Tasks determine what is learned in visual statistical learning

Timothy J. Vickery, Su Hyoun Park, Jayesh Gupta

University of Delaware

Marian E. Berryhill

University of Nevada, Reno

CORRESPONDING AUTHOR: Timothy J. Vickery

Department of Psychological and Brain Sciences

University of Delaware

108 Wolf Hall

Newark, DE 19716

Email: tim.vickery@gmail.com

Abstract

Visual statistical learning (VSL), the unsupervised learning of statistical contingencies across time and space, may play a key role in efficient and predictive encoding of the perceptual world. How VSL capabilities vary as a function of ongoing task demands is still poorly understood. VSL is modulated by selective attention and faces interference from some secondary tasks, but there is little evidence that the types of contingencies learned in VSL are sensitive to task demands. We found a powerful effect of task on what is learned in VSL. Participants first completed a visual familiarization task requiring judgments of face gender (female/male), or scene location (interior/exterior). Statistical regularities were embedded between stimulus pairs. During a surprise recognition phase, participants showed less recognition for pairs that had required a change in response key (e.g., female followed by male) or task (e.g., female followed by indoor) during familiarization. When familiarization required detection of 'flicker' or 'jiggle' events unrelated to image content, there was weaker, but uniform VSL across pair types. These results suggest that simple task manipulations play a strong role in modulating the distribution of learning over different pair combinations. Such variations may arise from task and response conflict, or because the manner in which images are processed is altered.

Keywords: Visual statistical learning; response selection; task switching

Perceptual systems face two severe challenges: the inverse problem of determining causes from noisy inputs, and capacity limitations. One proposed coping mechanism is to chunk information that consistently co-occurs spatially or temporally. When acquired in an unsupervised manner, this is called ‘statistical learning’ (SL). This phenomenon happens in infants (Saffran, Aslin, & Newport, 1996) and adults (Saffran, Johnson, Aslin, & Newport, 1999), who acquire and flexibly express learning of such contingencies across sensory domains. Demonstrations of *visual* SL (VSL) are prevalent for spatial (Fiser & Aslin, 2001, 2002b) and temporal (Fiser & Aslin, 2002a) contingencies. Temporal VSL permits recognition of stereotyped sequences following exposure. Learning effects are observed in explicit recognition rates as well as enhanced performance (Turk-Browne, Jungé, & Scholl, 2005).

VSL seems to occur without explicit awareness of contingencies present during familiarization. Thus, VSL is usually studied using minimal, unvarying demands during learning, and little is known about resource requirements and task interactions with VSL. Prior work probed relationships between temporal VSL and selective attention. Turk-Brown, Jungé, and Scholl (2005) found that participants viewing a stream of bicolored objects selectively learned the *attended* color stream, suggesting that selective attention ‘gates’ VSL (but see Musz, Weber, & Thompson-Schill, 2015). Pursuing a different hypothesis regarding shared mechanisms between VSL and statistical summary formation, Zhao, Ngo, McKendrick, and Turk-Browne (2011) showed statistical summary tasks performed during familiarization disrupted VSL compared with a non-summary task. A follow-up suggested that this may be due effects on focal attention and working memory (Hall, Mattingley, & Dux, 2015). These findings suggest that VSL is likely sensitive to familiarization task demands, although both effects could be due to changes in the distribution of selective attention within the sequence.

Is VSL subject to other forms of interference? VSL is correlated with caudate activity (Turk-Browne, Scholl, Chun, & Johnson, 2009), which is also implicated in response selection and task-set switching (Cools, Clark, & Robbins, 2004; Crone, Wendelken, Donohue, & Bunge, 2006; Seger, 2008). Stevens, Arciuli, and Anderson (2015) demonstrated that concurrent motor activity disrupts VSL compared to passive familiarization. They concluded that coordinated motor activity and extraction of statistical regularities rely on common resources. The role of the caudate in stimulus-response learning motivated us to test whether VSL is affected by simple response and task requirements during familiarization.

An additional question relevant to this work is whether stimulus similarity plays a role in VSL. Most prior research has employed homogenous stimuli, e.g., abstract shapes. Several paired faces and scenes (e.g., Turk-Browne, Scholl, Johnson, & Chun, 2010), but did not compare within-category to cross-category performance (although statistical relationships among conceptual categories can support SL-like recognition; Brady & Oliva, 2008).

The present study

We asked how changes in task or response modulate VSL. Surprisingly, there were robust effects of simple response-selection demands, such that the mapping of paired stimuli to the same or different responses during training had a significant effect on subsequent recognition (Experiments 1-2). A less reliable effect of task-switching also emerged, with weaker learning for cross-category compared to within-category pairings. Experiment 3 demonstrated that there is no notable influence of categorical similarity on VSL in the absence of varying training task demands. Experiment 4 randomly assigned participants to either categorize or detect “jiggle” events, finding deleterious effects of both task and response switching on VSL selective to the categorization group. Familiarization task demands affect VSL, affecting not only the overall strength of learning, but also the distribution of learning across types of stimulus combinations.

General Methods

Methods jointly summarize 4 experiments. All procedures were approved by the University of Delaware’s Institutional Review Board.

Participants. Participants (Experiment 1:N=20, Experiment 2:N=30, Experiment 3:N=25) were University of Delaware undergraduates and recruited online via Amazon Mechanical Turk (Experiment 4:N=104), reporting normal or corrected-to-normal visual acuity and color vision. They were compensated with course credit or cash and provided informed consent.

Apparatus. Experiments 1-3 employed Linux PCs with 17” CRT monitors (resolution: 1280x1024 at 75 Hz), operating MATLAB 2015a (Mathworks; Natick, MA) with Psychophysics Toolbox 3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Participants sat ~54 cm from the screen and responded via keyboard. Experiment 4 employed the participant’s web browser and computer using jsPsych 5.0.3 (de Leeuw, 2015) and Psiturk (Gureckis et al., 2016).

Stimuli. Face images were derived from the FERET database (Phillips, Wechsler, Huang, & Rauss, 1998), cropped to minimize background. Scene photos were collected from the Internet, depicting interior and exterior scenes. Images were 200x200 pixels (approximately 5.3°x5.3°).

Procedures. Participants completed a familiarization phase followed by a surprise recognition phase in one-hour sessions.

Familiarization phase. Prior to familiarization, 16 *AB* pairs of images (32 'paired' images) were randomly predetermined (Figure 1A) – *A* always preceded *B* during familiarization. Pairs were determined such that there was one of each unique combination of female, male, interior, and exterior (e.g., four pairs consisted of a different female (*A*), followed by (*B)* a female/male/interior/exterior image). For experiments 1-3, 16 unpaired 'singleton' images also appeared (4 per image type). All images appeared 4 times per each of 5 blocks, each block 192 trials (Experiment 1-3, Figure 1B), and 128 trials (Experiment 4, which excluded singletons during familiarization). Sequences were pseudo-randomized so pairs never immediately repeated. Images appeared for 1 s with an ITI of 1 s. A yellow circular frame fixation marker was superimposed at the center, remaining present during the ITI.

For Experiment 1, participants classified images as female/interior using the 'z' key and the left hand, or male/exterior using the 'm' key and the right hand. For Experiment 2, the sole difference was that 'n' and 'm' keys were used and participants used two fingers of one hand, to verify that response effects did not depend on using different hands. We encouraged accuracy and fast responses, which were accepted within a 2-s interval of stimulus onset. Correct and error responses resulted in turning the fixation marker green or red, respectively, until the end of the ITI.

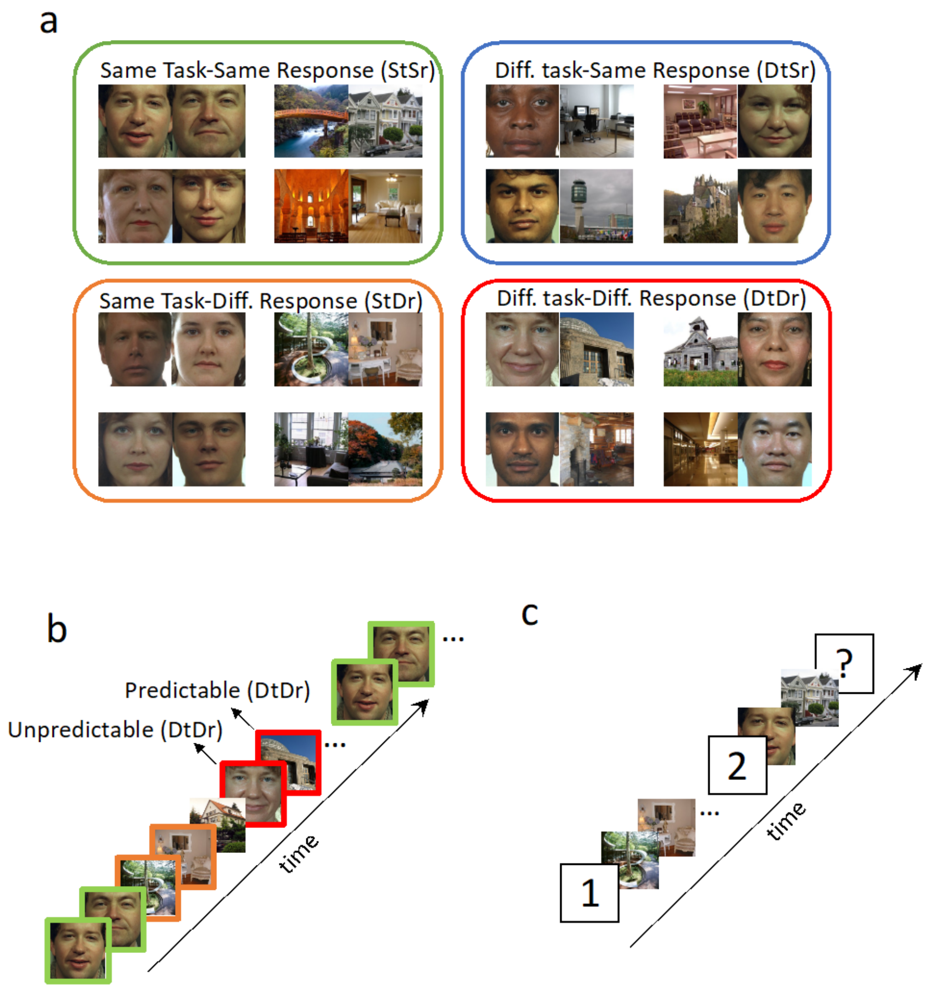

Figure 1. *(a) Example AB pairings for one subject. Not shown: 16 singleton items (4 of each type) that were not predictive or predictable in the training sequence (singletons did not appear in Experiment 4). (b) Sample trial sequence from the training sequence. Colored outlines are intended to depict the class of the pair, and were not shown to subjects. Each image appeared for 1 s and ISI was 1 s. A yellow fixation circle appeared overlaid on images and during ISI period (not depicted). The circle turned green after a correct response, red after an incorrect response. Trials after the first trial in the block were classified according to their relationship to the prior trial (Same/Different task, and Same/Different response required) and their membership (or not) in a paired set, which made them unpredictable or predictable. (c) Example of a recognition-stage trial.*

In Experiment 3, participants monitored images and pressed the space bar when an image 'flickered.' On 25% of image presentations, the image briefly turned off (53.3 ms) after 453.3 ms. Responses occurring until the following stimulus onset were hits.

In Experiment 4, participants were randomly assigned to groups. The 'categorization' group performed the task described in Experiment 1, except mappings of male/female and indoor/outdoor to the 'z' and 'm' keys were randomized across participants. The 'detection' group monitored streams of images for 'jiggle' events, occurring once per image per block, and pressed space when jiggle occurred. Jiggles occurred in non-adjacent pairs, and began 300 ms after image presentation with 2 cycles of displacement 5 pixels left/right of center, each cycle taking 200 ms. Participants received more feedback (e.g., "Error – press Z key for outdoor, M key for indoor"; 3000 ms) to account for possible inattention during instructions. Participants completed a practice familiarization phase of 32 trials with images that were not re-used.

Recognition phase. The recognition phase consisted of 64 forced-choice trials in which a target pair was matched against a foil pair (presented with same timing as familiarization and preceded by a 0.5-s sequence number label followed by a 0.5-s blank). Participants were informed before this phase about *AB* pairs, and indicated which sequence had appeared during familiarization. Trials were self-paced, responses unspeeded, and no feedback was provided.

Foil pairs had the same composition as targets but were re-combined *A* and *B* items. The *A* and *B* images in foil pairs remained consistently in the *A* or *B* position, but were swapped across pairs. All target and foil pairs appeared 4 times during the Recognition Phase.

Exclusion criteria. We excluded from analysis participants with familiarization accuracy <80% (Experiments 1-2, Experiment 4 categorization group), or a false alarm/miss rate >20%

during familiarization (Experiment 3, Experiment 4 detection group). One participant was excluded from Experiments 1 and 2, four were excluded from Experiment 3, and in Experiment 4, two were excluded from the categorization group, and none from the detection group.

Experiment 1

Experiment 1 examined whether VSL depends on response and task demands. *AB* pair images relied on the same or different task sets and responses (StSr, StDr, DtSr, and DtDr; the first letter indicates same/different task -- scene or face classification -- and the second letter indicates same/different response; see Figure 1A).

If response demands interfere with VSL, then pairs with the *same* response during familiarization (StSr, DtSr) should elicit superior recognition compared to pairs requiring different responses (StDr , DtDr). If executive demands matter, then same task pairs (StSr/StDr) pairs should be easier to learn than cross-task pairs (DtSr/DtDr). Similar effects might manifest if subcategorical and/or categorical distinctions influence learning, a possibility addressed in Experiments 3 and 4.

Results

Familiarization phase. We observed no effects of learning on performance (*p*'s>.29). There were effects on RTs of trial transition type (Figure 2A). Participants were slowed by two conditions: response switches (p=.006, e.g., female→male), and task switches (p=.01; e.g., male→outdoor), implying switching costs. For additional details see Supplementary Materials.

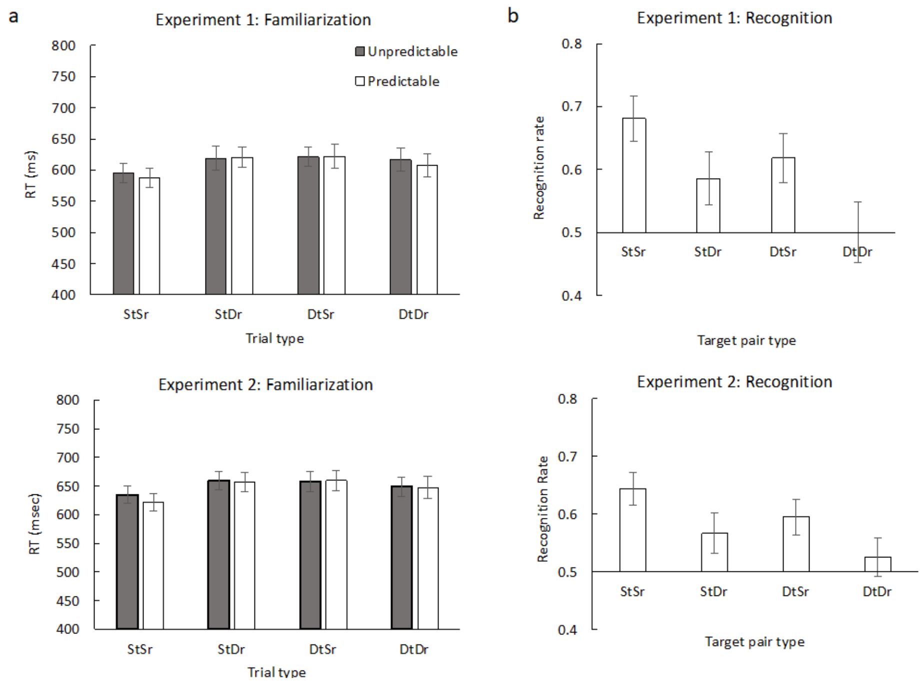

Figure 2. *a.) and c.) Mean RTs from Experiment 1, 2 familiarization stages averaged over blocks 2-5. b.) and d.) Mean correct recognition rates for the recognition stage for Experiment 1 and 2, respectively, separated according to pair type (Same or Different task, and Same or Different response). Dotted line reflects chance accuracy. Chance is .5. All error bars depict the standard error of the mean, in this and all other figures.*

Recognition phase. To assess learning, we compared recognition for each pair type (defined by training conditions: StSr/StDr/DtDr/DtSr) to chance (50%) using one-sample t-tests with Bonferroni correction (α=.05/4=.0125). Participants performed above-chance at recognizing StSr (t(18)=4.98, p<.001, *d*=1.14), and DtSr pairs (t(18)=3.05, p=0.007, *d*=0.70). Recognition for StDr (t(18)=2.02, p=.058, *d*=0.46), and DtDr (t<1) pairs did not differ from chance.

To compare learning across response and task, we conducted a 2 (response) x 2 (task) repeated-measures ANOVA on accuracy. There was a significant main effect of response ($F(1,18)=12.3$, $p=.002$, $\eta_p^2=0.41$), such that same-response were better recognized than different-response pairs. The main effect of task approached significance ($F(1,18)=4.28$, $p=.053$, $\eta_p^2=0.19$) -- same-task pairs were better recognized than different-task pairs. There was no interaction ($F<1$). A t-test contrasting DtSr with DtDr pairs provided direct evidence that switching responses across items of the pair impaired learning ($t(18)=2.46$, $p=.025$, $d=0.56$).

Discussion

During familiarization, task and response switching slowed responses. At recognition, participants performed above-chance when responses did not change. This suggest that the nature of pairings mattered and that maintaining the same response across a pair benefited subsequent recognition.

Experiment 2

Experiment 2 was a replication of Experiment 1, with higher power to elucidate borderline effects of task demands on VSL. We tested 30 participants based upon a power analysis suggesting that this sample would exceed 95% chance of detecting the task effect measured in Experiment 1.

Results

Familiarization stage. As in Experiment 1, responses were slower when response ($p=.03$) or task ($p=.02$) switched (see Figure 2C and Supplementary Materials for further details).

Recognition stage. Participants recognized StSr (t(28)=5.11, p<.001, *d*=0.95) and DtSr (t(28)=3.03, p=.005, *d*=0.56) pairs significantly above chance (corrected α=.0125), but performance did not differ from chance for StDr (t(28)=1.92, p=.065, *d*=0.36), or DtDr pairs, (t(28)=0.78, p=.44, *d*=0.14) (Figure 1D).

An ANOVA revealed a main effect of response during familiarization, as in Experiment 1, (F(1,28)=10.4, p=.003, η_p^2=0.27). The main effect of task showed a numerical but non-significant trend, such that pairs involving task switches were less recognizable (F(1,28)=2.30, p=.14, η_p^2=0.08). There was no interaction (F<1). We return to the question of whether task switches impair pair learning in Experiment 4.

Discussion

Experiment 2 replicated Experiment 1. During familiarization, response switches slowed reaction times. During recognition, those pairs acquired with consistent responses showed evidence of VSL. A weaker task switch penalty did not reach statistical significance. We reasoned that familiarization task demands interfered with VSL, particularly response switching. An alternative explanation could be that VSL is more effective for pairs with higher inter-item similarity. This possibility is addressed below.

Experiment 3

This experiment minimized familiarization task demands and ensured consistent responses across pair types. The categorization task was replaced with flicker detection to ensure sustained attention. If response and task demands explained the pattern of selective VSL in Experiments 1 and 2, then removing them should equalize VSL across learning conditions. In Experiment 3, we maintain the same terminology and classification

(StSr/StDr/DtSr/DtDr) to facilitate comparisons with Experiments 1-2, although DtSr and DtDr distinctions were arbitrary in this context. A sample size of 20 was planned, as this experiment was followed and designed to match Experiment 1, but 26 were collected due to accidental oversampling.

Results

Familiarization phase (Flicker Detection). Performance accuracy was high (Mean false alarm rate: 0.8%; Mean miss rate: 4.7%). We neither anticipated nor observed variation in accuracy or RT as a function of trial type given the simple detection task.

Recognition phase. All pair types were recognized at above-chance levels (SS: t(20)=3.53, p=.002, *d*=0.77; StDr: t(20)=2.47, p=.023, *d*=0.54; DtSr: t(20)=2.70, p=.014, *d*=0.59; DtDr: t(20)=3.21, p=.004, *d*=0.70; Figure 4), although StDr and DtSr did not pass Bonferroni correction (α=.0125). A 2x2 ANOVA revealed no main effects of category or subcategory (F<1) and no interaction of category and subcategory, (F(1,20)=1.19, p=.29, η_p^2=0.06). No pairwise differences between conditions approached significance (ps>.2).

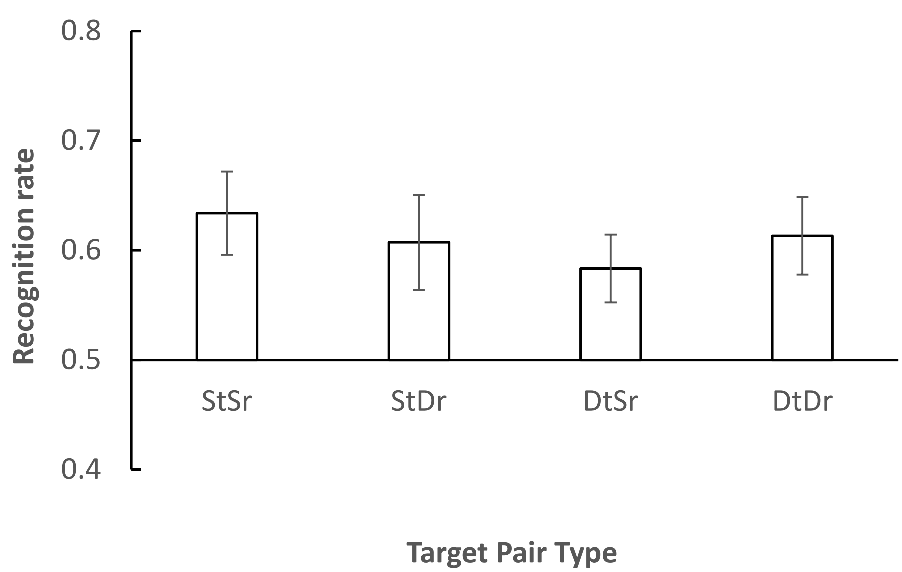

Figure 3. *Recognition rates for Experiment 3.*

Discussion

When the familiarization task was trivial, pairs showed statistically equivalent VSL. On the basis of this finding, we hypothesized that interference effects observed in Experiments 1-2 were likely due to task demands during familiarization.

Experiment 4

Our results suggest that response, and possibly task, demands influence what is learned in VSL. Yet, this interpretation requires comparisons across experiments. Experiment 4 compared a categorization with a detection task in a larger sample. We pseudo-randomly assigned participants to two equal-sized groups, one that completed a categorization task and another detected ‘jiggle’ events during familiarization.

Results

Familiarization phase. As expected, the Detection group who pressed a key when the shape ‘jiggled’ was highly accurate (Mean false alarm rate: 1.0%; Mean miss rate: 1.8%) and not significantly influenced by image type or predictiveness.

The Categorization group showed main effects of both task ($p=.002$) and response ($p=.003$). There was an interaction due to the subadditivity of those factors ($p<.001$); see Figure 4, and Supplementary Materials.

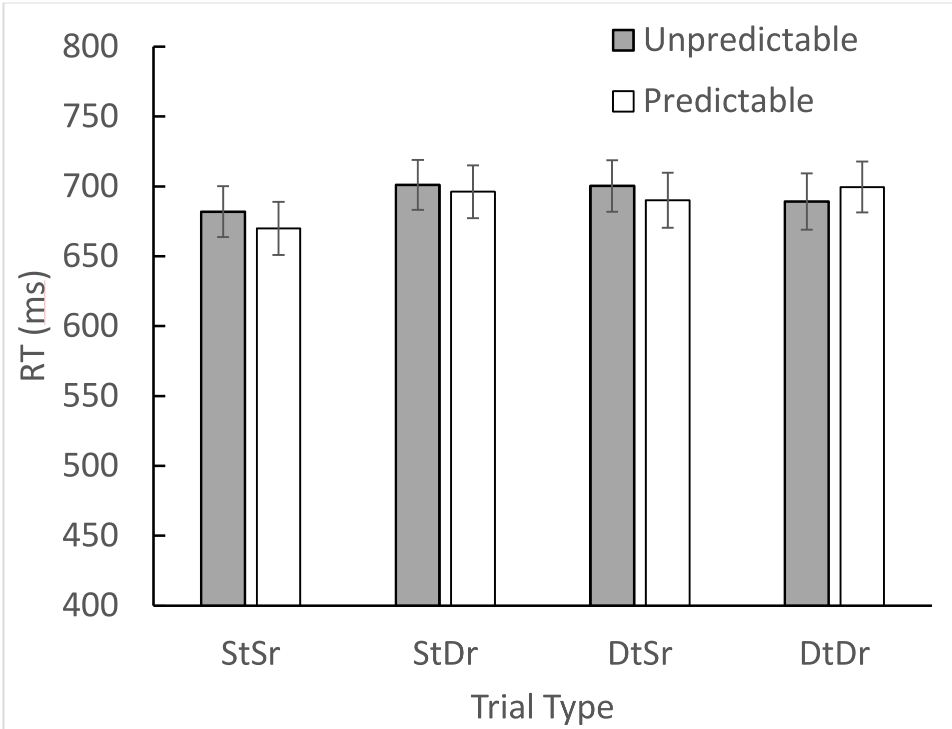

Figure 4. *Mean RTs from familiarization stage for the categorization group of Experiment 4, averaged over blocks 2-5.*

Recognition phase. As in Experiment 3, and to maximize comparability across groups, we separated Detection outcomes into StSr, StDr, DtSr, and DtDr bins, although the DtSr/DtDr distinction was arbitrary. We randomized assignment of image type to "response key" for the Detection group to equate distinctions across group assignment.

To identify significant group interactions, we conducted a mixed-effects ANOVA on accuracy at recognizing familiarized pairs, with the between-subjects factor of group, and task and response (same or different) as repeated-measures factors (Figure 5). Interactions of group x response ($F(1,100)=15.2$, $p<.001$, $\eta_p^2=.13$), and group x task ($F(1,100)=7.27$, $p=.008$, $\eta_p^2=.068$) reached significance. To understand these interactions, we conducted two follow-up repeated-measures ANOVAs for each group separately.

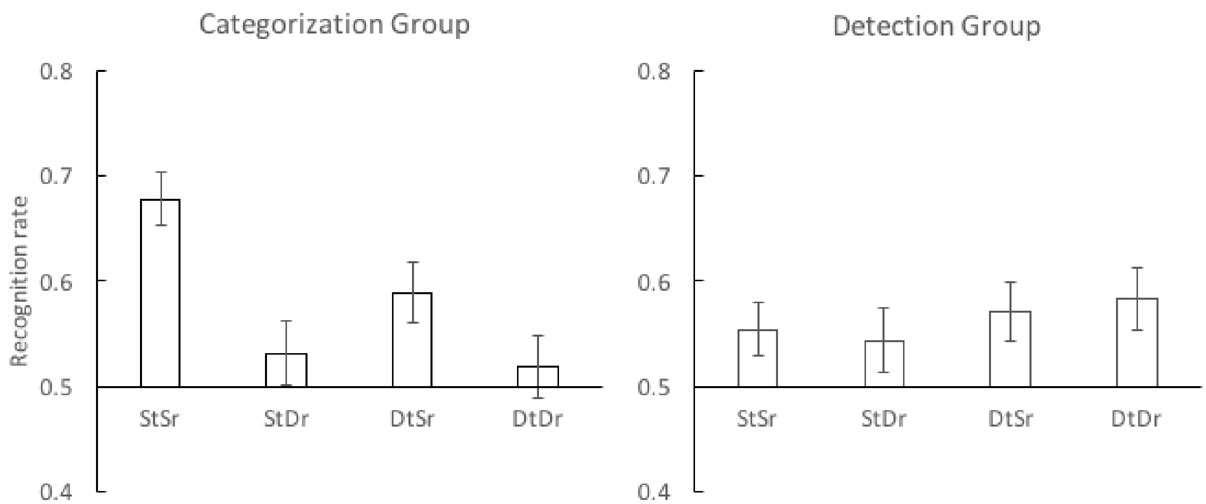

Figure 5. Experiment 4 *Recognition phase accuracy.*

For the Categorization group, results were similar to Experiments 1 and 2, albeit with significant task effects. One-sample t-tests comparing performance to chance recognition rates showed that *only* DtSr and StSr conditions exceeded chance (t(49)=3.13, p=.003, d=.44; and t(49)=7.01, p<.001, *d*=.99, respectively). The 2x2 ANOVA revealed a main effect of task was significant (F(1,49)=6.24, p=.016, η_p^2=.11); a task switch was associated with poorer subsequent memory, as was a response switch (F(1,49)=24.1, p<.001, η_p^2=.33). Finally, the interaction was just significant (F(1,49)=4.04, p=.05, η_p^2=.076), due to under additivity of task and response effects.

In contrast, the Detection group had significantly above chance recognition rates for DtDr and DtSr (t(51)=3.65, p<.001, *d*=.51 and t(51)=3.00, p=.004, *d*=.42). StDr (t(51)=1.93, p=.06, *d*=.27) and StSr (t(51)=2.06, p=.045, *d*=.29) were not significant following multiple-comparisons correction. A repeated-measures ANOVA revealed no significant main effects or interaction.

We contrasted group performance in the four conditions with independent samples t-tests. The groups significantly differed in StSr performance ($t(100)=3.38$, $p=.001$, $d=.67$), with better performance in the Categorization group, and the DtDr difference was trending ($t(100)=-1.72$, $p=.088$, $d=-.34$), favoring Detection group performance. Other differences were non-significant (p’s>.62).

Discussion

Experiment 4 provided evidence showing that the familiarization task influences what contingencies are learned. The passive, jiggle-detection familiarization task yielded more consistent learning across conditions. The Categorization group showed reduced learning in conditions that involved a task and/or response switch.

General Discussion

VSL depends on the familiarization task and on what types of statistical contingencies are learned. We found VSL disruption when stimulus categories were associated with different response sets and responses. Response-shift effects were more robust than task-shift effects, but both interfered with VSL. Importantly, VSL was *undifferentiated* among similar transitions under passive-viewing familiarization. Our results suggest that VSL operates equally well over similar and diverse stimuli when those combinations are familiarized under a passive task that does not confound response and task set differences with image content.

Why did task and response differences influence VSL recognition? Several possibilities deserve future investigation. We favor the interpretation that response and task selection demands interfere with acquisition. Even though switching response or task set imposed modest demands, consequences were evident in RTs during familiarization. VSL may rely, in part, on the mechanisms that support managing these sets; when demands on those mechanisms increase, learning suffers.

Alternatively, categorization task demands might have altered stimulus processing. For instance, the requirement to press a specific key in response to a female face may create a multimodal representation including the action and the face image. Similarities and differences across such representations might influence what is more or less likely to be learned. The removal of such contingencies at recognition may limit the expression of learned contingencies. Relatedly, the inclusion of categorical judgments may create event boundaries. Boundaries induced by semantic processing are known to influence subsequent order judgments, with better order memory “within” event boundaries (DuBrow & Davachi, 2013).

Might these results reflect variations of selective attention, thought to play a role in SL (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Turk-Browne, Jungé, & Scholl, 2005; Musz et al., 2015)? Our demonstrations were not based on dual-task procedures. Because interference occurred in the context of a task that required more than passive viewing, the results are unlikely attributable to simple variations in the overall strength of selective attention. Speculatively, categorization may have drawn selective attention to different features that the detection task, or it may have fluctuated under more demanding conditions of switches.

Our findings are consistent with observations of motor activity disrupting VSL (Stevens et al., 2015). Our data further suggest that interference effects may be due to response and task selection resources that may be critical to VSL. Our findings also make sense in light of neuroimaging data suggesting that VSL is correlated with striatal activity (Turk-Browne et al., 2009), a neural correlate of other forms of motor-related and categorical learning (Cincotta & Seger, 2007; Rauch et al., 1997).

'VSL' may be multiple dissociable phenomena. Bays, Turk-Browne, and Seitz (2016) provide evidence that recognition and performance effects may be dissociable. Here, we assessed learning by asking people to make judgments of familiarity. Other components of VSL may be inaccessible to report. Although we tried to access performance effects of learning during familiarization, we were largely unsuccessful. A different kind of transfer task, such as memory search through a temporal stream, might provide evidence of VSL where explicit recognition judgments fail, a possibility worth further investigation.

Conclusion

VSL is subject to multiple forms of interference. Here we found that different familiarization tasks produced different patterns of learning. Our findings highlight the fact that VSL is sensitive to the attention paid to stimuli during learning, and VSL manifests differently for different types of stimulus combinations depending on the task required.

Acknowledgments

This research was funded by NSF OIA 1632849 to TJV and MEB, NSF BCS 1558535 to TJV, and NSF OIA 1632738 to MEB.

References

Bays, B. C., Turk-Browne, N. B., & Seitz, A. R. (2016). Dissociable behavioural outcomes of visual statistical learning. *Visual Cognition*, *23*(9), 1072–1097. https://doi.org/10.1080/13506285.2016.1139647

Brady, T. F., & Oliva, A. (2008). Statistical Learning Using Real-World Scenes. *Psychological Science*, *19*(7), 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Cincotta, C. M., & Seger, C. a. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience*, *19*(2), 249–265. https://doi.org/10.1162/jocn.2007.19.2.249

Cools, R., Clark, L., & Robbins, T. W. (2004). Differential Responses in Human Striatum and Prefrontal Cortex to Changes in Object and Rule Relevance. *Journal of Neuroscience*, *24*(5), 1129–1135. https://doi.org/10.1523/JNEUROSCI.4312-03.2004

Crone, E. A., Wendelken, C., Donohue, S. E., & Bunge, S. A. (2006). Neural evidence for dissociable components of task-switching. *Cerebral Cortex*, *16*(4), 475–486. https://doi.org/10.1093/cercor/bhi127

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–

1286. https://doi.org/10.1037/a0034024

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2002b). Statistical Learning of New Visual Feature Combinations by Infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–15826. https://doi.org/10.1073/pnas.232472899

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., … Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842. https://doi.org/10.3758/s13428-015-0642-8

Hall, M. G., Mattingley, J. B., & Dux, P. E. (2015). Distinct Contributions of Attention and Working Memory to Visual Statistical Learning and Ensemble Processing. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(4), 1–36. https://doi.org/10.1037/xhp0000069

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, S14. https://doi.org/10.1068/v070821

Musz, E., Weber, M. J., & Thompson-Schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception &*

Psychophysics, *77*(1), 78–96. https://doi.org/10.3758/s13414-014-0757-5

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*. https://doi.org/10.1163/156856897X00366

Phillips, J. P., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(10), 1090–1104. https://doi.org/10.1109/34.879790

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, *16*(5), 295–306. https://doi.org/10.1016/S0262-8856(97)00070-X

Rauch, S. L., Whalen, P. J., Savage, C. R., Curran, T., Kendrick, A., Brown, H. D., … Rosen, B. R. (1997). Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. *Human Brain Mapping*, *5*(2), 124–132. https://doi.org/10.1002/(SICI)1097-0193(1997)5:2<124::AID-HBM6>3.0.CO;2-5

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by eight-month-old infants. *Science*, *274*(5294), 926–928. https://doi.org/http://dx.doi.org/10.1126/science.274.5294.1926

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

Saffran, J. R., 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. *Psychol Sci*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2007.07.010

Stevens, D. J., Arciuli, J., & Anderson, D. I. (2015). Concurrent Movement Impairs Incidental But Not Intentional Statistical Learning. *Cognitive Science*, *39*(5), 1081–1098. https://doi.org/10.1111/cogs.12180

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. https://doi.org/10.1037/0096-3445.134.4.552

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. *Dx.doi.org*, *21*(10), 1934–45. https://doi.org/10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit Perceptual Anticipation Triggered by Statistical Learning. *Journal of Neuroscience*, *30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Zhao, J., Ngo, N., McKendrick, R., Turk-Browne, N. B., Zhao, J., Ngo, N., … Turk-Browne, N. B. (2011). Mutual Interference Between Statistical Summary Perception and Statistical Learning. *Psychological Science*, *22*(9), 1212–1219. https://doi.org/10.1177/0956797611419304