University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Perfors, A;Kidd, E

Title:
The Role of Stimulus-Specific Perceptual Fluency in Statistical Learning

Date:
2022-02-01

Citation:
Perfors, A. & Kidd, E. (2022). The Role of Stimulus-Specific Perceptual Fluency in Statistical Learning. Cognitive Science, 46 (2), https://doi.org/10.1111/cogs.13100.

Persistent Link:
https://hdl.handle.net/11343/308516

License:
CC BY-NC-ND

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13100
© 2022 The Authors. Cognitive Science published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).
ISSN: 1551-6709 online
DOI: 10.1111/cogs.13100

The Role of Stimulus-Specific Perceptual Fluency in Statistical Learning

Andrew Perfors,[a] Evan Kidd[b]

[a]*School of Psychological Sciences, University of Melbourne, ARC Centre of Excellence for the Dynamics of Language*
[b]*Max Planck Institute for Psycholinguistics, Research School of Psychology, The Australian National University, ARC Centre of Excellence for the Dynamics of Language*

Received 3 September 2020; received in revised form 2 January 2022; accepted 6 January 2022

Abstract

Humans have the ability to learn surprisingly complicated statistical information in a variety of modalities and situations, often based on relatively little input. These statistical learning (SL) skills appear to underlie many kinds of learning, but despite their ubiquity, we still do not fully understand precisely what SL is and what individual differences on SL tasks reflect. Here, we present experimental work suggesting that at least some individual differences arise from stimulus-specific variation in perceptual fluency: the ability to rapidly or efficiently code and remember the stimuli that SL occurs over. Experiment 1 demonstrates that participants show improved SL when the stimuli are simple and familiar; Experiment 2 shows that this improvement is not evident for simple but unfamiliar stimuli; and Experiment 3 shows that for the same stimuli (Chinese characters), SL is higher for people who are familiar with them (Chinese speakers) than those who are not (English speakers matched on age and education level). Overall, our findings indicate that performance on a standard SL task varies substantially within the same (visual) modality as a function of whether the stimuli involved are familiar or not, independent of stimulus complexity. Moreover, test–retest correlations of performance in an SL task using stimuli of the same level of familiarity (but distinct items) are stronger than correlations across the same task with stimuli of different levels of familiarity. Finally, we demonstrate that SL performance is predicted by an independent measure of stimulus-specific perceptual fluency that contains

[Correction added on 09 May, 2022 after first online publication: CAUL funding statement has been added.]

Correspondence should be sent to Andrew Perfors, School of Psychological Sciences, University of Melbourne, ARC Centre of Excellence for the Dynamics of Language. Complex Human Data Hub 700 Swanston St, Level 8 Parkville, VIC 3010 Australia E-mail: perfors@gmail.com

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

no SL component at all. Our results suggest that a key component of SL performance may be related to stimulus-specific processing and familiarity.

Keywords: statistical learning; perceptual fluency; domain specificity; familiarity; complexity

1. Introduction

Statistical learning (SL) refers to the ability to adapt to and learn from the probabilistic structure of the environment. This ability is phylogenetically old (Hauser, Newport, & Aslin, 2001) and appears to be operational quite early in development (Bulf, Johnson, & Valenza, 2011; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009). It occurs across multiple modalities and across a wide variety of types of stimuli (Brady & Oliva, 2008; Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015; Fiser & Aslin, 2002; Gebhart, Newport, & Aslin, 2009; Kirkham, Slemmer, & Johnson, 2002; Krogh, Vlach, & Johnson, 2013; Saffran, Aslin, & Newport, 1996). Moreover, recent evidence suggests that individual differences in SL abilities are stable and reliable (Isbilen, McCauley, Kidd, & Christiansen, 2017, 2020; Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018). Taken together, these findings appear to indicate that SL is a flexible and broadly applicable capacity for extracting and using the regularities in the world.

Because of this broad applicability, for some time, the field implicitly adopted a "unitary" conception of SL under which it is a domain-general mechanism that operates similarly across modalities. However, as Frost, Armstrong, and Christiansen (2019) note, a unitary conception is inconsistent with more recent research, suggesting that SL involves substantial modality specificity. For instance, people find it easier to simultaneously acquire two artificial grammars when those grammars are presented in different modalities, a result that is difficult to accommodate under a single mechanism view (Conway & Christiansen, 2006). Similarly, Siegelman and Frost (2015) found no association between performance on SL tasks involving visual and auditory linguistic stimuli. Interestingly, Siegelman et al. (2018) reported a significant association between visual SL and SL for non-linguistic sounds, but once again not with linguistic stimuli. They suggested that these results reflect linguistic entrenchment, theorizing that while participants came to their studies without prior knowledge and expectations regarding how visual shapes and non-linguistic sounds were distributed, their life-long experience led to expectations regarding how syllables patterned to form words. Consistent with this assertion, independent ratings of the "wordiness" of novel linguistic stimuli positively predicted performance: stimuli that were more similar to native-language words were more easily acquired through SL.

These data suggest that both modality specificity and domain independence play a role in SL, with prior learning constituting an important yet understudied component of the process. As a result, SL has more recently been conceptualized as composed of multiple components (Arciuli, 2017; Frost et al., 2019; Siegelman, Bogaerts, & Frost, 2017) that are grounded in basic perception and cognition (Christiansen, 2019; Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Isbilen et al., 2020). However, the details of those components remail unclear,

as Siegelman et al. (2017) note: "in contrast with general intelligence … the dimensions of SL as an individual ability are yet to be empirically established" (p. 5). We explore this issue here.

Frost, Armstrong, Siegelman, and Christiansen (2015) suggested that variance in the performance on SL tasks comes from *two* main sources—not only (a) variability in the efficiency with which domain-general learning mechanisms detect the statistical properties of the input stream, but also (b) variability in the efficiency of perceptual encoding within each modality. This second component, which we refer to here as *perceptual fluency* (PF), reflects something like the ease and accuracy with which people can encode and recall the items whose statistical regularities are being learned. Relatively little is known exactly what factors contribute to PF and why—indeed, this is part of the motivation for our work here—but as a first approximation, we suggest that it likely reflects aspects of both processing efficiency and memory (e.g., encoding, retrieval, and span). The logic is as follows: if people with higher PF find it easier to process and/or remember individual items, one might expect them to have more capacity "left over" that they can use to acquire the statistical regularities between those items. Individual differences in PF across modality might thus explain how SL tasks could be uncorrelated across modalities even though they depend on similar domain-general computations.

One test of this dual-source theory came from presenting participants with visual SL tasks that manipulated probabilistic structure and speed of presentation (Bogaerts, Siegelman, & Frost, 2016). As predicted, both variables affected learning, with higher performance observed for stimuli with higher transition probabilities (TPs) and longer exposure periods. People who benefited most from the longer inspection times were also better at learning the underlying statistical structure. This suggests that an individual's PF may afford a significant advantage in SL, and is consistent with the observation that at least one measure of SL has been found to be associated with psychometric measures of processing speed (Kaufman et al., 2010).

But what makes a person have high PF on a given set of stimuli? One possibility, as Frost et al. (2015) theorized, is that it is modality-specific: different people's brains process stimuli differently in different modalities. Another possibility is that fluency increases with the amount of time available (Arciuli & Simpson, 2011; Bogaerts et al., 2016; Turk-Browne, Junge, & Scholl, 2005). These factors are well studied and likely to play an important role.

In this paper, we investigate another factor. We hypothesize that individual differences in perceptual fluency—and thus statistical learning—might also reflect differences in how people process *specific stimulus items* rather than differences due to task or modality. We focus on two possible properties of those items (complexity and familiarity) and explore how and to what extent they affect PF and—via that—SL in the visual modality.

There is already substantial evidence that both complexity and familiarity affect PF. More complex stimuli take longer to process and require more working memory capacity (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005; Kemps, 1999; Liu, Chen, Liu, & Fua, 2012; Luria, Sessa, Gotler, Jolicoeur, & Dell'Acqua, 2010). They are also more appealing (Madan, Bayer, Gamer, Lonsdorf, Sommer, 2018) and attentionally engaging (Sun Firestone, 2021). Complexity affects learning as well: new symbols are easier to acquire when

they are visually simpler (Pelli, Burns, Farell, & Moore-Page, 2006). However, it is difficult to disentangle complexity from familiarity (Popov & Reder, 2020; Zhang, Liu, So, & Reder, 2020). Ratings of the two are often negatively correlated, suggesting that familiarity causes people consider things to be less complex (Bonin, Peereman, Malardier, Meot, & Chalard, 2003; Cycowicz, Friedman, Rothstein, & Snodgrass, 1997). Moreover, the effects of complexity on processing and memory are often substantially reduced, if not eliminated entirely, once familiarity is taken into account (Bethell-Fox & Shepard, 1988; Jackson & Raymond, 2008; Ngiam, Khaw, & Goodbourn, 2019; Reder, Liu, Keinath, & Popov, 2016; Sun, Zimmer, & Fu, 2011; Xie & Zhang, 2017; Zhang et al., 2020). This may be because items use working memory resources in a way that is inversely proportional to their strength in long-term memory (Popov & Reder, 2020). Thus, a complex item that is highly familiar and therefore strongly represented in long-term memory may require less working memory capacity than one that has never been seen before.

Our main question is whether statistical learning is improved when using items that people have higher perceptual fluency for (like simpler or more familiar ones). We expect higher PF to improve SL by freeing up cognitive capacity that can then be devoted to learning statistical regularities. This yields several predictions. First, SL performance should be higher when the items involved are simpler and/or more familiar. Second, to the extent that PF reflects stimulus-specific characteristics like familiarity or complexity rather than modality-specific abilities, correlations between SL tasks should be higher when the stimuli share those characteristics (even when the modality is the same). Finally, SL performance should be predicted by performance on a task that measures stimulus-specific PF but requires no SL at all. We test and find support for all of these predictions here.

It is worth noting that although our hypothesis is that individual differences in SL are moderated by PF—which is shaped at least in part by familiarity—our work is distinct from other research that also suggests that prior experience plays a role in SL of linguistic information(for example, Endress & Mehler, 2009; Gebhart et al., 2009; Perruchet & Poulin-Charronnat, 2012; Siegelman et al., 2018; but see also Potter, Wang, & Saffran, 2017; Wang & Saffran, 2014). Our focus here is on the role that familiarity may play in making *individual stimulus items* easier to remember and parse; we thus predict that previous exposure should *facilitate* SL in any modality. The focus of previous work is on the role that prior experience plays in learning *statistical regularities between items*; it suggests that previous linguistic exposure may *impair* SL due to interference and entrenchment, and the focus is not on the items so much as the TPs between them. Our approach complements this other research because both focus on the role of prior experience in shaping SL; it is distinct because it has a different mechanism and makes different predictions.

In the first of our three experiments, we show (a) that SL is better for stimuli that are both simpler and more familiar; and (b) perceptual fluency on those stimuli predicts SL. Experiments 2 and 3 are aimed at disentangling the impact of complexity versus familiarity. We find that SL and PF are still correlated and both are higher when stimuli are familiar, regardless of their visual complexity.

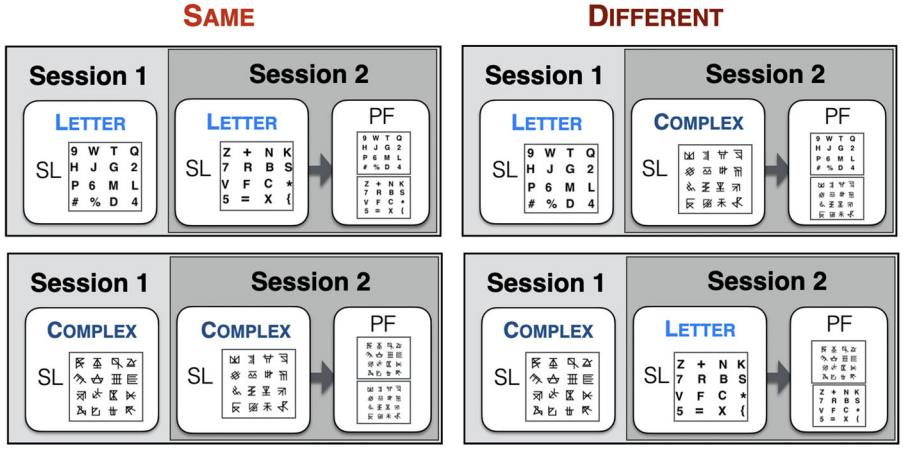


Fig. 1. Overview of Experiment 1 structure. Each participant was tested in two sessions. The first involved a visual statistical learning (SL) task with either LETTER or COMPLEX stimuli (see Figs. 2 and 3 for details). In the second session, the same participants completed another SL task with new stimuli that were either of the same or different type (although the specific stimulus items were always different). Following that, each person performed two consecutive tasks that measured their perceptual fluency (PF) on the two stimulus sets they had seen (Fig. 4).

2. Experiment 1: Method

This research tests the hypothesis that individual differences in SL are influenced by individual differences in PF. We investigated this by having the same individuals participate in two different visual SL tasks on two different days, as shown in Fig. 1. On each day, people were randomly assigned to one of two conditions (LETTER or COMPLEX) in which the SL task was the same and the only difference was the stimuli involved. The LETTER condition involved highly familiar and very simple stimuli (letters and common symbols), while the COMPLEX condition involved novel and visually complex shapes. On each day, people were randomly assigned to a condition; thus, some people saw the same type of stimuli on each day (although distinct items each time), while others saw the LETTER stimuli on 1 day and the COMPLEX stimuli on another. If complexity and/or familiarity play an important mediating role, we would expect SL performance to be higher for the LETTER stimuli. Additionally, we also expect there to be a higher test–retest correlation across sessions when the stimulus type is the same.

In addition to the SL tasks that occurred during both experimental sessions, Session 2 also presented participants with two measures of PF for the stimuli they had previously seen (the first PF task used the stimuli from Session 1, and the second used the stimuli from Session 2). The PF task, described in more detail below, was designed to measure the speed with which these particular stimuli could be parsed and encoded while involving no SL at all. If PF plays an important role in SL, we would expect that PF task performance would be correlated with SL task performance.

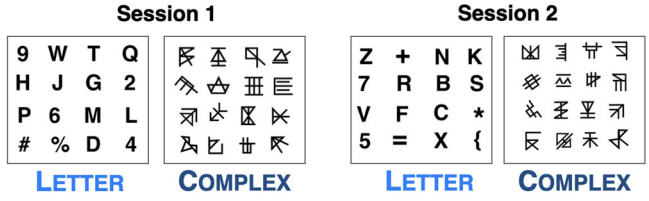


Fig. 2. Experiment 1 stimuli. At each session, participants saw 16 stimulus items that were either LETTER or COMPLEX, depending on randomly assigned condition. Although some people saw stimuli of the same type in both sessions, no single item occurred both times.

2.1. Participants

For Session 1, 160 participants were recruited from Amazon Mechanical Turk for the 15-minute task, for which they were paid $3.50USD. Of these, 14 were excluded for failing the attention check (described below). The remaining 146 were invited to return for a second 20-minute session a week later, for which they were paid $4USD. One hundred thirty five returned, three of whom failed the attention check. All analyses focus on the remaining 132 participants, 73 (55.3%) of whom were male and 129 (97.7%) were from the US. Ages ranged from 20 to 69 (mean 36.1). At each session, participants were randomly assigned to either LETTER or COMPLEX stimulus sets of 16 items each. This resulted in 53 people who saw the SAME stimulus complexity each time (17 LETTER, 36 COMPLEX) and 79 people who saw a DIFFERENT stimulus complexity each time (44 saw LETTER first, 35 saw COMPLEX first).

2.2. Materials

As shown in Fig. 2, the COMPLEX stimuli were created by combining between four and six straight lines, resulting in novel shapes that were perceptually discriminable but effortful to parse and remember. The LETTER stimuli consisted of letters and symbols. We chose these because they are highly overlearned and very familiar to any literate English speaker. To minimize the potential for "chunking" letter combinations into existing words, no vowels or vowel-like symbols like @ or & were included. Moreover, all participants in all conditions saw a different random combination of the stimuli into triples, as described below.

2.3. Procedure

2.3.1. Statistical learning task

The SL task in both sessions follows a standard design, as illustrated in Fig. 3. In it, participants see that symbols appear on the center of their screen one-by-one; SL involves learning the probability with which different symbols follow each other. Some symbols were embedded in "easy" triplets in which the TP of seeing one symbol given the previous one was 1.0. In the "hard" triplets, the TP was much lower, though still non-zero. After a training during which participants simply watched the sequence of symbols, they were given a test designed to measure the extent to which they had learned the TPs.

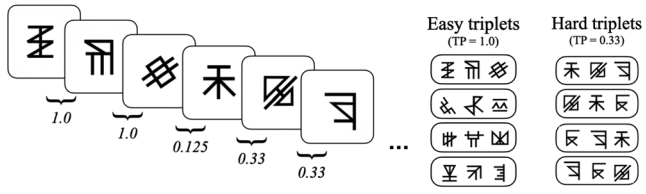


Fig. 3. Statistical learning task. All experiments used the same statistical learning task, a standard one adapted from Siegelman et al. (2017). In it, participants see individual symbols on the screen one by one for 800 ms each; statistical learning involves acquiring information about which symbols follow each other and with what probability. There are eight triplets of symbols, each randomly occurring 24 times within the 10-minute training sequence. For easy triplets, the TP between symbols is 1.0, meaning that a given symbol always follows the same other symbol. For the hard triplets, the TP between symbols is 0.33. Performing well on the test phase requires having learned these TPs. The triplets shown here are illustrative only; symbols were randomly assigned for each participant.

The specific details of our task were nearly identical to Siegelman et al. (2017), which was designed to have good psychometric properties including reasonable test–retest reliability and the ability to differentiate between individuals. Besides the specific stimuli used, because we were running on MTurk, the only difference between our task and theirs is that we included an attention check by embedding four or five English words in the training sequence. (In Session 1, these words were *train*, *boy*, *lion*, and *walk*. In Session 2, they were *boat*, *lady*, *koala*, *tree*, and *doll*.) Before training, participants were told that they should simply watch the sequence, but that to ensure attention we had included a few English words among the symbols; they were not informed about how many there would be. They were told to write down any English words as they occurred and report them in a text box at the end of training. People who got fewer than three correct in either session were excluded from all analyses.

Following Siegelman et al. (2017), the SL task itself consisted of 16 items combined into eight triplets, each repeated 24 times in random order. The task is divided into two phases: (a) a familiarization phase and (b) a test phase. During the familiarization phase, each item appeared alone at the center of the participant's screen for 800 ms with a 200 ms interstimulus interval. The eight triplets were designed so that four were easy, defined by TPs of 1.0 between each item (x_5-x_6-x_7, x_8-x_9-x_{10}, x_{11}-x_{12}-x_{13}, x_{14}-x_{15}-x_{16}), while the other four were harder, defined by TPs of 0.33 between items (x_1-x_2-x_3, x_2-x_1-x_4, x_4-x_3-x_1, x_3-x_4-x_2). The mapping of each image to each item, as well as the order of the triplets and the location of the embedded English words, was randomized for each participant at each session (thus the TP between triplets was $1/8 = 0.125$). The presentation of triplets was constrained so that triplets were never immediately repeated. The familiarization phase lasted 10 minutes.

The test phase immediately followed. We administered the exact same 42-item test as in Siegelman et al. (2017). In contrast to many common SL tasks, which typically use two-alternative forced choice (2AFC) as a sole measure of learning, the task uses an array of different method to measure learning. These were: (a) 16 2AFC trials, where participants had to choose whether they had seen either an attested triplet or an unattested foil; (b) eight 4AFC trials, where participants had to choose the correct triplet given three foils; (c) six 2AFC trails

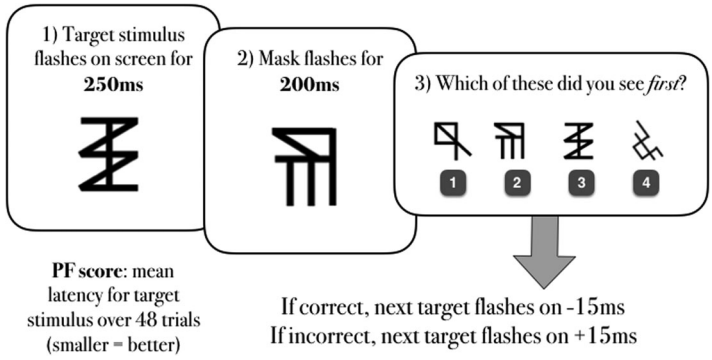


Fig. 4. Perceptual fluency task. During each trial of the PF task, a target stimulus is flashed briefly onscreen, followed by a masking stimulus for 200 ms. Participants are then presented with four items (the target, the mask, and two others randomly selected from the stimulus set). If they successfully identify the target, on the next trial, the duration of target presentation decreases by 15 ms; if not, it increases by 15 ms. The PF score (PFscore) reflects the mean target latency over all 48 trials. Smaller PFscore indicates better PF, because it indicates that the person can attain similar accuracy based on a shorter target presentation time.

where participants had to choose which pair of shapes they had seen, where the correct one had TPs of either 0.33 or 1.0 and the incorrect one had lower TPs; (d) four 4AFC "pair" trials, which were identical but had three distractor pairs; and (e) eight pattern completion trials, where participants were required to complete four triplets patterns and four pair patterns by selecting the correct symbol.

For each participant, we calculated an overall SL accuracy score, SLscore, which reflects the proportion of the 42 test items that they got correct. To evaluate whether our participants were engaging with the task in a way similar to the participants in Siegelman et al. (2017), we calculated the correlation for performance on each item between our participants and theirs. It was significant in both conditions (LETTER: $r = .8$, $p < .0001$; COMPLEX: $r = .68$, $p < .0001$), suggesting that our participants found the same items to be difficult as theirs did.

2.3.2. Perceptual fluency task

The PF task, as illustrated in Fig. 4, is a novel measure we designed to capture the facility with which people could encode and recall the specific stimuli in the experiment. Inspired in part by the Inspection Time literature (see O'Connor Burns, 2003 for an overview), each of the 48 trials in the task involves flashing an item onto the screen for a small amount of time (250 ms on the first trial) followed by a masking stimulus (always presented for 200 ms). The task is adaptive: if the participant successfully identifies the item, the target on the next trial flashes more quickly (with a duration of 15 ms less); if they do not, it flashes slower (with a duration of 15 ms more). Over the course of the task, participants who can achieve similar accuracy for targets displayed for a shorter duration have higher perceptual fluency for those stimuli. This is reflected in their PFscore, which consists of the mean target latency over all trials; a lower PFscore thus reflects higher PF.

Each of the 16 symbols in the relevant stimulus set was the target three times. Masks and distractors were selected from the same stimulus set and randomized by person, as was trial order.

3. Experiment 1: Results

The overall hypothesis being tested in this paper is that PF plays a mediating role in the performance on SL tasks. This hypothesis yields several testable predictions. First, it predicts that people will show different levels of performance on SL in tasks that differ only in terms of which stimuli they use, for which the SL component is identical. Second, it predicts that correlations between SL performance at two different time points should be greater if the stimuli are of the same type than if they are not (even though all are within the same modality). Third, it predicts that SL performance should be correlated with PF, even though no SL is involved in the PF task at all. We consider each prediction in turn. Unless stated otherwise, all analyses include data from both sessions.

3.1. Is performance better on LETTER stimuli?

Do people improve when the stimuli involved are LETTER rather than COMPLEX? We ask this question separately for both SL and PF tasks. Appendix A contains more detail.

3.1.1. Statistical learning

Accuracy on the SL tasks, as measured by SLscore, does appear to span the range of individual differences, ranging between 26.2% and 97.6% (mean: 59.6%) in the LETTER condition and 28.6% and 92.9% (mean: 52.7%) in the COMPLEX condition. In both conditions, performance is significantly above chance, which Siegelman et al. (2017) calculate as 40% (LETTER: $t(112) = 11.45$, $p < .0001$, $d = 1.08$; COMPLEX: $t(150) = 10.69$, $p < .0001$, $d = 0.87$). Moreover, the difference in accuracy between LETTER and COMPLEX conditions, as shown in Fig. 5(a), is significant ($t(211) = 3.24$, $p = .0014$, $d = 0.42$).

3.1.2. Perceptual fluency

As Fig. 5(b) reveals, PF was also significantly improved (as reflected in a lower PFscore) when the stimuli were LETTER than when they were COMPLEX ($t(240.1) = -3.46$, $p = .0006$, $d = 0.43$). This is reassuring both because it helps to support the notion that PFscore is indeed measuring something about perceptual fluency, and because it suggests that participants were behaving sensibly on the PF task in general.

3.2. Are SL correlations higher between stimuli of the same familiarity?

If PF influences performance on SL tasks, then we should expect that test–retest correlations between SL performance on two different sessions should be reasonably high when the stimuli are of the same type (LETTER–LETTER or COMPLEX–COMPLEX), even though no individual items are repeated. Conversely, we should expect lower correlations if they

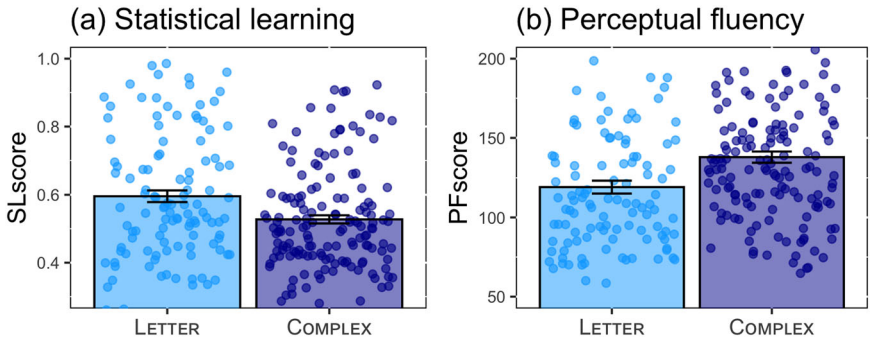


Fig. 5. Performance as a function of stimulus complexity. (a) In the statistical learning task, accuracy (SLscore) was significantly higher for LETTER stimuli than for COMPLEX stimuli, though both were far above chance at approximately 40%. (b) In the perceptual fluency task, mean latency of the target stimulus (PFscore) was significantly lower for LETTER stimuli, indicating that participants were better at quickly encoding and remembering the LETTER stimuli than the COMPLEX ones.

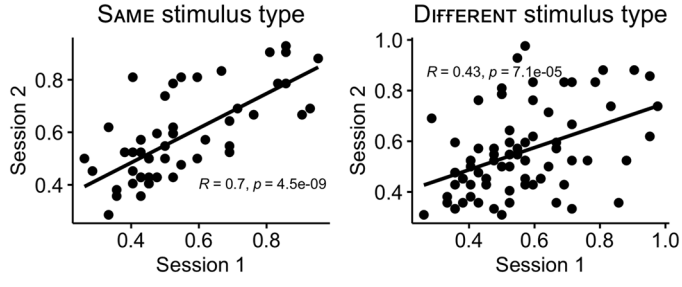


Fig. 6. Test–retest correlations in SLscore as a function of stimulus similarity. The left panel shows the correlation between SLscore on Sessions 1 and 2 for those participants who performed the statistical learning task using stimuli of the same type each time (either LETTER–LETTER or COMPLEX–COMPLEX). The right panel shows the same correlation but for participants who saw a different type of stimuli each time (either LETTER–COMPLEX or COMPLEX–LETTER). Correlations were significantly higher when the stimuli were of the same type, despite the fact that all were in the same modality and no specific symbols were repeated between sessions.

are different at different sessions (LETTER-COMPLEX or COMPLEX–LETTER), even though the modality is the same. We evaluate this prediction in Fig. 6. It is clear that the test–retest correlation between SAME stimuli was higher ($r = .7$, $p < .0001$) than between DIFFERENT sets ($r = .43$, $p < .0001$). The difference between these correlations is significant using a Fisher r-to-z transformation ($z = 2.26$, $p = .024$).

3.3. *Does perceptual fluency predict performance in the statistical learning task?*

A final prediction of the hypothesis that PF influences SL performance is that higher PF (as measured by a lower PFscore) should be associated with higher SL (as measured by a higher SLscore). The correlation between PF and SL accuracy within the same person was indeed significant both across the SAME stimuli ($r = -.37$, $p < .0001$) and DIFFERENT

stimuli ($r = -.28$, $p < .0001$); the difference between these correlations is not significant ($z = -0.76$, $p = .45$).

4. Experiment 2: Method

The results thus far suggest that SL is improved when PF is high, as when stimuli are simpler and more familiar. Moreover, test–retest correlations between SL tasks are higher when the stimuli on both days are of the same type, and performance on a test of PF is both higher for letters than unfamiliar complex shapes, and is correlated with SL performance.

However, an open question remains: to what extent do our results reflect *familiarity*, and to what extent do they reflect *complexity*? We deliberately confounded these two factors in Experiment 1, which means that we cannot tell whether they performed worse on the COMPLEX stimuli because they were novel or complicated or both. We investigate this question in Experiment 2, which presents another group of participants with stimuli that were designed to be simpler than the COMPLEX stimuli in Experiment 1, while still being unfamiliar (we call them SIMPLE). If complexity was driving the results in Experiment 1, we would expect SL performance using the SIMPLE stimuli to be similar to the LETTER stimuli, since both are relatively simple; if not, we would expect performance on the SIMPLE stimuli to be similar to the COMPLEX ones, since both are unfamiliar.

4.1. Participants

Eighty participants were recruited from Amazon Mechanical Turk and paid $3.50USD for the 15-minute task, which was a single session, identical to Session 1 of Experiment 1 but with SIMPLE stimuli. Of these, six were excluded for failing the attention check, which was the same as before. All analyses focus on the remaining 74 participants, 46 (62.2%) of whom were male, and 70 (94.6%) were from the United States. Ages ranged from 21 to 62 (mean 36.4). None were in Experiment 1.

4.2. Materials

Our goal in Experiment 2 was to design stimuli that were not as visually complex as the COMPLEX stimuli in Experiment 1. Defining visual complexity is a difficult task and to our knowledge, there is no fully agreed-upon approach (see, for example, Donderi, 2006; Miton & Morin, 2021; Pelli et al., 2006; and Sun & Firestone, 2021, for discussions of the issues involved). However, most measures capture similar general principles (e.g., that simpler items have fewer and more distinguishable features) and are highly correlated with each other. We therefore designed the SIMPLE stimuli to have fewer features than the COMPLEX ones (between two and four rather than four and six) as well as more distinguishable features (consisting of lines, arcs, and circles rather than just lines). The stimuli are shown in Fig. 7.

To ensure that these stimuli were indeed simpler, we calculated the objective complexity of each stimulus using a measure called *perimetric complexity*. It is defined as the ratio of inked surface to the perimeter of the inked surface and is both commonly used (Miton &

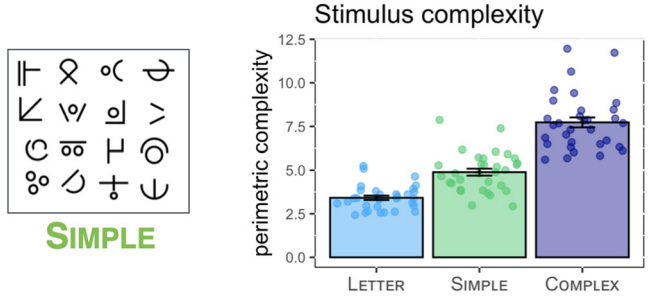


Fig. 7. Experiment 2 stimuli. All participants saw 16 unfamiliar SIMPLE stimulus items (left panel) that were designed to be visually less complex than the COMPLEX ones in Experiment 1. They had fewer and more distinguishable features and scored lower on an objective measure called perimetric complexity (right panel).

Morin, 2021) and correlated with human learning efficiency (Pelli et al., 2006). The calculation was implemented in Mathematica as in Watson (2012). As Fig. 7 shows, the perimetric complexity of the SIMPLE stimuli was significantly lower than that of the COMPLEX stimuli ($t(55.2) = 8.134$, $p < .0001$) and closer to the LETTER stimuli, although still significantly different ($t(52.1) = -6.269$, $p < .0001$). If complexity matters and familiarity does not, one would therefore expect SL performance on the SIMPLE stimuli to be better than on the COMPLEX stimuli and slightly worse than the LETTER stimuli. Conversely, if only familiarity matters, then one would expect that performance on the SIMPLE and COMPLEX stimuli should be similar to each other, and both substantially worse than on the LETTER stimuli.

4.3. Procedure

The procedure was completely identical to the SL task in Session 1 of Experiment 1, except with SIMPLE stimuli instead. As before, participants observed a 10-minute familiarization task in which they saw the stimuli appear individually on screen. Which symbols appeared in which position was again randomized for each participant, and all participants saw a combination of easy and hard triplets. Each of the triplets occurred 24 times in random order, as before, and the attention check was the same. The test phase was also identical, involving the same 42 questions, which measured how well they had learned the transitional probabilities between the SIMPLE stimuli.

5. Experiment 2: Results

As in the previous experiment, accuracy on the SL task (SLscore) using the SIMPLE stimuli spans the range of individual differences, with a mean of 50.5%, a minimum of 26.2%, and a maximum of 95.2%. Performance was significantly above the chance level of 40% ($t(73) = 6.2$, $p < .0001$, $d = 0.73$) and there was once again a high correlation between the item-level accuracy of the participants of Siegelman et al. (2017) and ours ($r = .66$, $p < .0001$).

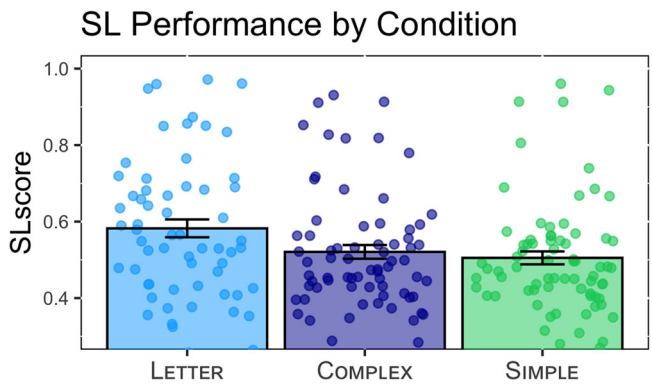


Fig. 8. Experiment 2: SL performance by condition. Accuracy on the SL task (SLscore) was significantly higher for LETTER stimuli than for either the unfamiliar complex stimuli (COMPLEX) or the unfamiliar simple stimuli (SIMPLE), suggesting that it is familiarity rather than complexity that improves statistical learning performance.

Most importantly, as shown in Fig. 8, overall SL performance on the SIMPLE stimuli was more similar to performance on the COMPLEX stimuli than the LETTER stimuli. To make the most appropriate comparison, we performed a one-way ANOVA on SLscore between SIMPLE and the Session 1 SL task performance in the LETTER and COMPLEX conditions (results were not qualitatively different if instead we used the SL scores from both sessions). There was a significant effect of condition ($F(2, 203) = 4.2743$, $p = .0152$, $\eta^2 = 0.04$). Follow-up t-tests indicated that the difference between LETTER and SIMPLE was significant ($t(113.8) = 2.68$, $p = .0085$, $d = 0.47$) but not the difference between COMPLEX and SIMPLE ($t(142.1) = -0.63$, $p = .5271$, $d = 0.11$). In other words, people showed similar levels of SL for both of the unfamiliar stimuli, despite the fact that they differed substantially in complexity. This suggests that *familiarity* rather than *complexity* underlays the SL performance that we observed in Experiment 1.

6. Experiment 3: Method

The results from Experiments 1 and 2 are suggestive, but despite our best efforts, it may have been impossible to fully measure or control for stimulus complexity. Accordingly, it is possible that the SIMPLE stimuli were not actually simpler or easier to process. If this were the case, it would mean that—Experiment 2 notwithstanding—these effects might have been driven by complexity rather than familiarity. This would still be interesting, but would have very different implications for the nature of the mechanisms that underlie SL. In addition, it would be reassuring to replicate these results on a different population of participants.

Motivated by these considerations, we conducted a third pre-registered experiment designed to disentangle complexity and familiarity. In it, we took the opposite approach as in Experiment 2, holding complexity constant but testing participants who varied in their familiarity with the stimuli. Specifically, we used two sets of stimuli: CHINESE characters and the

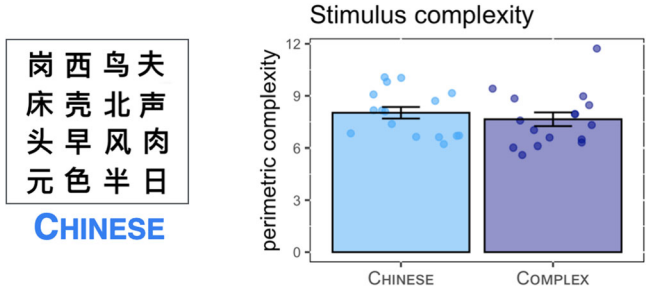


Fig. 9. Experiment 3 stimuli. All participants saw the CHINESE and COMPLEX items in two separate SL and PF tasks. CHINESE and COMPLEX stimuli did not differ in perimetric complexity.

COMPLEX stimuli from Study 1. We tested two different groups of participants: native Chinese speakers (for whom the Chinese characters were overlearned and highly familiar) and English speakers (who had little or no experience with Chinese). If familiarity were driving the effect, one would expect that the Chinese participants would show better SL and higher PF for the CHINESE stimuli, but that all participants would perform equally on the novel COMPLEX stimuli.

6.1. Participants

A total of 206 undergraduates enrolled in first-year psychology at the University of Melbourne participated in this 35–45 minute experiment to receive course credit. Of these, 29 were excluded for failing the pre-registered attention check, which was identical to the one in Experiment 1 and consisted of correctly reporting the English words that had been embedded in the training sequences of the SL tasks. All analyses focus on the remaining 177 participants, 48 (27.1%) of whom were male. Ages ranged from 17 to 27 (mean 19.4).

All participants took part in the exact same experiment as each other, seeing the same stimuli and undergoing the same procedure. The conditions were defined by the participants themselves according to the following pre-registered criteria. First, they took a short four-question quiz in which they were asked to choose the meaning of four different Chinese characters (for questions, see Appendix B). Participants who scored less than three correct were automatically assigned to the English group. Of those who got three or more correct, those who did not rate themselves as native speakers of Chinese were also put into the English group. This left 61 participants in the Chinese group and 116 in the English group. Since all came from the same undergraduate cohort, the groups were matched on education level and did not differ in age ($t(157.37) = 0.39$, $p = .694$).

6.2. Materials

As described below, every participant in Experiment 3 took part in the same two SL and PF tasks as each other. In one, they saw the COMPLEX stimuli from the first session of Experiment 1. In the other, the stimuli were common CHINESE characters, as shown in Fig. 9. The characters varied in the number of strokes and pronunciation (based on initial phoneme and

tone) and all corresponded to simple nouns. Appendix B contains the full list of characters. There was no difference in the perimetric complexity of the CHINESE and COMPLEX stimuli in this experiment ($t(29.2) = 0.722$, $p = .476$).

6.3. Procedure

The procedure for this experiment was similar to Experiment 1 but had a few important differences resulting from several considerations. First, in order to recruit from the undergraduate population pool, it was necessary to have an experiment that could be completed in one session; use of this pool was important as it contained many native Chinese speakers who matched the non-Chinese speakers on age and education level. Second, the important comparisons for this experiment were between groups rather than between tasks, which necessitated a different structure.

We therefore presented each participant with two SL tasks, one with each type of stimuli (COMPLEX and CHINESE), with the order randomized. Each person then also completed two PF tasks with the same stimuli, in the same order as in the SL tasks. Thus, a participant who saw the CHINESE stimuli on the first SL task saw them in the first PF task as well. This ensured that nobody performed two adjacent tasks with the same stimuli. The SL and PF tasks themselves were identical to those in Experiment 1, with the same attention checks, instructions, and questions, and so forth.

7. Experiment 3: Results

The purpose of this experiment was to further explore whether stimulus familiarity or complexity drove the effects in the previous experiment. If, as hypothesized, it is due to familiarity that yields several (pre-registered) predictions. First, it suggests that we should observe an interaction between speaker and stimulus on SL tasks, such that Chinese speakers will perform better on the CHINESE stimuli than in any other case. Second, it suggests that we should see the same interaction on the PF task. And third, it suggests that PF should be correlated with SL, especially on the same stimuli. We consider each prediction in turn.

7.1. Statistical learning

As in the previous experiments, accuracy on the SL task (SLscore) covered the range of individual differences, from 19.0% to 100% (mean: 48.9%). Performance was significantly above chance ($t(353) = 11.8$, $p < .0001$, $d = 3.43$) and there was again a high correlation between the item-level accuracy of the participants of Siegelman et al. (2017) and ours on both the first SL task ($r = .64$, $p < .0001$) and the second ($r = .75$, $p < .0001$).

If familiarity rather than complexity drives the effects we have seen so far, we should expect that there should be an improvement in SL performance only for those participants who saw familiar stimuli: the Chinese speakers shown CHINESE characters. That is, we should expect an interaction between participant and stimulus. Conversely, if stimulus complexity were driving the effect, we should expect everyone to perform equally on both stimulus sets.

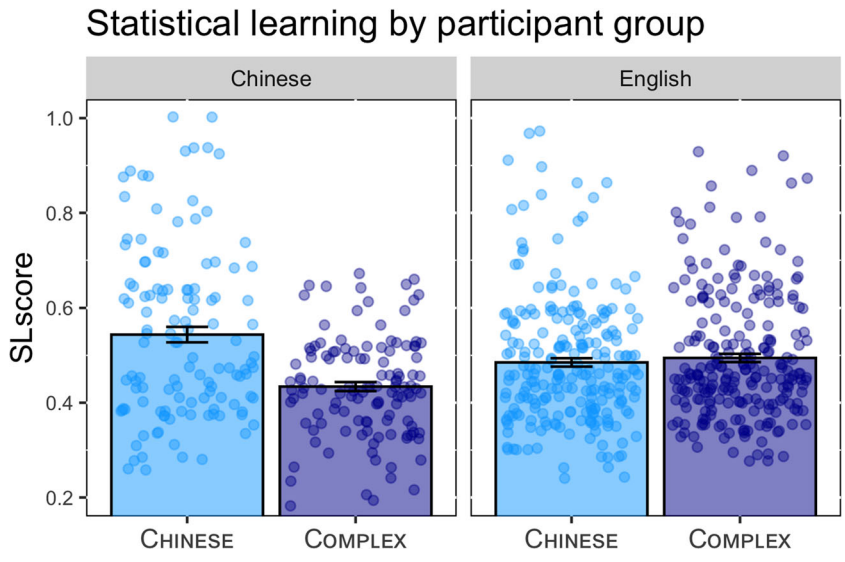


Fig. 10. Experiment 3 statistical learning. As predicted by the familiarity hypothesis, there was an interaction between participant and stimulus type: Chinese speakers (left panel) showed higher statistical learning performance on the CHINESE stimuli than did English speakers.

The results, as shown in Fig. 10, support the familiarity hypothesis: Chinese speakers show improved performance on CHINESE stimuli, and the interaction is significant (stimulus: $F(1, 704) = 9.1698$, $p = .0026$; participant: $F(1, 704) = 0.0062$, $p = .9375$; interaction: $F(1, 704) = 29.4098$, $p < .0001$). Post-hoc t-tests reveal that, as expected, English speakers performed similarly on the CHINESE and COMPLEX stimuli ($t(461.9) = 0.758$, $p = .449$), while Chinese speakers performed significantly better on the CHINESE stimuli than they did on the COMPLEX stimuli ($t(195.1) = -5.804$, $p < .0001$). These results indicate that stimulus familiarity, not absolute complexity, drives these individual differences in SL.

7.2. *Perceptual fluency*

By a similar logic as above, if familiarity underlies PF, we should expect an interaction between participant and stimuli on PFscore as well. As Fig. 11 shows, this is indeed what we found: Chinese speakers performed better (with a lower PFscore) on the CHINESE stimuli than anybody did on any of the other stimuli (stimulus: $F(1, 698) = 30.885$, $p < .0001$; participant: $F(1, 698) = 30.066$, $p < .0001$; interaction: $F(1, 698) = 38.7714$, $p < .0001$). Post-hoc t-tests reveal that, as in the SL task, English speakers performed similarly on the CHINESE and COMPLEX stimuli ($t(447.5) = 0.804$, $p = .422$), while Chinese speakers performed significantly better on the CHINESE stimuli than they did on the COMPLEX stimuli ($t(232.25) = 8.815$, $p < .0001$).

We can also investigate whether PF is associated with SL. If familiarity has something to do with this, one would expect that the correlation between PFscore and SLscore would be higher when the stimuli were the same (both CHINESE or both COMPLEX) than when they were

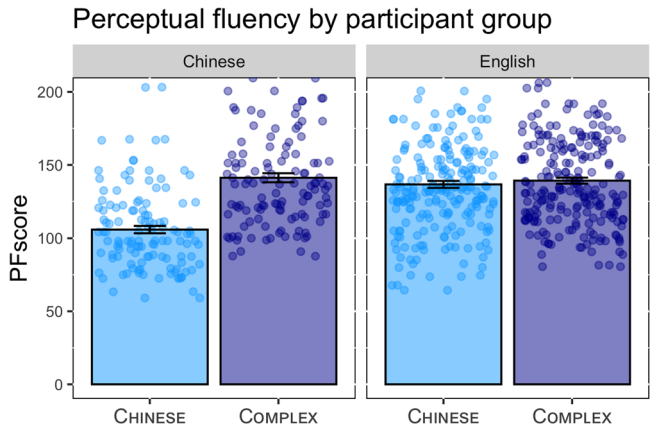


Fig. 11. Experiment 3 perceptual fluency. As predicted by the familiarity hypothesis, there was an interaction between participant and stimulus type: although all participants had poor perceptual fluency on the COMPLEX stimuli, the Chinese speakers (left panel) had better perceptual fluency (as reflected by a lower PFscore) on the CHINESE stimuli than did English speakers.

different (i.e., PF for CHINESE and SL for COMPLEX or PF for COMPLEX and SL for CHINESE). This was indeed the case (same: $r = -0.2$, $p < .0001$; different: $r = -.04$, $p = .45$). The difference between these correlations was significant ($z = -2.1$, $p = .036$).

8. General discussion

This work investigated the degree to which PF on specific stimuli influences SL involving those stimuli, and to what extent PF is driven by stimulus familiarity, complexity, or both. Recognizing that SL relies on basic memory processes (Frank et al., 2010; Frost et al., 2019; Isbilen et al., 2020), we hypothesized that participants would learn the same statistical regularities better when they found items to be easier to perceptually distinguish, parse, and remember. This hypothesis was uniformly supported across three experiments. In Experiment 1, participants were significantly better at learning statistical regularities when the stimuli involved were simple and familiar than when they were complex and novel, even though the modality was the same in both cases. This suggests that the ease with which statistical distributions are learned is moderated not only by a person's modality-specific level of perceptual fluency, but also by how easily they can process and remember specific *items*. These effects were also observable at the individual level: test–retest correlations were higher when the two tasks used items that were similarly complex and familiar, even though none of the specific individual items were repeated. Additionally, individual performance on both SL tasks was significantly associated with performance on a novel independent measure of PF. In Experiment 2, we found that SL performance was not improved when the stimuli were simpler but still unfamiliar. Finally, in Experiment 3, we held stimulus complexity constant and varied

participant experience with the experimental material, and once again found that participants performed best on the stimuli with which they were most familiar.

Overall, these data identify a fundamental role for perceptual fluency in statistical learning and demonstrate that perceptual fluency not only varies significantly between individuals but is also strongly driven by familiarity. Our findings are consistent with research showing that familiarity matters more than complexity in working memory and processing efficiency (Jackson & Raymond, 2008; Ngiam et al., 2019; Reder et al., 2016; Xie & Zhang, 2017; Zhang et al., 2020). Our results may also help elucidate why frequency of exposure—leading to high familiarity—matters in so many aspects of cognition, from language to decision-making to memory (for example, Hasher & Zacks, 1984; Sedlmeier & Betsch, 2002; Gries & Divjak, 2012; Baayen, Milin, & Ramscar, 2016). Given the ubiquity of frequency effects, it is perhaps not particularly surprising if SL is also improved when the items involved are highly familiar, having occurred frequently in prior experience. Indeed, this finding can also explain why people with specific expertise or training in a domain show better SL in that domain (Mandikal Vasuki, Sharma, Ibrahim, & Arciuli, 2017; Martire, Growns, & Navarro, 2018).

Our work thus suggests an important modification to existing theories of SL (for example, Frost et al., 2015). Rather than postulating that individual variation in the efficiency of perceptual processing and memory is entirely based on modality or timing, our results indicate that familiarity with the specific items involved plays an important moderating role. We do not suggest that there are no modality-specific individual differences in PF, but we *do* suggest that familiarity may be far more important than has previously been recognized; indeed, it is possible that some of the variation that has historically been attributed to modality differences may reflect familiarity differences instead.

Whether familiarity would have the exact same effect on SL in the auditory modality remains an open question, as to our knowledge, comparable studies to ours do not exist. However, our results are consistent with evidence that Mandarin learners have better statistical learning for artificial tonal stimuli than English speakers (Potter et al., 2017). Moreover, high frequency and therefore more familiar items have been found to play facilitative roles in speech segmentation in infants (Bortfield, Morgan, Golinkoff, & Rathbun, 2005; Monaghan & Christiansen, 2010) and adults (Frost, Monaghan, & Christiansen, 2016). Other research has shown an important role for high-frequency marker words in artificial grammar learning (for example, Valian & Coulson, 1988). In these studies, the high-frequency familiar items serve as cues to higher level structure, and their designs are interesting because they mimic asymmetrical (e.g., Zipfian) frequency distributions present in language. An auditory analog of our current study would show something even more fundamental, suggesting that familiarity with a lower level feature of language (e.g., syllables or phonemes) might provide an important basis for the statistical learning of higher level forms (e.g., words).

We have interpreted our results as reflecting the importance of familiarity, but one might question to what extent they reflect verbalisability instead: perhaps, people performed better for familiar stimuli not because they were familiar *per se*, but because they had names and were thus easier to verbalize. This possibility is consistent with research suggesting that SL is improved for stimuli that are easy to encode verbally (Conway, Karpicke, & Pisoni, 2007). It is probable that verbalisability does play some role in our results, and it is impossible to rule

it out entirely: any stimulus that is familiar enough to improve SL will also almost certainly become associated with a label or mental shorthand. That said, several considerations suggest that our results are not *entirely* due to an improved ability to verbalize familiar stimuli. First, verbaliseability cannot explain the significant correlations found between performance on the PF and SL tasks. The PF task was a stimulus-specific task that had no component of SL at all, and occurred too rapidly (on the order of 120–130 ms) for verbalization to be a factor in the performance. For similar reasons, verbaliseability cannot explain the improved PFscore Chinese people had for CHINESE stimuli. Second, verbaliseability cannot explain why in Experiment 3 the Chinese participants actually performed *worse* in the SL task on the COMPLEX stimuli than the English participants did ($t(298.7) = -4.7$, $p < .0001$); the COMPLEX should be equally unverbaliseable for both populations. This result might instead indicate a form of perceptual interference, with COMPLEX stimuli being harder for Chinese people to encode because of their similarity to the highly familiar Chinese characters. This interpretation is speculative, but regardless, it does suggest that at least part of the performance differential in the SL task reflects familiarity rather than verbaliseability. Additionally, even if part of the effect is due to verbaliseability, this does not change our main conclusion that stimulus-specific PF affects SL; it just means that PF is enhanced for verbaliseable stimuli.

Our work is complementary to entrenchment accounts of SL, which suggest that the prior experience of TPs can interfere with the learning of novel TPs over the same items (Endress & Mehler, 2009; Perruchet & Poulin-Charronnat, 2012; Siegelman et al., 2018). The results are an interesting contrast to these cases, because our focus has been on the experience with the *stimuli themselves* (rather than the TPs) and because (other than the Chinese participants with COMPLEX stimuli) we found a facilitative rather than disadvantageous effect of prior experience. The facilitative effect makes sense given our focus on the stimuli themselves, since the relevant prior experience did not contradict what participants saw in experiment (whereas in previous research, the TPs in the experiment contradicted the learned TPs in the language). In our work, people's previous experience with letters and characters may have made those stimuli easier to rapidly encode, thus allowing more processing resources to be devoted to the SL component of the task. The effect of stimulus familiarity seems to have been substantial enough to override any interference caused by any pre-existing TPs (letter–letter associations) that participants may have had. Although our stimuli were controlled such that no participant saw the exact same combinations of items, literate participants are likely to have implicit expectations of relative letter locations, which could have but did not influence pattern detection.

Overall, our results support the suggestion that, rather than being a purely abstract and domain-general process, SL is a multicomponential process grounded in basic cognition (Arciuli, 2017; Frost et al., 2019; Growns, Siegelman, & Martire, 2020; Siegelman et al., 2017). Indeed, we found that within-domain variation in PF (both within and between individuals) plays a significant role in learning. These data are important, in the first instance, because they broaden the scope of variables that influence SL. Frost et al. (2019) argued that while SL is likely implicated in most domains of cognition, research on the topic has pursued a more isolationist strategy, and as such its promise as an explanatory concept has not been realized. Our finding that PF influences SL naturally lends itself to Frost et al. (2019)'s

call for domain integration because it grounds learning in basic memory and perceptual processes.

A natural question, which we cannot fully answer here, is how PF and SL interact. On the one hand, it is clear that stimulus encoding is a pre-condition for learning probability distributions across items (Bogaerts et al., 2016); however, if SL plays a fundamental role in learning, then the relationship should be bidirectional, such that perceptual features may be sharpened by predictable patterns over time. By way of example, consider how the two might interact in reading acquisition. Once a child learns individual letters, they begin to coordinate those letters into words, which in languages like English is difficult because of the one-to-many grapheme-to-phoneme mappings. Initially, this process is labored for children, whose strategies vacillate between whole-form memorization and phoneme-by-phoneme pronunciation. However, an account where SL is important for reading (Arciuli, 2018) predicts that, as the learner identifies regularities in the visual input, orthographic representations will be strengthened, thus completing a virtuous circle.

This work suggests part of a solution to the question of exactly what SL *is* and how it relates to other cognitive skills. If SL is influenced by PF, and that, in turn, is affected by familiarity and prior experience, then we should expect a degree of modality-specific effects on learning, which is consistent with the bulk of past research. As such, we take a further step away from the conceptualization of performance on SL tasks as measuring domain-general computations over amodal abstract representations. These findings also suggest that one way to improve SL (and any learning that relies on it) is simply to increase exposure to the stimuli involved. We are particularly excited for the potential of this possibility, since statistical learning is important in so many different domains and increased exposure is in many cases a relatively easy and cheap intervention. Much remains to be done, but our work opens the door to a variety of advances, both theoretical and applied.

Acknowledgments

Research costs for AP were funded through ARC grants DP150103280 and DP180103600. We would also like to thank Jing Qian and Rachel Yam for their intellectual contributions to the experimental design, Rebecca Frost for helpful feedback, and Toby Elmhirst for contributing his Mathematica expertise.

Open access publishing facilitated by The University of Melbourne.

Open Research Badges


This article has earned Open Data and Open Materials badges. Data are available at https://github.com/perfors/vsl and materials are available at https://github.com/perfors/vsl.

Notes

1. https://github.com/perfors/vsl/ contains stimuli, experiment code, data, and analyses for all experiments.
2. We removed a single outlier that was more than 4SD from the mean of PFscore. Results are qualitatively the same if no outliers or outliers at 2SD or 3SD are removed. We opted to remove at 4SD because inspection of the data suggested that this particular data point appeared to be a high-leverage outlier with a very long relative latency.
3. The pre-registration can be found at http://aspredicted.org/blind.php?x=r76fb4.
4. We pre-registered running 160 participants, expecting more difficulty in recruiting than we had, and were encouraged to run more than 160 in order to ensure that all students were able to earn sufficient course credit. All analyses reported here were performed only once, on the full dataset, but in order to ensure that our results were not due to a larger sample size than anticipated, we also ran all analyses on the first 160 people in the dataset, with identical results; see https://github.com/perfors/vsl/ for details of these analyses.
5. We followed this classification because it is what we pre-registered, but upon examination of the dataset, we saw that numerous participants did not rate themselves as native speakers but still got three or more correct on the quiz. We therefore replicated all analyses with a classification that put those participants in the Chinese group. The results, which can be found in the Supporting Information, are qualitatively the same to those reported here.
6. As pre-registered and as in Experiment 1, we removed the six PFscore values that were more than 4SD from the mean. Including them does not change any of our qualitative results.

References

Alvarez, G., & Cavanagh, P. (2004). Capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B*, *372*(1711), 1–9.

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*, 634–643.

Arciuli, J., & Simpson, I. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, *14*, 464–473.

Baayen, H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174–1220.

Bethell-Fox, C., & Shepard, R. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 12–23.

Bogaerts, L., Siegelman, N., & Frost, R. (2016). Splitting the variance of statistical learning performance: A parametric investigation of exposure duration and transitional probabilities. *Psychonomic Bulletin and Review*, *23*, 1250–1256.

Bonin, P., Peereman, R., Malardier, N., Meot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 158–167.

Bortfield, H., Morgan, J., Golinkoff, R., & Rathbun, K. (2005). Mommy and me: Familiar names help babies into speech segmentation. *Psychological Science*, *16*, 298–304.

Brady, T., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*(7), 678–685.

Buchsbaum, D., Griffiths, T., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology*, *76*, 30–77.

Bulf, H., Johnson, S., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132.

Christiansen, M. (2019). Implicit statistical learning: A tale of two theories. *Topics in Cognitive Sciences*, *11*, 468–481.

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

Conway, C., Karpicke, J., & Pisoni, D. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of Deaf Studies and Deaf Education*, *12*(3), 317–334.

Cycowicz, Y., Friedman, D., Rothstein, M., & Snodgrass, J. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*(2), 171–237.

Donderi, D. (2006). An information theory analysis of visual complexity and dissimilarity. *Perception*, *35*(6), 823–835.

Endress, A., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*, 351–367.

Eng, H., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin and Review*, *12*, 1127–1133.

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

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in human speech segmentation. *Cognition*, *117*, 107–125.

Frost, R., Armstrong, B., & Christiansen, M. (2019). Statistical learning research: A critical review and possible directions. *Psychological Bulletin*, *145*, 1128–1153.

Frost, R., Armstrong, B., Siegelman, N., & Christiansen, M. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125.

Frost, R., Monaghan, P., & Christiansen, M. (2016). Using statistics to learn words and grammatical categories: How high frequency words assist language acquisition. In A. Papafragou, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 81–86). Austin, TX: Cognitive Science Society.

Gebhart, A., Newport, E., & Aslin, R. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin and Review*, *16*, 486–490.

Gries, S., & Divjak, D. (Eds.). (2012). *Frequency effects in language learning and processing*. Berlin/Boston: De Gruyter Mouton.

Growns, B., Siegelman, N., & Martire, K. (2020). The multi-faceted nature of visual statistical learning: Individual differences in learning conditional and distributional regularities across time and space. *Psychonomic Bulletin & Review*, *27*, 1291–1299.

Hasher, L., & Zacks, R. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372–1388.

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

Isbilen, E., McCauley, S., Kidd, E., & Christiansen, M. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.),

Proceedings of the 39th Annual Conference of the Cognitive Science Society (pp. 564–569). Austin, TX: Cognitive Science Society.

Isbilen, E., McCauley, S., Kidd, E., & Christiansen, M. (2020). Statistically-induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*, e12848.

Jackson, M., & Raymond, J. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 556–568.

Kaufman, S., DeYoung, C., Gray, J., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Kemps, E. (1999). Effects of complexity on visuo-spatial working memory. *European Journal of Cognitive Psychology*, *11*(3), 335–356.

Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.

Krogh, L., Vlach, H., & Johnson, S. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, *3*, 598.

Liu, T., Chen, W., Liu, C., & Fua, X. (2012). Benefits and costs of uniqueness in multiple object tracking: The role of object complexity. *Vision Research*, *66*, 31–38.

Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience*, *22*, 496–512.

Madan, C., Bayer, J., Gamer, M., Lonsdorf, T., & Sommer, T. (2018). Visual complexity and affect: Ratings reflect more than meets the eye. *Frontiers in Psychology*, *8*, 2368.

Mandikal Vasuki, P., Sharma, M., Ibrahim, R., & Arciuli, J. (2017). Statistical learning and auditory processing in children with music training: An ERP study. *Clinical Neurophysiology*, *128*(7), 1270–1281.

Martire, K., Growns, B., & Navarro, D. (2018). What do the experts know? Calibration, precision, and the wisdom of crowds among forensic handwriting experts. *Psychonomic Bulletin and Review*, *25*(6), 2346–2355.

Miton, H., & Morin, O. (2021). Graphic complexity in writing systems. *Cognition*, *214*, 104771.

Monaghan, P., & Christiansen, M. (2010). Words in puddles of sound: Modelling psycholinguistic effects in language acquisition. *Journal of Child Language*, *37*, 545–564.

Ngiam, W., Khaw, K., & Goodbourn, P. (2019). Visual working memory for letters varies with familiarity but not complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1761–1775.

O'Connor, T., & Burns, N. (2003). Inspection time and general speed of processing. *Personality and Individual Differences*, *35*(3), 713–724.

Pelli, D., Burns, C., Farell, B., & Moore-Page, D. (2006). Feature detection and letter identification. *Vision Research*, *46*, 4646–4674.

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language*, *66*(4), 807–818.

Popov, V., & Reder, L. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review*, *127*(1), 1–46.

Potter, C., Wang, T., & Saffran, J. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive Science*, *41*, 913–927.

Reder, L., Liu, X., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin and Review*, *23*(1), 271–277.

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

Sedlmeier, P., & Betsch, T. (Eds.). (2002). *Etc: Frequency processing and cognition*. New York, NY: Oxford University Press.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213.

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.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120.
Sun, H., Zimmer, H., & Fu, X. (2011). The influence of expertise and of physical complexity on visual short-term memory consolidation. *Quarterly Journal of Experimental Psychology*, *64*(4), 707–729.
Sun, Z., & Firestone, C. (2021). Curious objects: How visual complexity guides attention and engagement. *Cognitive Science*, *45*, e12933.
Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 127–132.
Turk-Browne, N., Junge, J., & Scholl, B. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564.
Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, *27*, 71–86.
Wang, T., & Saffran, J. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*, 953.
Watson, A. (2012). Perimetric complexity of binary digital images: Notes on calculation and relation to visual complexity. *Vision Research*, *14*, 1–41.
Xie, W., & Zhang, W. (2017). Familiarity increases the number of remembered Pokemon in visual short-term memory. *Memory and Cognition*, *45*(4), 677–689.
Zhang, J., Liu, X., So, M., & Reder, L. (2020). Familiarity acts as a reduction in objective complexity. *Memory and Cognition*, *48*, 1376–1387.

Appendix A: Further analyses for Experiments 1 and 2

Figs. A1–A4.

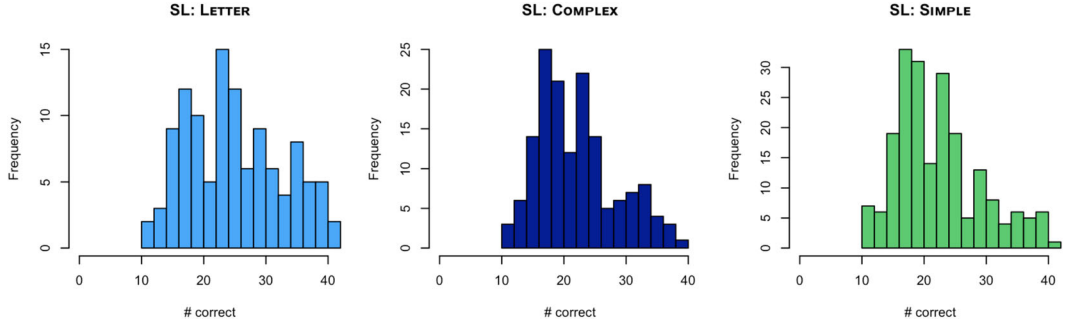


Fig. A1. Histograms of accuracy on the SL tasks. Each histogram plots the individual overall SLscore for each participant, broken down by condition. It is evident that there is a wide range of individual variation within each condition, and that the histograms appear approximately normal and similar in character to those reported in Siegelman et al. (2017).

Appendix B: Stimuli for Experiment 3

Fig. B1

Appendix C: Additional analyses for Experiment 3

Figs. C1–C3

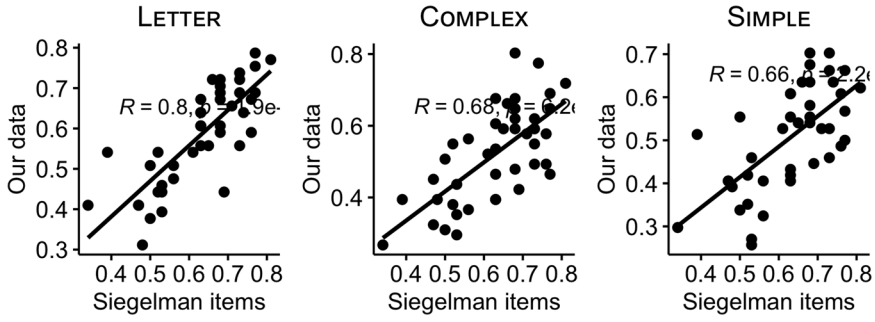


Fig. A2. Item-level correlations with data from Siegelman et al. (2017). Each plot shows the correlation of performance on each of the 42 test items in the first SL task in our study and that of Siegelman et al. (2017). Correlations are strong and highly significant whether calculated over the LETTER stimuli (left), the COMPLEX stimuli (middle), or the SIMPLE ones. This demonstrates that our participants were approaching the task similarly to theirs.

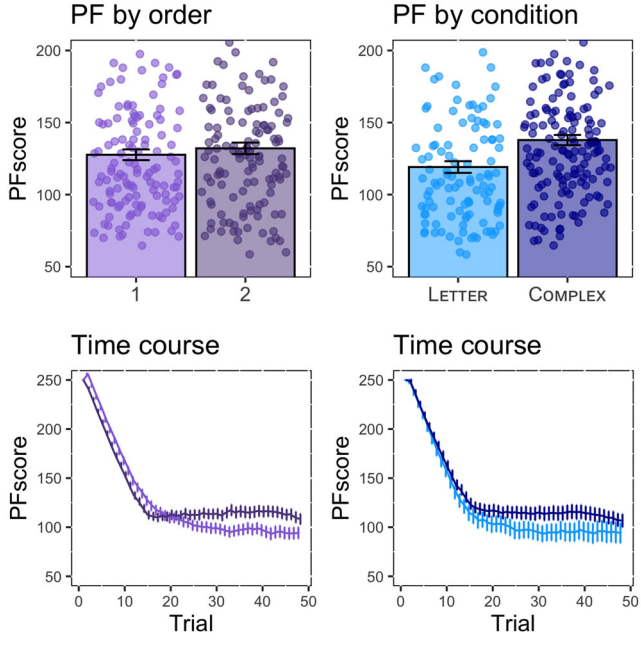


Fig. A3. Performance in perceptual fluency task. The panels on the left show PFscore as a function of whether the PFtask was the first or the second the participant did. The time course (bottom) reveals that there is a small effect of both learning (improved performance initially on the second task) and fatigue (improved performance at the end of the first task). However, as the bar graph at the top reveals, these effects cancel each other out and there is no significant difference by order of the task. Conversely, the panels on the right show that there is consistently better performance with the LETTER stimuli than with the COMPLEX stimuli.

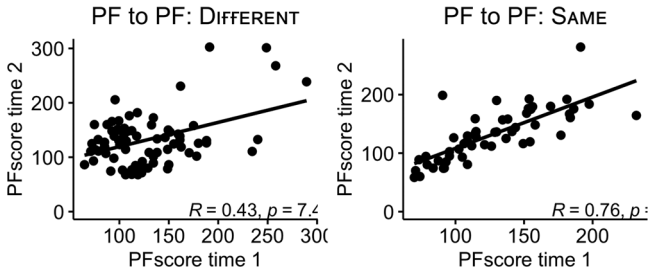


Fig. A4. Correlations between PFscores. Each plot shows the correlation of performance on the two PF tasks each person took, broken down by whether the stimuli were from the same set (both COMPLEX or both LETTER). Correlations are significant for both, but higher when the stimuli were from the same set. This suggests that the PF task reflects aspects of both stimulus-specific encoding efficiency as well as more general speed of encoding or processing.

1. 他们
   (a) metal door *(b) they* (c) government (d) her

2. 球
   (a) spread (b) beg (c) play *(d) ball*

3. 没有
   (a) yes (b) present *(c) not* (d) sister

4. 红
   (a) red (b) blue (c) green (d) yellow

Character	Pronunciation	Meaning
日	rì	day/ date
夫	fū	husband
元	yuán	dollar
风	fēng	wind
头	tóu	head
半	bàn	half
皮	pí	skin
西	xī	East
色	sè	colour
早	zǎo	morning/ early
肉	ròu	meat/ flesh
声	shēng	sound
壳	ké	shell

Fig. B1. Stimuli used in Experiment 3. On the left are the four test questions used to evaluate Chinese proficiency; correct answers are italicized. The table on the right shows all 16 Chinese characters along with their pronunciation and meaning. All corresponded to nouns and varied widely in initial phoneme and tone.

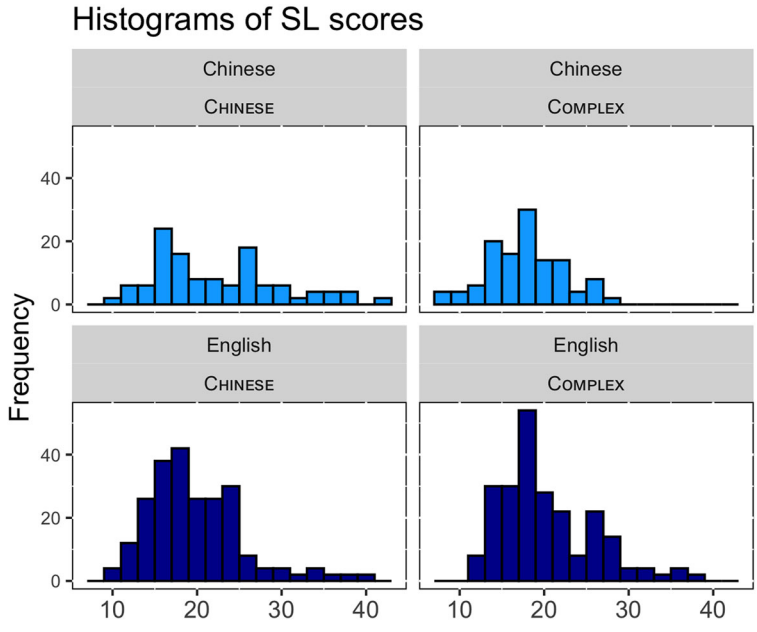


Fig. C1. Histograms of accuracy on the SL tasks. Each histogram plots the individual overall SLscore for each participant, broken down by condition. As before, there is a wide range of individual variation within each condition, and the histograms appear approximately normal and similar in character to those reported in Siegelman et al. (2017).

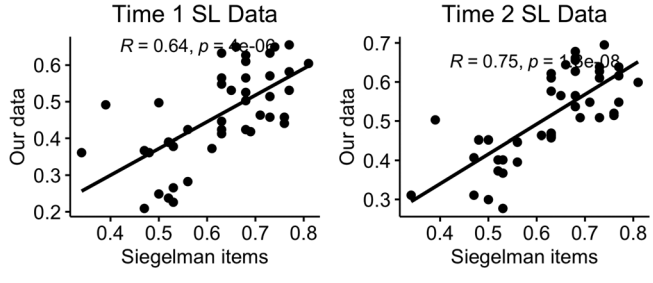


Fig. C2. Item-level correlations with data from Siegelman et al. (2017). Each plot shows the correlation of performance on each of the 42 test items in the SL task in our study and that of Siegelman et al. (2017), broken down by whether it was the first or the second task. As before, correlations are strong and highly significant, which this demonstrates that our participants were approaching the task similarly to theirs.

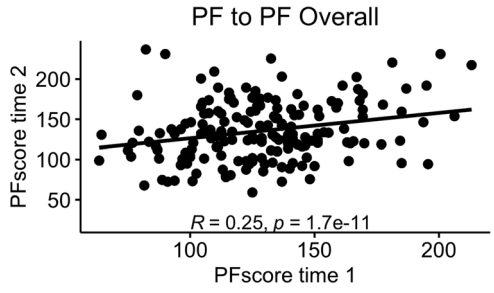


Fig. C3. Correlations between PF scores. Correlation of performance on the two PF tasks each person took. As before, the correlation is positive and significant.