Splitting the variance of statistical learning performance:

A parametric investigation of exposure duration and transitional probabilities

Louisa Bogaerts[1], **Noam Siegelman**[2], **& Ram Frost** [2,3,4]

[1] CNRS & Aix-Marseille University, France

[2]The Hebrew University of Jerusalem, Israel

[3]BCBL, Basque center of Cognition, Brain and Language, San Sebastian, Spain

[4]Haskins Laboratories, New Haven, CT

Correspondence Address:

Louisa Bogaerts

CNRS & Aix-Marseille University, Laboratoire de Psychologie Cognitive

3, place Victor Hugo, Case D, 13331 Marseille Cedex 3, France

E-mail: bog.louisa@gmail.com

Acknowledgements: This paper was supported by the Israel Science Foundation (Grant 217/14 awarded to Ram Frost), and by the National Institute of Child Health and Human Development (RO1 HD 067364 awarded to Ken Pugh and Ram Frost, and PO1-HD 01994 awarded to Haskins Laboratories). Louisa Bogaerts is a research fellow of the Fyssen Foundation. We are indebted to Steve Frost for his valuable comments.

Abstract

What determines individuals' efficacy in detecting regularities in visual statistical learning? Our theoretical starting point assumes that the variance in performance in statistical learning can be split into variance related to efficiency in encoding representations within modality, and variance related to the relative computational efficiency of detecting distributional properties of the encoded representations. Using a novel methodology we dissociated encoding from higher-order learning factors, by independently manipulating exposure duration and transitional probabilities in a stream of visual shapes. Our results show that the encoding of shapes and retrieving their transitional probabilities are not independent and additive processes, but interact to jointly determine SL performance. The theoretical implications of these findings for a mechanistic explanation of SL are discussed.

Introduction

Statistical learning (SL), the learning of distributional properties of sensory input across time and space, is the mechanism by which cognitive systems discover the underlying regularities in the environment. As such, SL plays a key role in segmentation, discrimination, and categorization of the input, shaping the basic representations for a wide range of sensory, motor, and cognitive abilities (see Frost, Armstrong, Siegelman, & Christiansen, 2015, for a discussion). The term "SL" thus refers to the ability to learn and assimilate an array of possible statistical properties of sensory events. These include their aggregated relative frequency, their variance, and mostly, the extent of their co-occurrence (see Erickson & Thiessen, 2013, for a review). The present paper is concerned with the latter form of computation.

Starting from Saffran's original work (Saffran, Aslin, & Newport, 1996), which revealed that infants are able to segment speech on the basis transitional probabilities (TPs), a large number of studies have demonstrated that people often display remarkable sensitivity to the co-occurrence of items embedded in a continuous stream. This was shown across ages from adults to newborns, (e.g., Bulf, Johnson, & Valenza, 2011), across sensory modalities (*Visual*: e.g., Fiser & Aslin, 2001; Kirkham, Slemmer, & Johnson, 2002; Turk-Browne, Jungé, & Scholl, 2005; *Auditory*: e.g., Saffran, Newport, Aslin, Tunick & Barrueco, 1997; Gebhart, Newport & Aslin, 2009; *Tactile*: e.g., Conway & Christiansen, 2005), with both adjacent (e.g., Endress & Mehler, 2009) and non-adjacent contingencies (e.g., Newport & Aslin, 2004; Gómez, 2002). Interestingly, although in all these studies the tested sample as a group showed clear evidence of learning, not all individuals were shown to perform better than chance (see Siegelman & Frost, 2015, for a discussion). What determines the efficacy of detecting co-occurrence of events in a stream? Why do some individuals show clear evidence of learning in typical SL tasks, whereas others seem to perform at chance? What are the cognitive operations underlying this capacity? These complex questions hold the promise of revealing critical insights regarding the mechanisms driving SL, leading to deeper comprehension of what SL abilities could predict and why.

In a recent theoretical discussion of the factors influencing the variance of SL performance, Frost and his colleagues (2015) have suggested that this variance should be split into two main sources: 1. Variance related to *efficiency in encoding* the individual elements in the stream within the modality of their presentation, that is, the ability to create internal representations of each element of the continuous perceptual input. 2. Variance related to the relative computational *efficiency of detecting* the distributional properties of the encoded representations registering their TPs. Whereas the efficacy of creating detailed and reliable internal representations of individual elements appearing in a fast sequential input could be traced to the neuronal mechanisms that determine the effective resolution of one's sensory system, computational efficiency of detecting the TPs of these elements could be traced to capacities in binding temporal and spatial contingencies by the medial-temporal lobe memory system (Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014; Karuza et al., 2013; Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014). This view suggests that both encoding and binding abilities constrain learning of regularities, and they *jointly* determine the actual performance of an individual in a given SL task. Moreover, it presupposes some form of temporal processing modularity, where the internal representations computed from the inputs are subject to higher-level computations that bind them to register their distributional properties. Here we explore for the first time the possible predictions of this theoretical framework. We orthogonally manipulated factors related to encoding and binding constraints, and measured their relative contribution to SL performance on the group and individual level.

In the study we report here, we focused on performance in the extensively used, visual-statistical-learning (VSL) task (Arciuli & Simpson, 2011, 2012; Frost, Siegelman, Narkiss, & Afek, 2013; Siegelman & Frost, 2015; Turk-Browne et al., 2005). In the VSL task, participants are presented with a stream of complex visual shapes organized in pairs or triplets, when their constituent shapes follow each other in a predictable sequence (typically, TPs=1). Following a familiarization phase, participants are tested to assess their ability to report which shapes appeared in the stream in the original order. The VSL task allows a unique opportunity to experimentally address our theoretical question by disentangling *1)* the encoding and *2)* the learning of statistical dependencies components in SL. In any continuous stream of shapes experimenters can independently manipulate *1)* shape exposure duration (*ED*), that is, the amount of time that the stimulus is physically available for processing and *2)* the transitional probabilities (TPs) within the shapes. Whereas ED is a parameter affecting the efficacy of processing the visual stimuli for encoding them into internal representations (e.g., Potter & Levy, 1969, Loftus & Kallman, 1979), TPs is a parameter related to efficiency in registering their distributional properties. Jointly manipulating these parameters within-subjects could thus provide important information regarding individual susceptibility to encoding constraints vs. individual sensitivity to correlational transparency (see Frost et al., 2015, for discussion).

In the present study we did exactly this. Our participants, university students, participated in a series of VSL tasks, in all of which they watched evenly-paced streams of complex visual shapes. However, rather than using a fixed ED or a fixed TPs, as in most current SL studies (but see, Hunt & Aslin, 2001), in each session we manipulated the ED and TPs of shapes in the stream in a within-subject factorial design. EDs were set to 200, 600 or 1000ms per shape. TPs between shapes could be quasi-regular (0.6, 0.8), or fully regular (1.0). Following the familiarization phase, participants were tested to assess how well they learned the statistical contingencies of the shapes in each of the streams given the different presentation constraints. Thus, by looking at the change in performance across the tasks we could examine the independent influence of ED and TPs within the stream on SL performance, as well as their possible interaction.

This design allowed us to address, in parallel, critical theoretical questions that have not been addressed so far: What is the impact of incremental ED and TPs on SL? Do they impact SL independently (and additively) of each other, as predicted by a temporal processing modularity assumption, or do they show substantial interaction? If they do interact, what is the nature of this interaction? Finally, what does the distribution of individual sensitivity to both factors look like in the population? More generally, we ask how is human performance in extracting regularities from the input affected by different constraints (Linearly? Logarithmically? Inversed-U shaped?), when constraints are imposed on the time allocated for encoding events, and on varying the extent of event predictability.

Method

Participants

Fifty adults (12 males), all students at the Hebrew University, participated in the study for course credit or payment. Their age ranged from 21 to 27 (mean=23.4). Subjects were all native Hebrew speakers.

Design and Materials

The experiment required each participant to perform nine VSL tasks. The VSL tasks included 22 complex visual shapes (Turk-Browne et al., 2005). In each condition and for each subject, 16 of the 22 shapes were randomly chosen and randomly organized to create eight ordered pairs (the remaining six shapes were used for the screening items, see below). The eight pairs were presented continuously, one after the other, in a random order, to create a familiarization stream in which each pair appeared 24 times, with the constraint that the same pair could not be repeated twice in a row.

ED (the time that the shape appeared on the screen) and TP (the conditional probability of the second shape of each pair appearing after the first) were manipulated so that each factor included three levels: ED of 200, 600, and 1000ms, and TPs of 0.6, 0.8, and 1. Combining the three levels of each factor created nine tasks overall (see Figure 1). The interval between shapes was always fixed to 100ms to avoid introducing any chunking bias. Manipulation of TPs was done by including random noise in the 0.6 and 0.8 conditions: for example, for each pair AB during familiarization in the TPs=0.8 subtest, shape B appeared after shape A 80% of the time, while in 20% of the time shape B was randomly replaced by another shape X, while avoiding immediate repetition of shapes. Shape X was never the start of a new pair (i.e., only after its presentation, a new pair was presented).

Depending on the ED used, keeping the number of repetitions and shape interval constant, the familiarization phase of each of the tasks lasted from 2 to 7 minutes. Participants were asked to attend to the stream, and were not told that the stream was constructed of pairs. Following the familiarization stream, participants were instructed that they would now see two pairs of shapes on the screen (see Figure 1), and that their task would be to report which pair was more familiar to them. They were then tested with 38 2AFC trials. Thirty-two of the test-trials contrasted *1)* "true pairs"—two shapes that appeared as a pair during the familiarization phase (TPs between shapes being 0.6, 0.8 or 1.0 depending on the condition)—and *2)* "foils"—two shapes that did not appear as a pair during familiarization. Foils were constructed without violating the position of the shapes within the original pairs (e.g., for two true pairs AB and CD, possible foils could be AD or CB, but not AC or DB). Scores in the SL task ranged then from 0 to 32, calculated as the number of correct identifications of pairs during the test phase. The remaining 6 test-trials were aimed to identify and screen participants who did not attend the familiarization stream. These trials contrasted "true pairs" with a pair containing a novel shape that did not appear at all during familiarization (see Romberg & Saffran, 2013, for a similar procedure). Participants who missed 18 or more of the total of 54 screening items (six in each of the nine tasks) were excluded from the analyses. Following this screening, eight participants were excluded from the analysis. Due to a technical problem the data of two subjects in the ED=200, TP=1 condition were not saved. All subsequent analyses are based on the remaining 42 participants.

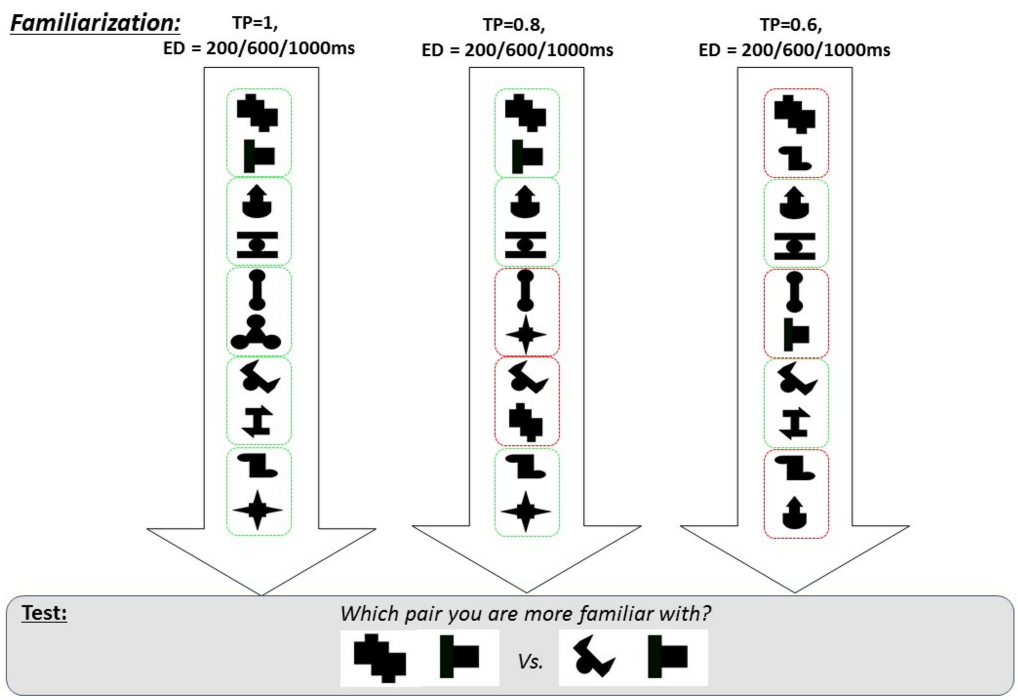

Figure 1. The nine VSL tasks: combining three TP levels (0.6, 0.8, 1) and three ED levels (200, 600, and 1000 ms).

General Procedure

The nine SL subtests were initiated by the participants from home, through an on-line platform. All nine subtests had to be completed in a period of 30 days, with no less than 24 hours between sessions. The mean time interval between sessions was 2.3 days (SD=1.2). Participants were instructed to do the task alone in a quite room and to avoid external distractions (turn-off their cell phone, and turn off music) and were asked to have only the experiment window open. The order of the tasks was random.

Results

Group level

Table 1 summarizes performance in the nine tasks. Performance in all nine conditions was significantly better than chance (one-sample t-test comparing mean performance to 50% chance, all *p's*<0.01). This suggests that successful learning was present even with the very fast presentation rates and lower TPs. A one-way repeated measures ANOVA with session number (1-9) as a predictor, revealed that level of performance did not significantly change across the nine sessions (F(8,312)=1.49, p=0.16, see also Siegelman and Frost, 2015, for similar findings).

	TP=0.6	**TP=0.8**	**TP=1**
ED=200ms	54.8% (11%)	56.3% (14%)	58.4% (14%)
ED=600ms	59.5% (13%)	59.7% (15%)	67.6% (17%)
ED=1000ms	61.8% (12%)	65.6% (17%)	72.8% (18%)

Table 1. Mean performance rate in each of the nine tasks (in parenthesis – standard deviations)

In order to examine the influence of TPs, ED, and their interaction on SL performance we conducted a logistic mixed-effect analysis, using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). The dependent variable was accuracy in the forced-choice test (excluding the screening items). The model included fixed effects of TPs, ED, their interaction, the position of the target-pair within the forced choice question (whether the target was first or second), and trial number in the test[1]. The random effect structure included a by-subject random intercept and random slopes for TPs, ED, and their interaction. The predictors TPs and ED were both centered and standardized, trial number was centered, and the variable of target position was dummy coded (target in first position – 0; second position – 1). The model included N=12032 observations, and had a log-likelihood of -7664.7.

We found a significant main effect of TPs ($\beta=0.18$, $SE=0.04$, $p<0.001$) and of ED ($\beta=0.23$, $SE=0.04$, $p<0.001$). These effects are presented in Figure 2: as can be seen, the effect of ED was found to be linear (with an improvement of 6.0% from 200 to 600 and a similar improvement of 4.5% from 600 to 1000). A paired t-test on the individual differences scores between the adjacent levels of ED confirmed that the extent of improvement between the lower two and top two ED levels did not differ significantly ($t(41)=0.61$, $p=0.54$).[2]

In contrast, the effect of TPs seems to deviate from linearity (with a small difference of 1.9% between the lower TPs of 0.6 and 0.8, which both implicated quasi-regularities, and a larger difference of 5.9% between 0.8 and full regularity TP=1). Indeed, a paired t-test revealed a

[1] Trial number was not a significant predictor of performance, $\beta=-0.002$, $SE= 0.002$, $p=0.18$ suggesting that the repeating the same target pairs and foils in the test phase did not alter performance.

[2] As in our procedure the ISI between shapes remained constant (see p. 7), the possibility that our results reflect rate (or length) of presentation rather than ED per se, cannot be overruled, and should be acknowledged.

marginally significant difference between TP=0.6 to TP=0.8 and TP=0.8 to TP=1 ($t(41)=1.78$, $p=0.08$), suggesting a possible non-linearity in the influence of TPs on SL performance. This suggests that there is a possibly qualitative difference between full regularity and quasi-regularity of shapes in the stream.

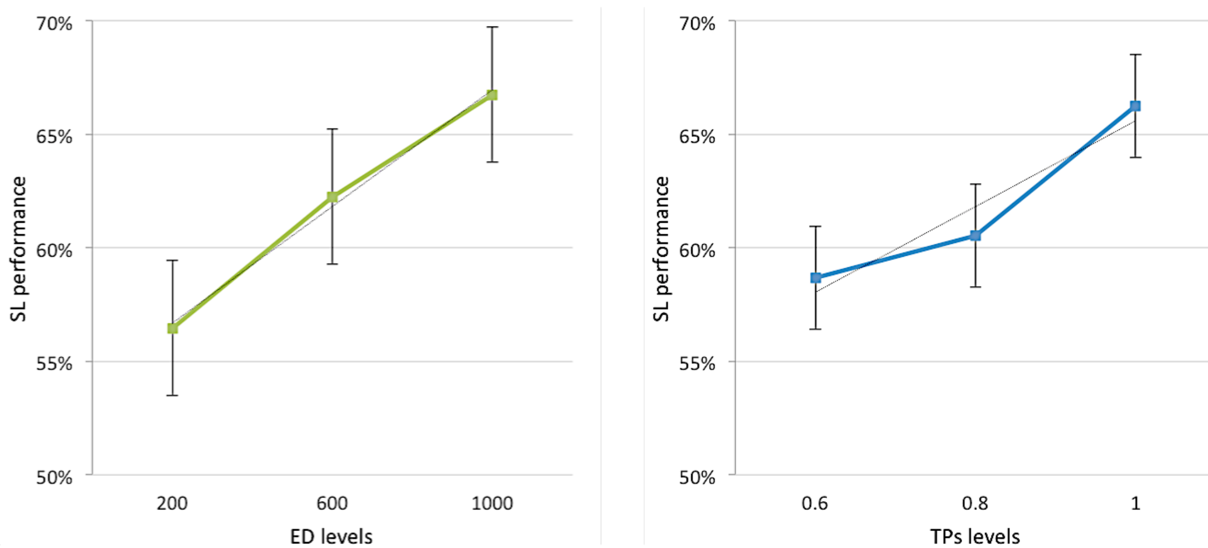

Figure 2. Mean scores in function of the TPs (left) and ED (right) manipulation. Error bars denote standard errors. The thin grey lines represent the best linear fit.

In addition to the two significant main effects of TPs and EDs, we found a significant interaction between TPs and ED (β=0.09, SE=0.03, $p<0.01$), which is depicted in Figure 3. In the fast ED of 200ms, performance was low even with high TPs (although above chance in all three conditions), suggesting that the fast presentation rate impaired detection of co-occurrence of shapes. For the intermediate and slow ED levels (600ms and 1000ms), the lines reflecting the three levels of TPs diverge – whereas performance was found to be similar in the two conditions that implicated quasi-regularities (TP=0.6 and TP=0.8), a different trend was revealed with full regularity (TP=1) with high performance already with ED=600ms. These data suggest an interesting interplay between encoding constraints and extent of regularity in determining SL performance. We further discuss this in the general discussion.

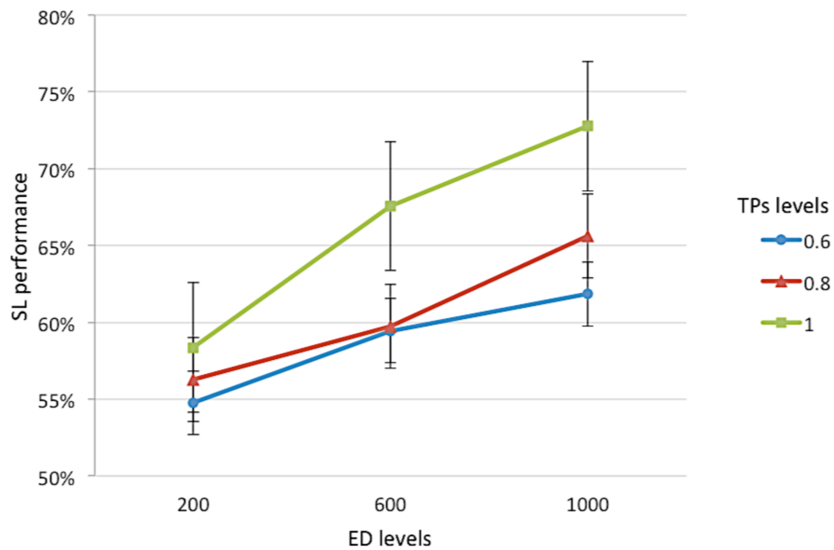

Figure 3. Interaction between TPs and ED. Error bars denote standard errors.

Individual level

We now turn to the insights revealed by considering patterns of individual performance. For each individual we extracted his/her slope for the effect of ED and TPs from the mixed logit model by looking at the by-subject random slopes. Note, that given the small number of manipulated levels of ED and TP per individual (three levels in each factor only), these slopes inevitably incur substantial noise. Nevertheless, in spite of this noise, some patterns stand out, pointing to interesting directions for future research. Figure 4A presents the scatterplot of slopes for ED and TPs of all participants in our study. The striking result is the high correlation between slopes within participants ($r=0.55$; $p<0.001$). Two outlier observations stand out, and if removed the correlation between slopes increases to $r=0.75$ ($p<0.001$). Thus, it seems that participants who showed greater sensitivity to changes of ED, tended to show greater sensitivity to changes of TPs, and vice versa. Note that some of this observed correlation could be driven by participants who did not show any evidence of learning: because these participants failed to exhibit any learning, they produced similar flat slopes for ED and TPs affecting the size of the correlation. However, even if individuals who did not exhibit significant learning at the individual level across the tasks (n=9 excluded participants) are removed from the analysis[3], a substantial correlation between susceptibility to ED and TPs still remains (r=0.48; p<0.01, see Figure 4B). These intriguing findings suggest that from an individual differences perspective, the ability to benefit from extended event duration, and greater event regularity, seems to be a unified individual capacity.

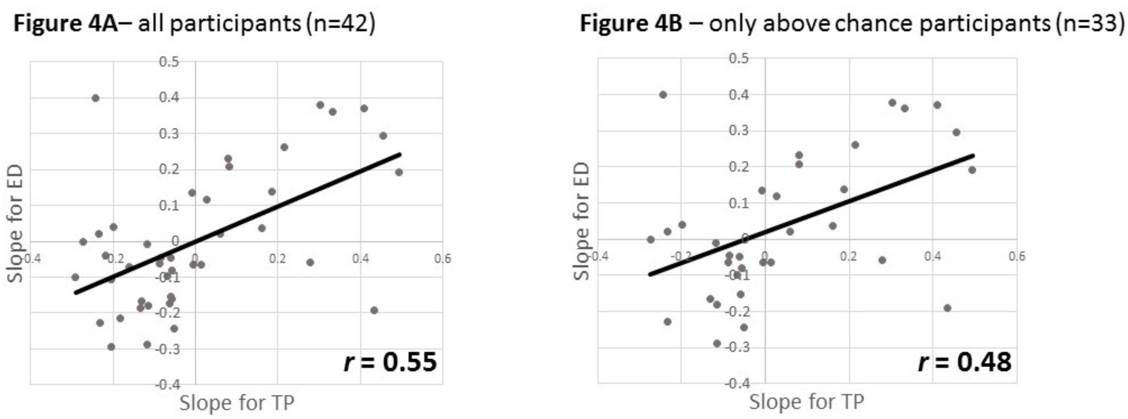

Figure 4. Correlation between slopes for ED and TP, over the whole sample (4A, left) and only for participants exhibiting learning at the individual level (4B, right).

Discussion

In the current study we independently manipulated TPs and ED in a visual SL task to dissociate factors related to encoding of visual shapes from the higher-order process of learning their distributional properties. We asked how each type of constraint affects SL, and how their joint interaction determines performance in the task. Our results provide a set of critical findings. First, we found that, at least within the range of 200 to 1000ms, ED impacts SL performance in a linear way, so that longer exposure of shapes results in better learning of their conditional

[3] The exclusion criterion for this analysis was set to success in 159 trials out of 288, across the nine tasks (i.e., a mean success rate of 55.2%). According the binomial distribution, this is the minimal number of success trials needed to present significantly above chance learning at the individual level.

probabilities. This converges with the earlier evidence provided by Turk-Browne et al. (2005) and Arciuli and Simpson (2011) who manipulated ED between subjects and reported improved SL performance at slower presentation rates (see also Emberson, Conway, & Christiansen, 2011). Second, we found that introducing quasi regularity in the stream impacts learning in a trajectory that seems to deviate from linearity. Although this deviation was marginally significant, our results show relatively small changes in SL performance when TPs increase from 0.6 to 0.8, but a substantially large improvement in performance when TPs implicate full regularity. This pattern of performance suggests that, at least on the group level, full regularity of shapes in the streams may be qualitatively different from any quasi-regularity, in terms of improving SL performance.

However, importantly, our experimental design allowed us to go beyond the independent influence of ED and TPs within the stream on SL performance, and examine their interaction. The striking finding of our study is the interplay between encoding constraints and extent of regularity in determining learning outcomes. Overall, our results suggest that the sensitivity to extent of TPs in the stream was modulated by exposure duration and vice-versa. Although our findings show that even very short exposure durations (ED of 200ms) were sufficient to encode the visual shapes resulting in above-chance learning, the extent of regularity of shapes in the stream had a relatively smaller impact on learning, so that SL performance was relatively low even with full regularity. With additional exposure (ED of 600ms), a large difference in performance between full regularity (TPs=1) and quasi-regularity (TPs=0.6, 0.8) emerged, but no difference between the two levels of quasi regularity. Full sensitivity to the full range of TPs was found only with the longer exposures of shapes. This pattern of findings suggests that encoding of shapes and retrieving their transitional probabilities are not independent and additive processes. Rather, the distributional properties of shapes in the stream and their predictability may serve to *facilitate* their encoding in case of suboptimal shorter exposure durations, and conversely, increase of exposure time enhances sensitivity to fine differences in TPs. These findings have implications to a mechanistic description of the cognitive events occurring in the typical VSL task. Rather than considering temporal processing modularity where encoding of shapes into internal representations feeds into the subsequent phase of extracting their distributional properties, encoding and extraction of TPs seem to be a two-way street, one affecting the other. Whether this bidirectional dependency is causal in one direction or another requires further investigation.

The present findings are relevant to current debates regarding the extent of modality specificity in SL (see Frost et al., 2015, for review and discussion), and the relations between the sub-processes involved in SL (e.g., Thiessen, Kronstein, & Hufnagle, 2013). In the context of visual shapes, recent imaging studies implicate, on the one hand, higher-level visual networks (Nastase, Iacovella, & Hasson, 2014), and on the other hand, the domain-general hippocampus and the medial temporal lobe (MTL) memory system (Schapiro et al., 2014; Turk-Browne, Scholl, Chun, & Johnson, 2009). Our findings thus offer possible constraints for understanding how both modality specific (encoding of visual shapes) and modality general (extracting distributional properties) computations result in extent of learning regularities in the visual modality. These processes do not seem to be independent and sequential, where the completion of one initiates the launching of the other.

Our discussion so far has focused on the group level performance, yet from an individual differences perspective, another striking result is the high correlation between ED and TPs trajectories within participants. This high correlation suggests that individuals who showed greater sensitivity to changes of ED tended to show also greater sensitivity to changes of TPs, and vice versa (note that this correlation is based on a relatively large number of participants, and was held even after removing individuals who did not exhibit significant learning). This is an intriguing finding as it suggests that the individual ability to overcome encoding constrains (here operationalized as limitations of event duration), and to overcome learning constraints (here operationalized as noise related to event regularity), are interrelated. A possible interpretation of this finding is that, perhaps, the high correlation between sensitivity to ED and TPs was driven by peripheral factors such as a general state of attentiveness to the task. To investigate this hypothesis we calculated the partial correlation between individual ED and TP slopes, controlling for average performance on screening items, a proxy of attentiveness. The partial correlation between slopes within participants are still large and significant (full sample: $r_{partial}$=0.43, p<0.01; after removing outliers: $r_{partial}$=0.68, p<0.001; only above chance participants: $r_{partial}$=0.40, p<0.05).

A point that deserves some attention is the presence of a number of negative individual slopes (see Figure 4). Whereas a negative slope associated with the ED manipulation intuitively makes sense (some individuals who are fast encoders might fail to allocate attention when shapes in the stream are presented at a slow rate), the negative slopes for TPs is harder to explain. Although it is possible that the negative slopes for TPs represent simple noise, which is inevitable in such experimental design, a possible insight for this phenomenon can be drawn from recent studies suggesting that different populations of neurons encode full regularity and quasi regularity (Nastase et al., 2014). From the perspective of information theory, quasi-regularity is more informative than full regularity. Indeed, Kidd, Piantadosi, and Aslin, (2012) have recently shown that infants maximally attend to stimuli that are neither too predictable nor too unpredictable. The negative slopes of some of the participants in our sample may possibly hint towards individual differences in the point of optimal degree of extent of regularity for learning. This, however, requires further investigation, aiming to establish whether individual slopes for ED and TP are indeed stable characteristics of an individual (see Siegelman & Frost, 2015, for measures of reliability in SL tasks).

In conclusion, the present study suggests that manipulating task parameters in a within-subject parametric design provides considerable insight regarding the cognitive operations underlying visual SL. Research using similar methodology has the promise of establishing how encoding and higher-order learning factors account for the variance in performance in other modalities, leading to a better understanding of the mechanisms of SL.

References

Arciuli, J., & Simpson, I. C. (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. C. (2012). Statistical Learning Is Related to Reading Ability in Children and Adults. *Cognitive Science*, *36*(2), 286–304.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8. Retrieved from http://cran.r-project.org/package=lme4

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

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology (2006)*, *64*, 1021–1040.

Endress, A. D., & 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. C., & Thiessen, E. D. (2013). Beyond word segmentation: A two-process account of statistical learning. *Current Directions in Psychological Science, 22(3)*, 239-243.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*, 499–504.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243–52.

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

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

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology. General*, *130*, 658–680.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *Plos One*, *7*(5), e36399.

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8997–9002.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Loftus G. R., & Kallman H. J. (1979). Encoding and use of detail information in picture recognition. *Journal of Experimental Psychology: Human Learning and Memory, 5(3)*, 197-211.

Nastase, S., Iacovella, V., & Hasson, U. (2014). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Human Brain Mapping*, *35*, 1111–1128.

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

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81(1)*, 10-15.

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*, 1290–1320.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. Psychological Science, 8(2), 101–105.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The Necessity of the Medial Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*.

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

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology-General*, *134*(4), 552–564.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–45.