What drives individual differences in statistical learning? The role of perceptual fluency and familiarity

Amy Perfors
School of Psychological Sciences
University of Melbourne
ARC Centre of Excellence for the Dynamics of Language

Evan Kidd
Max Planck Institute for Psycholinguistics
Research School of Psychology
The Australian National University
ARC Centre of Excellence for the Dynamics of Language

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 variation in perceptual fluency — the ability to rapidly or efficiently code and remember the stimuli that statistical learning occurs over. We show 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, we find that test-retest correlations of performance in a statistical learning task using stimuli of the same level of familiarity (but distinct items) are stronger than correlations across the same task with different levels of familiarity. Finally, we demonstrate that statistical learning performance is predicted by an independent measure of stimulus-specific perceptual fluency which contains no statistical learning component at all. Our results suggest that a key component of SL performance may be unrelated to either domain-specific statistical learning skills or modality-specific perceptual processing.

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 (Teinonen, Fellman, Näätänen, Alku, &

Huotilainen, 2009; Bulf, Johnson, & Valenza, 2011). It occurs in both auditory and visual modalities and across a wide variety of types of stimuli (Saffran, Aslin, & Newport, 1996; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Brady & Oliva, 2008; Gebhart, Newport, & Aslin, 2009; Krogh, Vlach, & Johnson, 2013; Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015). Moreover, recent evidence suggests that individual differences in statistical learning abilities appear to be stable and reliable (Isbilen, McCauley, Kidd, & Christiansen, 2017; Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018). Taken together, these findings appear to indicate that statistical learning is a broad, domain-general ability for extracting and using the regularities in the world.

Consistent with this idea, statistical learning abilities do appear to be tied to real-world skills, especially in the domain of language: there are significant associations between performance on SL tasks and first and second language acquisition (Kaufman et al., 2010; Romberg & Saffran, 2010; Shafto, Conway, Field, & Houston, 2012; Kidd & Arciuli, 2016; Hamrick, Lum, & Ullman, 2018), adult language processing and skill (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak & Christiansen, 2012; Daltrozzo et al., 2017), and literacy (Arciuli & Simpson, 2012; Frost, Siegelman, Narkiss, & Afek, 2013). Intriguingly, the relationship between language and SL has been demonstrated even when SL is measured in both the verbal and visual modality. This is often taken to suggest that statistical learning and language build on shared skills and computational principles – such as the ability to attend to the statistical relationships (transitional probabilities, or TPs) between items and/or the capacity to use those as cues to structure.

The evidence thus far paints a picture of statistical learning as a broad-based, domain-general cognitive learning ability. However, some puzzles remain. One is that there is no apparent link between statistical learning abilities and working memory or IQ (Reber, Walkenfeld, & Hernstadt, 1991; Kaufman et al., 2010; Siegelman & Frost, 2015; Kidd & Arciuli, 2016). Given the centrality of IQ and working memory to cognition, this is somewhat disconcerting. Perhaps even more confusing, multiple studies have found no relationship between SL performance in different modalities: that is, high accuracy in a visual SL task does not predict higher accuracy in an auditory SL task, even if the underlying statistical regularities are exactly the same (Conway & Christiansen, 2006; Siegelman & Frost, 2015; Erickson, Kaschak, Thiessen, & Berry, 2016). Moreover, preliminary evidence suggests that statistical learning shows different developmental trajectories in different modalities (Raviv & Arnon, 2018). If SL is truly domain-general, why is this?

Frost, Armstrong, Siegelman, and Christiansen (2015) suggested that the reason for this is that variance in performance on SL tasks comes from *two* main sources – not just (i) variability in the efficiency in which domain-general learning mechanisms detect the statistical properties of the input stream, but also (ii) variability in the efficiency of perceptual encoding within each modality. Under this theory, the relationship between SL and language arises because both draw on the same domain-general learning mechanisms, but SL tasks are uncorrelated across modalities because of differences in perceptual efficiency within the visual and auditory domains. One test of the theory came from presenting participants with nine different visual SL tasks that factorially manipulated speed of presentation (200ms, 600ms, or 1000s) and probabilistic structure (within-pair TPs were 0.6, 0.8, or 1.0) (Bogaerts, Siegelman, & Frost, 2016). As predicted, both variables affected learning, with higher performance observed at slower exposure rates and for high TPs. Participants who were faster to encode the stimuli were also better at learning the underlying statistical structure. This suggests that an individual's *perceptual fluency* may afford a significant advantage in statistical learning, and is consistent with the observation that at least one measure of statistical learning has been found to be associated with psychometric measures of processing speed (Kaufman et al., 2010).

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

However, there is yet another, less well-studied possible factor : stimulus familiarity. Because familiar items are easier to quickly encode and accurately remember (Jackson & Raymond, 2008; Xie & Zhang, 2017), this might free up cognitive capacity to devote to other tasks (like statistical learning). As a result, one might expect SL performance to be higher when the items involved are familiar rather than unfamiliar, even if the underlying statistical regularities are the same in both cases. Moreover, to the extent that perceptual fluency reflects familiarity rather than modality, one might expect that correlations between SL tasks would be affected by the perceptual similarity of the stimuli even when the modality is held constant. Finally, if statistical learning is affected by perceptual fluency, one might expect SL performance to be predicted by a performance on a purely perceptual stimulus-specific speeded memory task. We test all of these predictions in this work.

It is worth noting that although our hypothesis is that individual differences in statistical learning are moderated by perceptual fluency – which is shaped by familiarity – our work is distinct from other research that also suggests that prior experience plays a role in statistical learning of linguistic information (e.g., Gebhart et al., 2009; Endress & Mehler, 2009; Perruchet & Poulin-Charronnat, 2012; Siegelman et al., 2018). Our focus 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 statistical learning in any modality, not interfere with statistical learning in the linguistic one. 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 affect SL due to interference and entrenchment, and the focus is not on the items so much as the transition probabilities between them. Our approach complements this other research because both focus on the role of prior experience in shaping statistical learning; it is distinct because it has a different mechanism and makes different predictions.

Experiment 1: Method

This research tests the hypothesis that individual differences in statistical learning are substantially mediated by individual differences in perceptual fluency, which is itself shaped by stimulus familiarity. We investigated this by having the same individuals participate in two different visual SL tasks on two different days, as shown in Figure 1. On each day people were randomly assigned to one of two conditions (FAMILIAR and UNFAMILIAR) in which the SL task was the same and the only difference was whether the stimuli involved were novel to the participant or not. Thus, some people saw stimuli with the same level of novelty on each day (although distinct items each time) while others saw the FAMILIAR stimuli on one day and the UNFAMILIAR ones on another. If perceptual fluency plays an important mediating role, we would expect both that SL performance is higher when the stimuli are FAMILIAR, and that there would be a higher test-retest correlation across sessions when the stimulus novelty is the same.

In addition to the statistical learning tasks that occurred during both experimental sessions, Session 2 also presented participants with two measures of perceptual fluency (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 was designed to measure the speed with which these particular stimuli could be parsed and encoded. If perceptual fluency plays an important role in statistical learning, we would expect that PF task performance would be correlated with SL task performance.

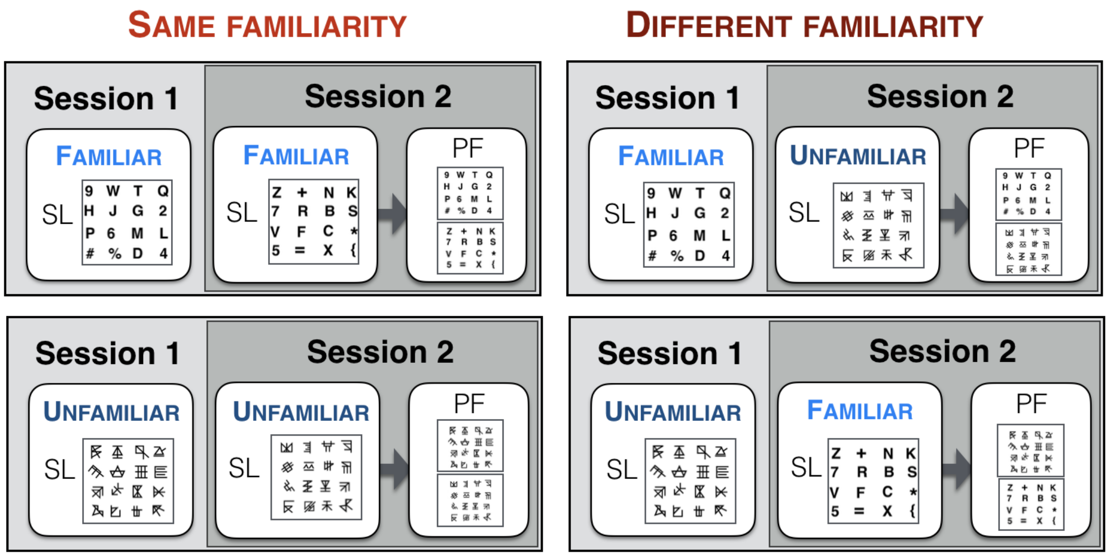

Figure 1. Overview of experiment structure. Each participant was tested in two sessions approximately one week apart. In the first session, people performed an embedded triplet visual statistical learning (SL) task using either FAMILIAR or UNFAMILIAR stimuli. In the second session, the same participants completed another SL task with new stimuli that either matched or did not match in familiarity (although in all cases the specific stimulus items were different). Following the SL task, each person performed two consecutive tasks that measured their perceptual fluency (PF) on the two stimulus sets they had seen overall.

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. 135 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 FAMILIAR or UNFAMILIAR stimulus sets of 16 items each. This resulted in 53 people who saw the SAME stimulus complexity each time (17 FAMILIAR, 36 UNFAMILIAR) and 79 people who saw a DIFFERENT stimulus complexity each time (44 saw FAMILIAR first, 35 saw UNFAMILIAR first).

Materials

As shown in Figure 2, the UNFAMILIAR 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 FAMILIAR stimuli consisted of letters and symbols. We chose these because they are highly over-learned and very familiar to any literate English speaker. In order to minimise 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.

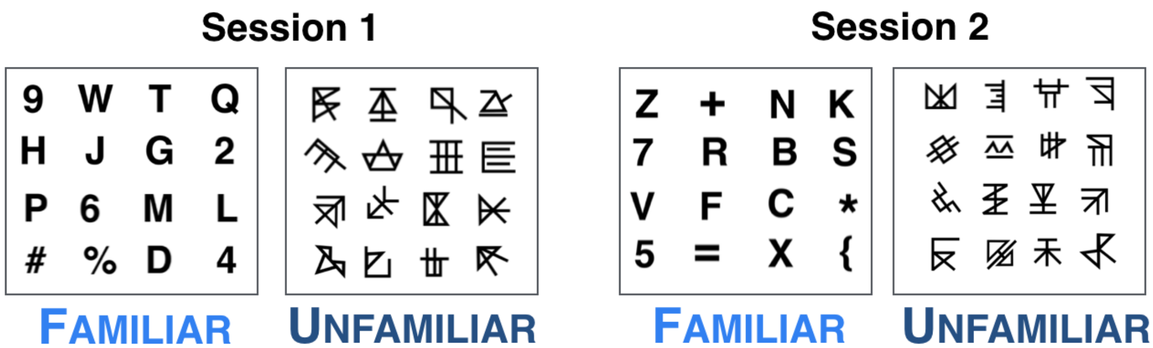

Figure 2. *Experiment 1 stimuli.* At each session, participants saw 16 stimulus items that were either FAMILIAR or UNFAMILIAR, depending on randomly-assigned condition. Although some people saw stimuli of similar familiarity in both sessions, no single item occurred both times.

Procedure

Statistical Learning (SL) task. The statistical learning task in both sessions was nearly identical to Siegelman, Bogaerts, and Frost (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, 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 amongst 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 statistical learning task itself consists of 16 items x combined into 8 triplets, each repeated 24 times in random order. Each item appears alone for 800ms with 200ms between items. The eight triplets were designed so that some are easy, defined by transition probabilities 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 some were harder, defined by transition probabilities 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 randomised for each participant at each session.

After the ten-minute training sequence, we administered the exact same 42-item test as in Siegelman et al. (2017). Test items varied in many ways amongst each other, from response required (picking the familiar pattern or filling a missing shape in a pattern) to TPs (whether the target had low TPs of 0.33 or high TPs of 1.0, as well as the TPs and number of position violations of the foils) to number of distractors (two to four) to whether they are on pairs or triples. As such, the items span a wide range of difficulty with the aim of capturing individual differences in performance.

For each participant we calculate an overall statistical learning score, SLscore, which reflects the proportion of the 42 test items that they got correct. In order to evaluate whether our participants found the same items to be difficult as 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 (FAMILIAR: $r = 0.8, p < 0.0001$; UNFAMILIAR: $r = 0.68, p < 0.0001$), suggesting that our participants were approaching the task similarly to theirs.

Perceptual Fluency (PF) task. The Perceptual Fluency (PF) task, illustrated in Figure 3, 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 (250ms on the first trial) followed by a masking stimulus (always presented for 200ms). If the participant successfully identifies the item, the target on the next trial flashes more quickly (with a duration of 15ms less); if they do not, it flashes slower (with a duration of 15ms 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 improved perceptual fluency.

Each of the 16 items in the relevant stimulus set was the target three times over the course of the 48 trials. Masks and distractors were selected from the same stimulus set and completely randomised for each person, as was the order of trials.

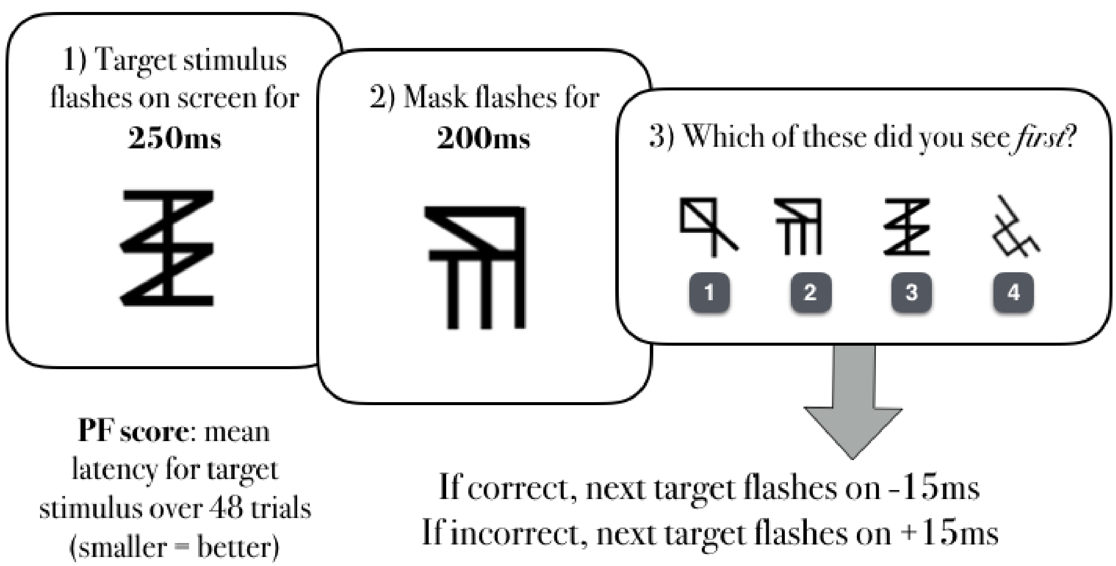

Figure 3. Perceptual fluency task. During each trial of the PF task, a target stimulus is flashed briefly onscreen, followed by a masking stimulus for 200ms. Participants are then presented with four items (the target, the mask, and two others randomly selected from the stimulus set). If they successfully identify the target, on the next trial the duration of target presentation decreases by 15ms; if not, it increases by 15ms. The perceptual fluency score (PFscore) reflects the mean target latency over all 48 trials. Smaller PFscore indicates better perceptual fluency, because it indicates that the person can attain similar accuracy based on a shorter target presentation time.

Experiment 1: Results

The overall hypothesis being tested in this paper is that perceptual fluency plays an important mediating role in performance on statistical learning tasks. This hypothesis yields several testable predictions. First, it predicts that people will show different levels of performance on statistical learning in tasks that differ only on familiarity, 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 similar familiarity than if they are not (even though all are within the same modality). Third, it predicts that statistical learning performance should be correlated with perceptual fluency, even though no statistical learning is involved in the perceptual fluency task at all. We consider each of these predictions in turn.

Is performance better on Familiar stimuli?

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

Statistical learning. Accuracy on the statistical learning tasks, as measured by SLscore, does appear to span the range of individual differences, ranging between 26.2% to 97.6% (mean: 59.6%) in the FAMILIAR condition and 28.6% to 92.9% (mean: 52.7%) in the UNFAMILIAR condition. In both conditions, performance is significantly above chance, which Siegelman et al. (2017) calculate as 40% (FAMILIAR: $t(112) = 11.45, p < 0.0001, d = 1.08$; UNFAMILIAR: $t(150) = 10.69, p < 0.0001, d = 0.87$). Moreover, the difference in accuracy between FAMILIAR and UNFAMILIAR conditions, shown in Figure 4(a), is highly significant ($t(211) = 3.24, p = 0.0014, d = 0.42$).

Perceptual fluency. As Figure 4(b) reveals, perceptual fluency was also significantly improved, as reflected in a lower PFscore, when the stimuli were FAMILIAR than when they were UNFAMILIAR ($t(240.1) = -3.46, p = 0.0006, d = 0.43$). 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.

Are SL correlations higher between stimuli of the same familiarity?

If perceptual fluency mediates performance on statistical learning 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 familiarity (FAMILIAR-FAMILIAR or UNFAMILIAR-UNFAMILIAR), even though no individual items are repeated. Conversely, we should expect lower correlations if they are different at different sessions (FAMILIAR-UNFAMILIAR or UNFAMILIAR-FAMILIAR), even though the modality is the same. We evaluate this prediction in Figure 5. It is clear that the test-retest correlation between SAME familiarity versions was higher ($r = 0.7, p < 0.0001$) than between DIFFERENT versions ($r = 0.43, p < 0.0001$). The difference between these correlations is itself significant using a Fisher r-to-z transformation ($z = 2.26, p = 0.024$).

Does perceptual fluency predict performance in the statistical learning task?

A final prediction of the hypothesis that perceptual fluency mediates statistical learning performance is that higher perceptual fluency (as measured by a lower PFscore) should be associated with higher statistical learning (as measured by a higher SLscore). To evaluate this, we removed a

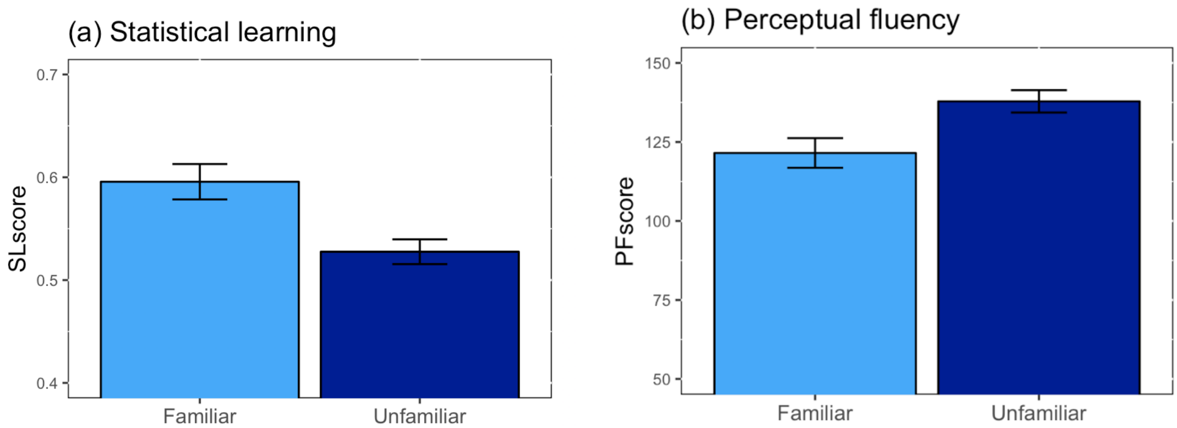

Figure 4. *Performance as a function of stimulus complexity.* (a) In the statistical learning task, accuracy (SLscore) was significantly higher for FAMILIAR stimuli than for UNFAMILIAR stimuli, though both were far above chance at approximately 40%. (b) In the perceptual fluency task, mean latency of the target stimulus (PFscore) was significantly lower for FAMILIAR stimuli, indicating that participants were better at quickly encoding and remembering the FAMILIAR stimuli than the UNFAMILIAR ones.

single outlier that was more than 4SD from the mean of PFscore.[1] The correlation between perceptual fluency and statistical learning accuracy within the same person on the same stimuli was indeed significant ($r = -0.27, p < 0.0001$). It remained significant even when restricted to only FAMILIAR stimuli ($r = -0.26, p = 0.005$) or only UNFAMILIAR stimuli ($r = -0.23, p = 0.005$).

Experiment 2: Method

The results thus far suggest that perceptual fluency influences statistical learning, and that people have higher perceptual fluency for familiar items. Statistical learning performance is higher when the items involved are overlearned and highly familiar; test-retest correlations between SL tasks are higher when the stimuli on both days are of the same level of familiarity; and performance on a test of perceptual fluency is both higher for familiar than unfamiliar items and correlated with statistical learning performance.

However, one open question remains: to what extent do our results reflect *familiarity*, and to what extent do they reflect *complexity*? Perhaps participants performed worse on the UNFAMILIAR stimuli not because they were novel, but because they were more visually complex. After all, each item contained on average more distinct lines as well as fewer other shapes than letters do; they thus may have required finer-grained parsing than the FAMILIAR stimuli. The possibility that complexity rather than familiarity is driving these effects is especially interesting in light of the fact that complexity is thought to affect working memory capacity (Alvarez & Cavanagh, 2004) and retrieval (Hofmeister, 2011).

[1]Results are qualitatively similar if no outliers are removed or if outliers at 2SD or 3SD are removed instead: the main correlation is significant at $p < 0.001$ regardless. We opted to remove them at 4SD because visual inspection of the scatterplots suggested that this particular data point appeared to be a high-leverage outliers with a very long relative latency, suggesting that fatigue or a lack of understanding of the task was driving it.

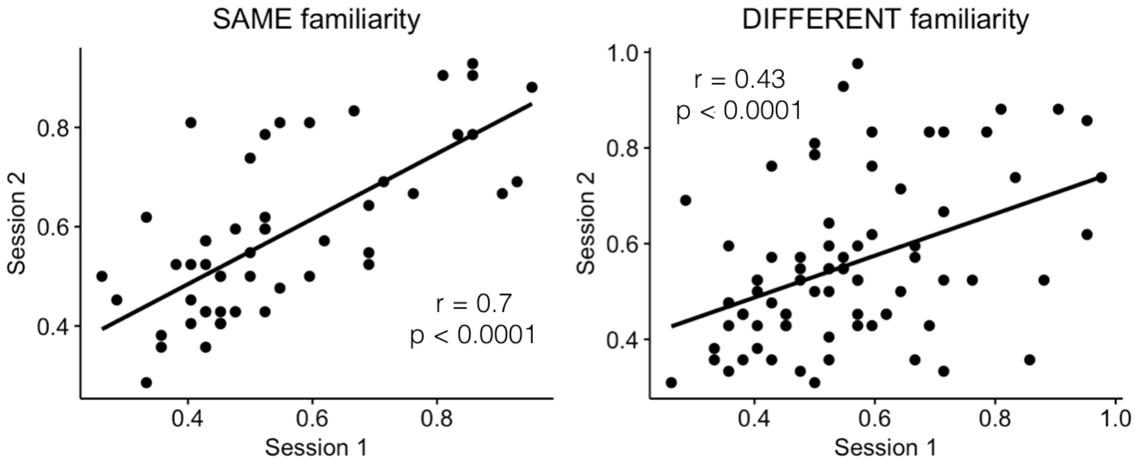

Figure 5. *Test-retest correlations in* SLscore *as a function of stimulus complexity.* The left panel shows the correlation between SLscore on Session 1 and Session 2 for those participants who performed the statistical learning task using stimuli of equal complexity each time (either FAMILIAR-FAMILIAR or UNFAMILIAR-UNFAMILIAR). The right panel shows the same correlation but for participants who saw stimuli of *different* complexity each time (either FAMILIAR-UNFAMILIAR or UNFAMILIAR-FAMILIAR). Correlations were significantly higher when the stimuli were of the same complexity, despite the fact that all were in the same (visual) modality and none of the specific items were repeated between sessions.

We tested this possibility in Experiment 2, which presents another group of participants with stimuli that were designed to be simpler than the UNFAMILIAR stimuli in Experiment 1, while still being unfamiliar (we call them UNFSIMPLE). If complexity was driving the results in Experiment 1, we would expect SL performance using the UNFSIMPLE stimuli to be similar to the FAMILIAR stimuli, since both are relatively simple; if not, we would expect performance on the UNFSIMPLE stimuli to be similar to the UNFAMILIAR ones, since both are unfamiliar.

Participants

80 participants were recruited from Amazon Mechanical Turk for the 15-minute task, for which they were paid $3.50USD. Of these, 6 were excluded for failing the attention check, which was identical to the one in Experiment 1. All analyses focus on the remaining 74 participants, 46 (62.2%) of whom were male and 70 (94.6%) were from the US. Ages ranged from 21 to 62 (mean 36.4). All participants were assigned to a single condition involving the UNFSIMPLE stimuli, described below. None of the participants were in Experiment 1.

Materials

Our goal in Experiment 2 was to design stimuli that were not as visually complex as the UNFAMILIAR stimuli in Experiment 1. Defining visual complexity is a difficult task and to our knowledge there is no agreed-upon answer (see, e.g., Luck & Vogel, 1997; Alvarez & Cavanagh, 2004); we thus followed the general principle that simpler items are those with both *fewer* and *more* *distinguishable* features. The UnfSimple stimuli were therefore composed of fewer features than the Unfamiliar ones (between two to four rather than four to six) and those features were more distinguishable (lines, arcs, and circles, not just lines). They are shown in Figure 6.

Figure 6. *Experiment 2 stimuli.* All participants saw 16 unfamiliar UnfSimple stimulus items that were designed to be visually less complex than the Unfamiliar ones in Experiment 1, with fewer and more distinguishable features.

Procedure

The procedure was identical to the statistical learning task in Session 1 of Experiment 1, except with UnfSimple stimuli instead.

Experiment 2: Results

As in the previous experiment, accuracy on the statistical learning task (SLscore) using the UnfSimple 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% ($t(73) = 6.2, p < 0.0001, d = 0.73$) and there was once again a high correlation between the item-level accuracy of the participants of Siegelman et al. (2017) and ours ($r = 0.66, p < 0.0001$).

Most importantly, as shown in Figure 7, overall statistical learning performance on the UnfSimple stimuli was more similar to performance on the Unfamiliar stimuli than the Familiar stimuli. In order to make the most appropriate comparison, we performed a one-way ANOVA on SLscore between UnfSimple and the Session 1 SL task performance in the Familiar and Unfamiliar conditions (results were not qualitatively different if instead we used the SL scores from both sessions). There was a significant effect of condition ($F(2,203) = 4.2743, p = 0.0152, \eta^2 = 0.04$). Follow-up t-tests indicated that the difference between Familiar and UnfSimple was significant ($t(113.8) = 2.68, p = 0.0085, d = 0.47$) but not the difference between Unfamiliar and UnfSimple ($t(142.1) = -0.63, p = 0.5271, d = 0.11$). In other words, people performed more similarly to the complex and unfamiliar stimuli than to the familiar ones. This suggests that *familiarity* rather than *complexity* underlay the differences in perceptual fluency and concomitant increase in SL performance observed in Experiment 1. We consider these issues further in the Discussion.

General Discussion

This work investigated the degree to which perceptual fluency influences statistical learning, and to what extent perceptual fluency reflects stimulus familiarity vs complexity. Recognising that

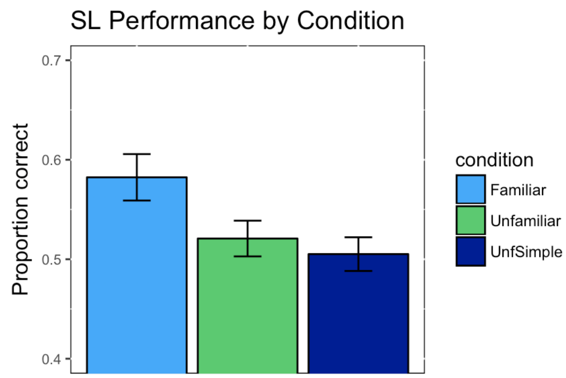

Figure 7. *Experiment 2: SL performance by condition.* Accuracy on the SL task (SLscore) was significantly higher for FAMILIAR stimuli than for either the unfamiliar complex stimuli (UNFAMILIAR) or the unfamiliar simple stimuli (UNFSIMPLE), suggesting that it is familiarity rather than complexity that improves statistical learning performance.

SL relies on basic memory processes, we hypothesised that participants would learn the same statistical distribution better when they found items to be easier to perceptually distinguish, parse, and remember. This hypothesis was supported: while participants performed significantly above chance on both SL tasks, they performed significantly better on the task containing overlearned items, even though the modality was the same in both cases. This suggests that the ease with which statistical distributions are learned is moderated not just by a person's modality-specific level of perceptual fluency, but also by their familiarity with the set of items. These effects were also observable at the individual level: test-retest correlations were higher when the two tasks used items that were equally familiar as each other, 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 perceptual fluency. Finally, we found that SL performance was not improved when the stimuli were *simpler* but still unfamiliar. 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 moderated by familiarity.

Our findings are consistent with the overwhelming evidence that frequency matters in almost every aspect of cognition, from language to decision making to memory (e.g., 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 statistical learning is also improved when the items involved are highly familiar (i.e., occurred frequently in previous experience). Our work thus suggests an important modification to existing theories of statistical learning (e.g., Frost et al., 2015). Rather than postulating that individual variation in the efficiency of perceptual coding is entirely based on modality or timing, our results indicate that familiarity plays an important moderating role. We do not suggest that there are no modality-specific individual differences in perceptual fluency, but we *do* suggest that familiarity may be far more important than has previously been recognised; indeed, it is possible that some of the variation that has historically been attributed to modality differences may reflect familiarity differences instead.

In Experiment 2 we attempted to disentangle the question of whether frequency *per se* was driving these effects, or whether it was actually an influence of stimulus complexity. Either way, the result would support the hypothesis that perceptual fluency plays an important role: but it would have different implications about what drives differences in perceptual fluency. We found that performance was far more affected by familiarity than complexity; when shown unfamiliar stimuli that were visually less complex than the unfamiliar stimuli in Experiment 1, people performed nearly identically to the UNFAMILIAR condition of Experiment 1. Of course, we cannot be certain that we succeeded in making our Experiment 2 UNFSIMPLE stimuli as visually simple as the FAMILIAR stimuli, but if raw visual complexity played a large role one might have expected at least a little improvement in performance on them. These results are consistent with research indicating that familiarity affects visual working memory (Jackson & Raymond, 2008; Xie & Zhang, 2017, 2018).

Whether familiarity would have the exact same effect on statistical learning in the auditory modality remains an open question, as to our knowledge comparable studies to ours do not exist. However, there is evidence in the auditory domain that high frequency and therefore more familiar items can play facilitative roles in speech segmentation in both 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 (e.g., 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 analogue of our current study would show something even more fundamental, such as familiarity with a lower-level features of language (e.g., a syllable inventory) providing an important basis for learning higher level forms (e.g., words).

Our work is interesting when considered in combination with entrenchment accounts of statistical learning, which suggest that the prior experience of transition probabilities can interfere with the learning of novel transition probabilities over the same items. For instance, Siegelman et al. (2018) showed that statistical learning of auditory linguistic stimuli (but not non-linguistic stimuli) was affected by prior linguistic experience, probably due to interference from native-language patterns of entrenchment. Similar interference effects may also explain opposite behaviour in similar SL tasks between speakers of different languages (Endress & Mehler, 2009; Perruchet & Poulin-Charronnat, 2012). Our results are an interesting contrast to these cases, because our focus has been on the familiarity of the *stimuli themselves* and because we found a facilitative rather than disadvantageous affect of prior experience. Indeed, in our studies it appears that the effect of stimulus familiarity was strong enough to override any interference caused by pre-existing 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 segmentation.

This work suggests part of a solution to the question of exactly what statistical learning *is* and how it relates to other cognitive skills. Perhaps SL tasks are poorly correlated with IQ and across modality, at least in part, because statistical learning is heavily mediated by perceptual fluency, and that in turn is strongly affected by familiarity and prior experience. One implication is that we should rethink the extent to which any particular statistical learning *tasks* are primarily measuring a domain-general statistical learning *factor*. These findings also suggest that one way to improve statistical learning (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 SL underlies learning 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 grant DP150103280. We would also like to thank Jing Qian for her intellectual contributions to the experimental design.

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

Arciuli, J., & Simpson, I. (2012). Statistical learning is related to reading ability in children and adult. *Cognitive Science*, *36*(2), 286-304.

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

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.

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.

Conway, C., Bauernschmidt, A., Huang, S., & Pisoni, D. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356-371.

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

Daltrozzo, J., Emerson, S., Deocampo, J., Singh, S., Freggens, M., Branum-Martin, L., & Conway, C. (2017). Visual statistical learning is related to natural language ability in adults: An ERP study. *Brain and language*, *166*, 40-51.

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.

Erickson, L., Kaschak, M., Thiessen, E., & Berry, C. (2016). Individual differences in statistical learning: Conceptual and measurement issues. *Collabra*, *2*(1), 14.

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.

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* (p. 81-86). Austin, TX: Cognitive Science Society.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243-1252.

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.* De Gruyter Mouton.

Hamrick, P., Lum, J., & Ullman, M. (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.

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.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376-405.

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* (p. 564-569). Austin, TX: Cognitive Science Society.

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

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

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184-193.

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.

Luck, S., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279-281.

Misyak, J., & Christiansen, M. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331.

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

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

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.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, *21*(4).

Reber, A., Walkenfeld, F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning Memory and Cognition*, *17*(5), 888-896.

Romberg, A., & Saffran, J. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 906-914.

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.

Shafto, C., Conway, C., Field, S., & Houston, D. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*, 247-271.

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.
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.
Xie, W., & Zhang, W. (2017). Familiarity increases the number of remembered pokemon in visual short-term memory. *Memory and Cognition*, *45*(4), 677-689.
Xie, W., & Zhang, W. (2018). Familiarity speeds up visual short-term memory consolidation: Electrophysiological evidence from contralateral delay activities. *Journal of Cognitive Neuroscience*, *30*(1), 1-13.

Appendix A: Further analyses

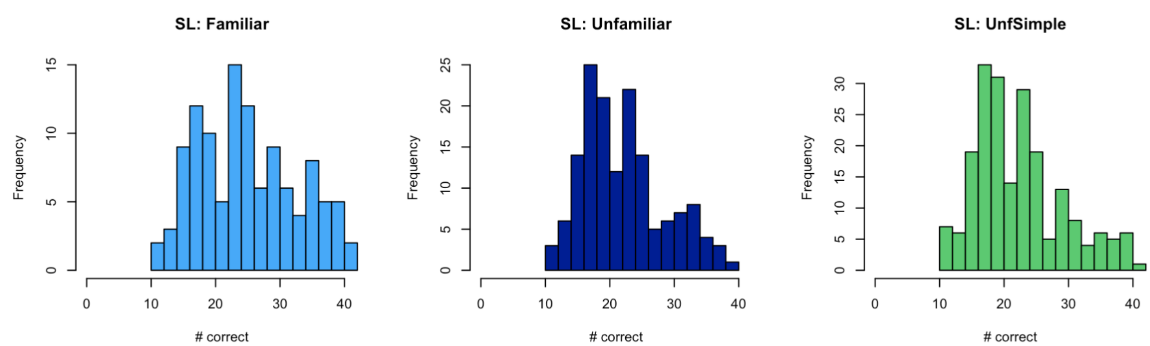

Figure 8. *Histograms of accuracy on the SL tasks.* Each histogram plots the individual overall SLscore for each participant, broken down by condition. It is evident that there is a wide range of individual variation within each condition, and that the histograms appear approximately normal and similar in character to those reported in Siegelman et al (2017).

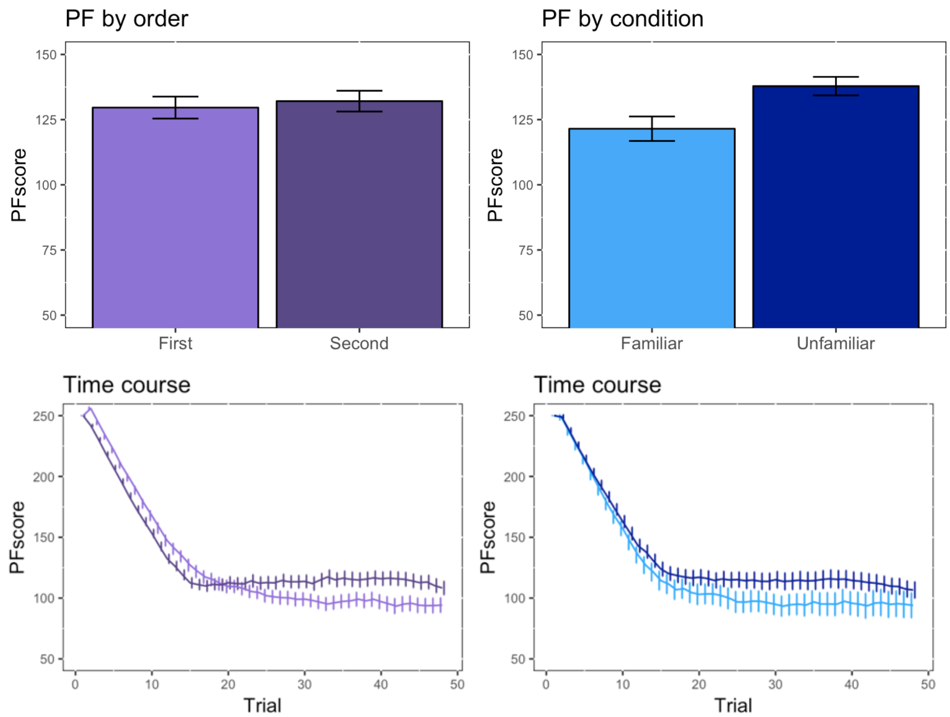

Figure 9. Performance in perceptual fluency task. The panels on the left show PFscore as a function of whether the PFtask was the first or the second the participant did. The time course (bottom) reveals that there is a small effect of both learning (improved performance initially on the second task) 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 FAMILIAR stimuli than with the UNFAMILIAR stimuli.