Suzuki, Y., DeKeyser, R., & Huang, Y. T. (2023). Implicit (not explicit) learning aptitude predicts the acquisition of difficult (not easy) structure. A visual-world eye tracking study. In W. Zhisheng, P. Skehan, & R. L. Sparks (Eds.), *Language aptitude theory and practice* (pp. 381–413). Cambridge University Press. https://doi.org/10.1017/9781009076463.020

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792–814. https://doi.org/10.1037/a0030801

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 953–968). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

Ullman, M. T. (2020). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition* (pp. 128–161). Routledge. https://doi.org/10.4324/9780429503986-7

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Walker, N., Monaghan, P., Schoetensack, C., & Rebuschat, P. (2020). Distinctions in the acquisition of vocabulary and grammar: An individual differences approach. *Language Learning*, *70*(S2), 221–254. https://doi.org/10.1111/lang.12395

West, G., Shanks, D. R., & Hulme, C. (2021). Sustained attention, not procedural learning, is a predictor of Reading, language and arithmetic skills in children. *Scientific Studies of Reading*, *25*(1), 47–63. https://doi.org/10.1080/10888438.2020.1750618

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *New handbook of second language acquisition* (pp. 319–353). Emerald Group Publishing Limited.

Williams, J. N., & Rebuschat, P. (2023). Implicit learning and second language acquisition: A cognitive psychology perspective. In A. Godfroid & H. Hopp (Eds.), *The routledge handbook of second language acquisition and psycholinguistics* (pp. 281–293). Routledge. https://doi.org/10.4324/9781003018872-27

Wilson, K., & Korn, J. H. (2007). Attention during lectures: Beyond ten minutes. *Teaching of Psychology*, *34*(2), 85–89. https://doi.org/10.1080/00986280701291291

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review*, *119*(1), 21–39. https://doi.org/10.1037/a0026182

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer. https://doi.org/10.1007/978-0-387-87458-6