



Stimulus variability and task relevance modulate binding-learning

Nithin George¹ · Tobias Egner^{2,3}

Accepted: 31 May 2021 / Published online: 19 July 2021
© The Psychonomic Society, Inc. 2021

Abstract

Classical theories of attention posit that integration of features into object representation (or feature binding) requires engagement of focused attention. Studies challenging this idea have demonstrated that feature binding can happen outside of the focus of attention for familiar objects, as well as for arbitrary color-orientation conjunctions. Detection performance for arbitrary feature conjunction improves with training, suggesting a potential role of perceptual learning mechanisms in the integration of features, a process called “binding-learning”. In the present study, we investigate whether stimulus variability and task relevance, two critical determinants of visual perceptual learning, also modulate binding-learning. Transfer of learning in a visual search task to a pre-exposed color-orientation conjunction was assessed under conditions of varying stimulus variability and task relevance. We found transfer of learning for the pre-exposed feature conjunctions that were trained with high variability (Experiment 1). Transfer of learning was not observed when the conjunction was rendered task-irrelevant during training due to pop-out targets (Experiment 2). Our findings show that feature binding is determined by principles of perceptual learning, and they support the idea that functions traditionally attributed to goal-driven attention can be grounded in the learning of the statistical structure of the environment.

Keywords Feature binding · Perceptual learning · Habitual attention · Visual search · Variability

Introduction

The visual system is organized hierarchically. Lower levels of the hierarchy are tuned to respond to basic stimulus attributes, such as orientation, color opponency, and motion direction. Higher levels of the hierarchy are involved in the recognition of complex, multi-feature objects. This hierarchical architecture necessitates the binding of individual features (such as shape and color) into an object representation. In spite of a long history of empirical research into feature binding in human visual cognition, typically in the context of visual search experiments, the underlying mechanisms remain debated. The field has moved from

a preoccupation with the role of attention in feature binding (e.g., Treisman & Gelade, 1980) to granting a key role to perceptual learning mechanisms in mediating efficient search for conjunctive features (e.g., Yashar & Carrasco, 2016). In the present study, we aimed to advance our understanding of the underlying mechanisms of this learning process by testing novel predictions for the learning of feature conjunctions derived from a prominent perceptual learning model (Yashar & Carrasco, 2016; Ahissar, Nahum, Nelken, & Hochstein, 2009). We begin with a brief review of the relevant background literature.

Feature binding and attention

The historically most influential account of feature binding is feature integration theory, which is grounded in a two-stage model of perception (Schneider & Shiffrin, 1977; Treisman & Gelade, 1980). According to this theory, primitive stimulus features like color and shape are processed in parallel in an initial “pre-attentive stage”, whereas the subsequent “attentive stage” binds features together to mediate object perception. The main contention of this model is that the deployment of attention is necessary for

✉ Nithin George
n.george@iitgn.ac.in

¹ Centre for Cognitive and Brain Sciences, Indian Institute of Technology Gandhinagar, 382355, Gujarat, India

² Department of Psychology and Neuroscience, Duke University, Durham, NC 27708, USA

³ Center for Cognitive Neuroscience, Duke University, Durham, NC 27708, USA

feature binding (e.g., to combine color and shape information to perceive a green triangle). In the seminal study by Treisman and Schmidt (1982), participants performed an identification task on a briefly presented display of objects defined by a conjunction of features (colored shapes) and a concurrent digit recognition task. Performing the attentionally demanding digit recognition task resulted in binding errors in the object recognition task, with participants reporting feature combinations that were not actually present in the display. Binding errors during an attentionally challenging task support the claim that attention is required for feature binding (Treisman & Schmidt, 1982). Lesions to brain areas associated with attention have also been shown to impair feature binding. For example, a Bálint's syndrome patient with bilateral parietal lobe lesions committed errors in binding shapes and colors, even though the recognition of individual shapes and letters was not impaired (Robertson, Treisman, Friedman-Hill, & Grabowecky, 1997; Robertson, Treisman, et al. 1995). Similarly, in patients with visual hemineglect, target detection in the contralesional hemifield is much more severely impaired when the target is differentiated from distractors by a conjunction of features compared to a single feature (Eglin, Robertson, & Knight, 1989; Esterman, 2000).

However, other data suggest that even very complex feature combinations present in more natural or familiar stimuli can be integrated in the absence of attention. For instance, it has been shown that an attentionally engaging concurrent letter discrimination task does not interfere with visual search performance when the targets are faces (Reddy, Reddy, & Koch, 2006; Reddy, Wilken, & Koch, 2004) or vehicles (Li, VanRullen, Koch, & Perona, 2002). The identification speed of familiar conjunctions, but not novel ones, has also been shown to be unaffected by transcranial magnetic stimulation (TMS) to right parietal cortex (implicated in hemineglect; Walsh et al., 1998). These findings suggest that feature binding for familiar objects, as opposed to say arbitrary shape–color combinations, can operate in the absence of focal attention. Our ability to combine features into object representations in very briefly presented natural scenes supports this claim. The deployment of focal spatial attention takes around 300 ms (Carrasco, 2011); however, observers have the ability to categorize complex natural scenes in 150 ms (Thorpe, Fize, & Marlot, 1996), faces in 100 ms (Crouzet, Kirchner, & Thorpe, 2010), and animals in 120 ms (Kirchner & Thorpe, 2006). Finally, task-irrelevant, unattended objects in a natural scene have been shown to be decodable from neural activity in object-selective areas in visual cortex (Peelen, Fei-Fei, & Kastner, 2009), suggesting that common, complex, multi-feature object information is present in the visual brain even when attention is occupied elsewhere.

Feature binding and perceptual learning

One way to reconcile these findings is to assume two distinct “modes” of binding, one being a “hardwired”, automatic (and thus, fast) integration of features in frequently encountered objects, and the other being a slow and attention-requiring “on-demand” binding process that can be applied to arbitrary feature conjunctions (VanRullen, 2009). However, a strong version of this dual-process view seems untenable in light of recent findings by Yashar and Carrasco (2016), who demonstrated that even a relatively short period of repeated exposure to arbitrary object feature conjunctions in very briefly presented search arrays (117 ms) can greatly enhance subsequent search performance for those objects. Participants performed a visual search task to report the presence/absence of a target defined by a color-orientation conjunction (e.g., a 50° tilted red line). During the training phase, one set of participants was pre-exposed to an alternate color-orientation conjunction (e.g., an 80° tilted blue line) that served as a distractor, while another set of participants was not pre-exposed to the alternate conjunction. The pre-exposed conjunction then served as the target in a subsequent test phase for both groups. The results showed a significant transfer of learning in the pre-exposed group, with perceptual sensitivity (d') for the former distractor in the test phase being comparable to the sensitivity for the prior target in the final block of the training.

This transfer of learning through brief pre-exposure of a (non-target) feature conjunction suggests that even the binding of arbitrary features can be learned quite rapidly without contributions from top-down attention. This form of perceptual learning—or “binding-learning”—may thus mediate our ability to quickly and effortlessly integrate features into objects that are encountered repeatedly (Yashar & Carrasco, 2016). This view, when applied to the divergent findings on the need for attention in feature binding reviewed above, suggests that participants in classic studies using arbitrary stimuli would have had little benefit from perceptual learning, whereas search for naturalistic stimuli (faces, natural scenes) would have benefited from vast prior perceptual learning experience (see Frank et al., 2014). In the current study, we strove to connect this new, perceptual learning perspective on feature integration directly to an integrative model of how perceptual learning takes place, by generating and testing novel predictions of that model for the manner in which feature conjunctions may be learned.

According to the reverse hierarchy theory (RHT; Ahissar et al., 2009), perceptual learning in tasks that recruit higher-order association areas is expedited by stimulus variability (or precision) and task relevance of the trained stimulus. RHT posits that perceptual learning always begins in higher-level areas, but can progress gradually to lower-level areas,

depending on task demands. This basic pattern of reverse progression of learning from higher-level to lower-level areas has been corroborated by a number of behavioral (Sowden, Davies, & Roling, 2000) and neuroimaging studies (Furmanski, Schluppeck, & Engel, 2004; Schwartz, Maquet, & Frith, 2002; Sigman et al., 2005). RHT holds that the transition of learning-induced plasticity from higher to lower areas in supporting perceptual learning depends on the variability of the stimulus features that people are exposed to during perceptual learning; in particular, high stimulus variability deters the reverse progression of learning from higher- to lower-level areas, whereas low variability promotes that progression. The reverse progression from generalizable learning in high-level areas to stimulus-specific learning in low-level area also necessitates task relevance of the trained stimulus.

The underlying assumption is that when trial-by-trial stimulus variability is high, for instance, when using a wide range of orientations in an orientation-discrimination task, the low-level neuronal populations that code for a specific stimulus orientation cannot be consistently tracked. Consequently, performance will depend primarily on adaptation of higher-level factors, such as decision criterion. By contrast, if training exemplars span a smaller range, learning can be achieved by fine-tuning responses in the more reliably involved lower-level populations. The critical empirical prediction from RHT is that perceptual learning should differ in the level of generalizability, or transfer, which should be high when learning takes place under high stimulus variability (promoting learning at a more abstract, generalizable level) but low when stimulus variability is small (promoting learning via changes in specific low-level populations). Additionally, learning at the abstract, generalizable level requires the trained stimulus be task-relevant. In the present study, we applied these predictions to the context of the learning of feature conjunctions in visual search.

The present study

The present study applied the above logic of RHT to the process of feature integration, by testing whether stimulus variability and task relevance modulate binding-learning. Feature binding is assumed to involve the recruitment of higher-level sensory areas that integrate information from the lower-level ones. We can therefore derive from RHT the prediction that binding-learning under high stimulus variability should enhance the transfer of learning compared to low stimulus variability conditions. To test this hypothesis, we manipulated the variability of the relationship between individual stimulus features in a feature conjunction. In Experiment 1, participants performed an adapted version of the task in Yashar and

Carrasco (2016), involving the detection of a target defined by a color-orientation conjunction (e.g., a left-tilted red line). As in Yashar and Carrasco (2016), one color-orientation conjunction was pre-exposed as a distractor during the training phase, and participants were tested on the pre-exposed item in the subsequent test phase. Crucially, the variability associated with that feature conjunction was manipulated across two groups. For the low-variability group, the trial-by-trial orientation values for the pre-exposed item was sampled from a (Gaussian) distribution with low variance, and for the high-variability group, the orientation values were sampled from a distribution with high variance. After training, participants performed a detection task where the pre-exposed item (previously serving as a distractor) was now employed as the target. Importantly, variability was not manipulated in the test phase, and all lines were presented at a fixed clockwise (CW) or counter-clockwise (CCW) orientation. That is, in the test phase, the visual statistics of the search display were identical for both high and low-variability groups. If the target detection sensitivity (d') in the test phase were the same as the final block of the training phase, it would indicate transfer of learning. On the other hand, if the sensitivity dropped in the test phase as compared to the sensitivity in the final training block, it would indicate a lack of transfer (or, conversely, a high specificity of learning). Experiment 2 tested if the task relevance of a pre-exposed distractor during training determines the transfer of learning.

Experiment 1

Experiment 1 investigated the role of variability associated with the conjunction of features on binding-learning. In the initial training phase, participants searched for a target (line segment) defined by a conjunction of two features (color and orientation) among distractors. They performed four blocks of trials in this phase, where they reported whether the training target (e.g., a CW tilted red line) was present or absent in the search display. A training distractor that did not share any feature with the target was also displayed (pre-exposure). This pre-exposed item (e.g., CCW tilted green line) was then employed as the test target that participants searched for in the subsequent test phase. Participants were randomly assigned to one of two groups, where the training distractor had either high or low variability. Importantly, the test phase was identical across the two groups, and there was zero variability in the orientation of any item in the test phase search display. Based on the assumptions that feature binding requires the involvement of higher-level visual areas, and that perceptual learning in those areas is promoted by high stimulus variability, we predicted

that the high-variability group's (but not the low-variability group's) performance in the test phase would be comparable to their performance in the last block of the training phase, indicating a transfer of binding-learning. A reduction in performance in the test phase, by contrast, would imply specificity.

Method

Participants

Twenty-four students from the Indian Institute of Technology Gandhinagar participated in the experiment (mean age = 21.00, SD = 2.34, five females). The number of participants was preset to 24 (12 in each group). The sample size was determined through power analysis in R using the *pwr* package (Champely et al., 2018). Based on Yashar and Carrasco's (2016) data, our sample size would be appropriate to detect an effect size of 0.5 (Cohen's *d*) with 80 % power ($\alpha = 0.05$; SD = 0.55; one-sample *t* test). All participants were naive to the purpose of the study and reported normal or corrected to normal visual acuity and normal color vision. All participants were compensated for their time and provided written consent. All the experiments reported in the study followed guidelines and regulations approved by the Institutional Ethics Committee of the Indian Institute of Technology Gandhinagar (IEC/2018-19/04/MS/019).

Apparatus

Participants were seated in a dimly lit room in front of a 17-inch LCD monitor connected to a PC controlled by Windows 7. The monitor resolution was 1366×768 with a refresh rate of 60 Hz. The experimental tasks were coded and run in MATLAB 2013a (Mathworks Inc., Natick, MA) using PsychToolbox 3 (Brainard, 1997). Responses were collected via a standard keyboard.

Stimuli

A sample stimulus display is presented in Fig. 1. The search display consisted of 24 tilted line segments presented on a grey background (162 cd/m^2). Each line segment (0.8°) appeared inside the cells of an imaginary 5×5 matrix, except the center cell where a fixation-cross was presented. The lines were oriented either clockwise (CW; right-tilt) or counterclockwise (CCW; left-tilt) from vertical and were colored red (255, 0, 0) or green (0, 255, 0). There were four possible feature combinations; CW-red, CW-green, CCW-red, CCW-green. Prior evidence suggests that color-orientation conjunctions remain unbound even in the presence of crowding (Yashar, Wu, Chen, &

Carrasco, 2019). Out of the four feature combinations, one combination was randomly chosen as the training target (e.g., CW-red). The test target (that was pre-exposed during the training phase) did not share color or orientation with the training target (e.g., CCW-green). In the training phase, there were two types of distractors: (1) distractors that shared at least one feature (color or orientation) with the training target, and (2) distractors that did not share any feature with the training target, and they subsequently served as the test target (Fig. 2A).

Procedure

The task consisted of four training blocks of 150 trials (600 trials in total), followed by a test block of 150 trials. At the beginning of each block, the identity of the target (line segment) was presented to the participants. The identity of the target was fixed during the four training blocks. Participants were instructed to search for this target and report whether it was present or absent in the search display. The target was present in half of the trials. At the beginning of the test block, participants were informed of the change in the target. The (new) test target was the previous distractor with non-overlapping features, that is, it had a different color and orientation from the training target. For example, if the training target was CW-red, then the test target was CCW-green.

Each trial began with the presentation of the fixation-cross (+) at the center of the screen for 500 ms. The fixation was followed by the presentation of the search display (117 ms). After the offset of the search display, a blank screen was shown until the participant's response. Participants were instructed to press the *z* key if the target was present and the *m* key if the target was absent. Participants were told to respond as accurately as possible without a constraint on the speed of the response. Participants were also instructed to fixate their gaze on the central fixation-cross. A 500-ms error feedback tone followed an incorrect response. The experimental session lasted 40 minutes. Participants performed 100 practice trials with black and white lines at the beginning of the experimental session.

Design

Training phase Participants were randomly assigned to high/low-variability groups. The mean orientation of each line segment was 45° (CW or CCW). The target (if present) had an orientation of precisely 45° . Variability was applied only to the orientation of the distractors during the training blocks. The distractors that shared at least one feature with the target had identical variability (mean = 45° ; SD = 1°) across the two groups (high/low). The variability of the other distractor with non-overlapping features (the test

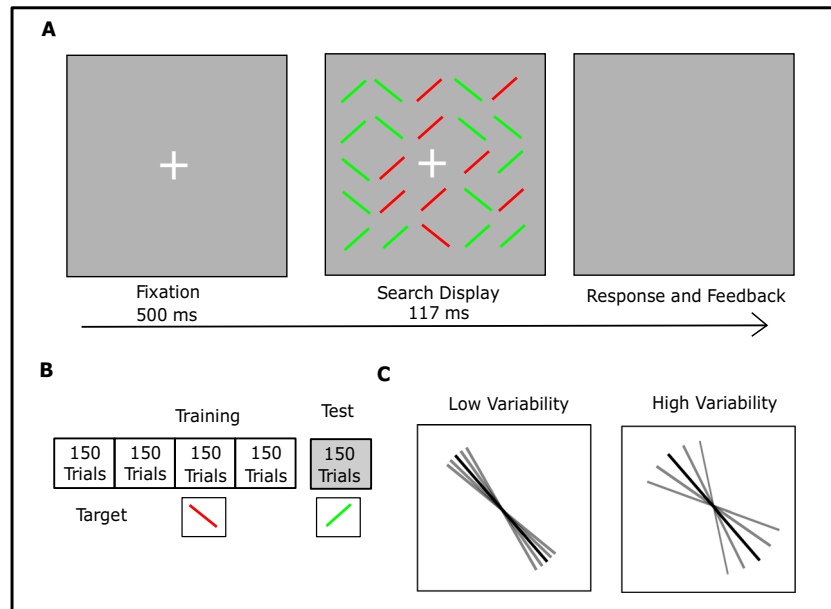


Fig. 1 Experimental Paradigm. **A** Trial structure. The search display consisted of red and green lines. Participants searched for a pre-defined target. They were instructed to press *m* key when the target is present and *z* key when the target is absent. The target was present in half of the total trials. A feedback tone followed an incorrect response. **B** Block structure Participants performed four training blocks and one test block. The identity of the training target and test target was

chosen randomly for each participant. The test target was pre-exposed as a distractor during the training. The variability of the test target was manipulated during the training. **C** Schematic of the variability manipulation. In low-variability condition, the variance of the distribution from which the orientation values were chosen was low ($SD = 1^\circ$). In high-variability condition, the distribution had high variance ($SD = 6^\circ$)

target) was manipulated across the two groups. For the low-variability group, the orientation of the test target was chosen from a Gaussian distribution with a mean of 45° and a standard deviation of 1° (Fig. 1C). For the high-variability group, the orientation of the test target was chosen from a Gaussian distribution with a mean of 45° and a standard deviation of 6° . Orientation was chosen as the variable dimension due to the convenience it provided in terms of allowing us to precisely manipulate and quantify statistical variability. In target-absent trials, eight items of each (distractor) feature-combination were presented (for a total of 24 items). In target-present trials, the target replaced a distractor other than the test target. The identities of search items (both target and distractors) were not tied to a specific location.

Test phase The test phase was identical for both high/low-variability groups. In the test block, all the items were presented precisely at 45° (CW or CCW).

Analysis

Perceptual sensitivity (d') and response criterion (c) were estimated for each block [$d' = z(\text{hit rate}) - z(\text{false-alarm rate})$, $C = -0.5[z(\text{hit rate}) + z(\text{false-alarm rate})]$]. To assess and contrast learning effects between groups, an ANOVA

was performed on these dependent measures with Group (high vs. low-variability) as a between-subjects factor and Block (1, 2, 3, 4, 5) as a within-subjects factor. To quantify possible transfer effects in each group, a transfer index (T) was estimated for each participant and tested against zero at the group level (Ahissar & Hochstein, 1997; Lu, Chu, Doshier, & Lee, 2005; Jeter, Doshier, Petrov, & Lu, 2009; Zhang et al., 2010; Hung & Seitz, 2014; Yashar & Carrasco, 2016).

$$T = 1 - \frac{(d'_4 - d'_5)}{(d'_4 - d'_1)}$$

where d'_4 denotes the sensitivity in the last training block and d'_1 denotes sensitivity in the first training block, d'_5 denotes the sensitivity in the test block. A T of 1 indicates complete transfer and T of 0 indicates complete specificity (no transfer at all).

Results

Sensitivity

The Group (high vs. low-variability) by Block (1, 2, 3, 4, 5) ANOVA revealed a significant main effect of Block, $F(4,88) = 13.10$, $p < 0.001$, $\eta_p^2 = 0.37$, as sensitivity

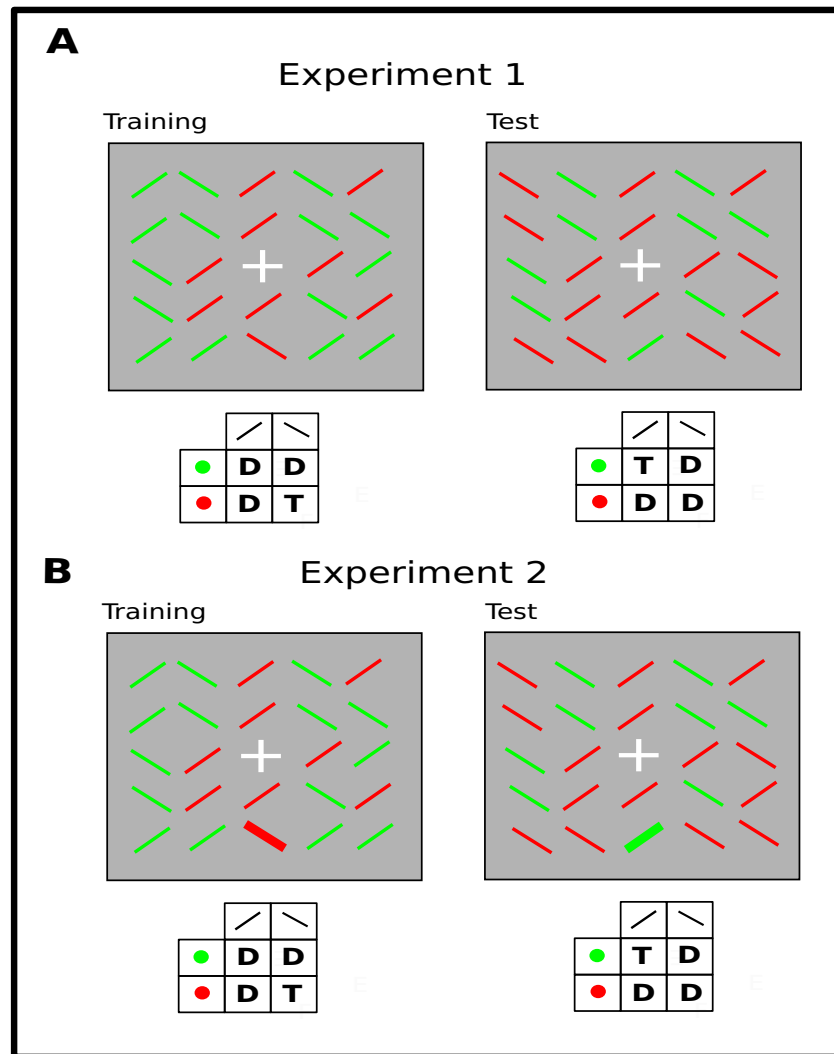


Fig. 2 Search display. Training and test search displays in Experiment 1 and Experiment 2. The target in the test block did not share a feature with the target in training block in both color and orientation dimension. D - Distractor, T - Target

improved significantly across the blocks (Fig. 3). There was a marginal interaction between Group and Block, $F(4,88) = 2.37$, $p = 0.058$, $\eta_p^2 = 0.10$. The sensitivity in the low-variability group was lower in the test block (Block 5; mean = 1.84) as compared to the sensitivity in Block 4 (mean = 2.30; $t(22) = 3.08$, $p = 0.01$, $r^2 = 0.54$), whereas there was no significant difference in sensitivity between Block 4 (mean = 2.03) and test block (mean = 1.91) in the high-variability group ($p = 0.49$). To test whether performance differed between the two groups in the training blocks, a 2 (Group: high vs. low-variability) \times 4 (Block : 1, 2, 3, 4) ANOVA was performed. There was a main effect of Block, $F(3, 66) = 21.76$, $p < 0.001$, $\eta_p^2 = 0.50$, but no significant main effect of Group ($p = 0.14$) or Group by Block interaction was

found ($p = 0.28$). A linear fit between training blocks and sensitivity showed significant change in sensitivity across the four training blocks for both high-variability, $F(1,46) = 12.12$, $p = 0.001$, $r^2 = 0.21$ and low-variability groups $F(1,46) = 5.95$, $p = 0.018$, $r^2 = 0.11$ (Fig. 6A).

Transfer index

To test the significance of the transfer of learning in each group, the transfer index (T) was estimated for each participant. Separate one-sample t tests were performed for both high and high-variability groups to test the respective hypotheses of transfer taking place or not in each group (see Yashar & Carrasco, 2016). A significant difference

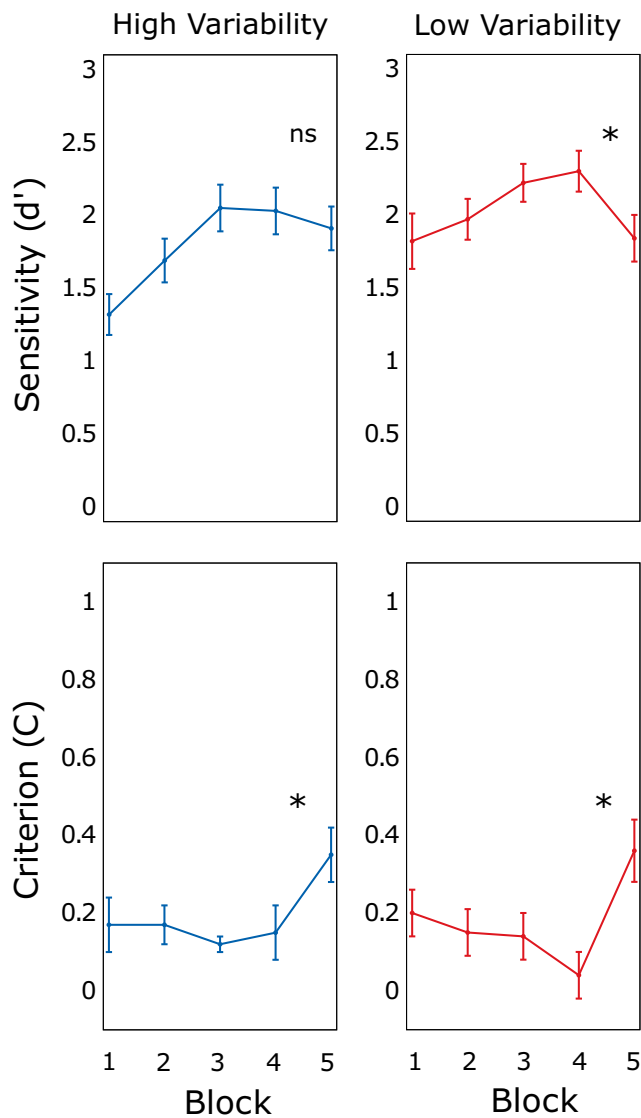


Fig. 3 Results of Experiment 1: Sensitivity and criterion as a function of blocks. Error bars show standard error. * $p < 0.05$

from zero would indicate transfer. Significant transfer was observed in the high-variability group, $t(11) = 3.80, p = 0.002$. No transfer was observed in low-variability group¹, $p = 0.34$ (Fig. 4).

¹Higher standard deviation was observed for transfer index data in the low-variability group and the mean of this group could likely be distorted by outliers. A re-analysis on the low-variability group data was performed after removing any transfer index exceeding 2.5 SDs from the mean (mean = -1.52; SD = 5.37; range = -14.95 to 11.90). One participant with a transfer index of -17.64 was removed based on this criterion. As in the main analysis reported in the Results section, a one-sample t test was performed on this data to test if the transfer index significantly differed from 0. Consistent with the observation in the main analysis, transfer index did not significantly differ from zero (mean = 0.062; $p = 0.83$).

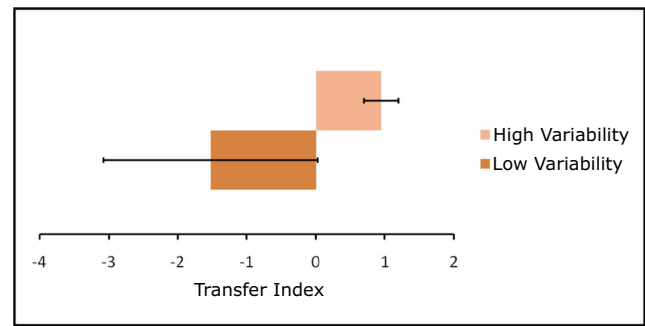


Fig. 4 Results of Experiment 1. Transfer index for high versus low-variability conditions

Criterion

The Group (high vs. low-variability) by Block (1, 2, 3, 4, 5) ANOVA revealed a significant main effect of Block, $F(4,88) = 6.77, p < 0.001, \eta_p^2 = 0.24$. No significant main effect of Group ($p = 0.83$) or Group by Block interaction ($p = 0.65$) was observed. The criterion was highest in the test blocks in both groups. No significant correlation was found between criterion and d' , $r(23) = 0.26, p = 0.22$, thus ruling out an effect of criterion on the observed d' .

Reaction time

To test if the observed effects on sensitivity were due to speed-accuracy trade-off, a Group (high vs. low-variability) by Block (1, 2, 3, 4, 5) ANOVA was performed on mean RT. Training reduced RT in both groups, as indicated by a significant main effect of Block, $F(3,66) = 16.48, < 0.001, \eta_p^2 = 0.43$. No significant main effect of Group ($p = 0.36$) or Group by Block interaction ($p = 0.91$) was observed.

Discussion

Experiment 1 investigated whether the variability of a pre-exposed color-orientation conjunction would influence feature-binding-learning. We found that the transfer of binding-learning was in fact dependent on the variability of the pre-exposed conjunction, thus supporting the notion that binding-learning is guided by the principles of perceptual learning proposed in RHT. Specially, the results suggest that binding-learning is associated with changes in higher-level areas involved in abstract representation of the stimulus, and learning is generalized when the participants are trained in a broad range of stimuli, but not when they are trained over a narrow range. Participants in both (high and low-variability) groups showed significant and comparable learning in the training phase (improved search performance

over blocks 1–4). In other words, the manipulation of variability did not influence performance in the training phase. The transfer of learning to the test target (block 5), however, was observed only in the high-variability group. Although overall sensitivity was marginally higher in the high-variability group, neither the main effect of Group (variability) nor the Group by Block interaction reached statistical significance. There was a significant change in criterion when moving from the training phase to the test phase for both high and high-variability groups. In the test phase, participants were more biased towards responding 'target absent'. Importantly, the variability manipulation had no effect on the criterion values. No significant correlation between d' and c was observed, suggesting that the observed changes in sensitivity were not dependent on the response strategies.

Although the pre-exposed distractor item was task-relevant in the sense that its feature conjunction had to be processed for identifying the search target, the brief presentation of the search array would have prevented the deployment of goal-driven attention (Carrasco, 2011). Any potential effect of attentional guidance after the offset of the search display (for instance, by an afterimage) is also unlikely to be driving changes in sensitivity, as effects of post-stimulus cuing have been found to be restricted to modulating response-bias (Carrasco, 2011). Thus, the observed transfer effects under high variability are unlikely to be attributable to goal-driven attention. Instead, the feature-binding-learning observed in the present experiment is likely to be driven by incidental learning. The incidental learning of the statistical structure of stimuli has been found to influence target detection (Jiang, Swallow, & Rosenbaum, 2013; Wang & Theeuwes, 2018a; Wang & Theeuwes, 2018b) and is proposed have an effect on the attentional priority map that is qualitatively different from traditional goal-driven attention (reviewed in Awh et al., 2012). For instance, Jiang, Swallow, and Rosenbaum (2013) showed that participants could incidentally extract the probability of a target being presented at a specific location and then use this information to guide attention to that location. This form of attentional biasing is referred to as a "selection history effect", reflecting a lingering bias to the statistical regularities in the stimulus environment (Awh, Belopolsky, & Theeuwes, 2012). However, studies that demonstrated selection history effects on attentional priority have not characterized the precise time-course of such effects. It is therefore unclear whether attentional biasing due to incidental learning can be operational in a briefly presented search display as in Experiment 1. In Experiment 2 we set out to test whether task relevance, a critical factor that determine perceptual learning (Ahissar & Hochstein, 1993), is necessary for binding-learning.

Experiment 2

The results of Experiment 1 showed that binding-learning in a serial search task is modulated by the variability of the pre-exposed conjunction. Even though the feature conjunction is pre-exposed as a distractor during training in Experiment 1, the serial nature of the search task renders both the target and the distractor conjunctions relevant to the task. That is, the discrimination of the target from distractors requires an initial binding of orientation and color features for all items in the search display, and in this sense the distractors are relevant to the task. Accordingly, Yashar and Carrasco (2016) proposed that binding-learning in their protocol reflects "task-based learning", and it has been suggested that the transfer of learning is expedited by the behavioral relevance of the trained stimuli (Ahissar & Hochstein, 1993). However, perceptual learning effects can also be found when the trained stimuli are entirely irrelevant to the task. For example, when participants were exposed to a task-irrelevant, sub-threshold motion signal while performing a letter identification task, their performance in a subsequent motion discrimination task improved (Watanabe, Náñez, & Sasaki, 2001). Contrary to RHT, studies that report task-irrelevant learning have led to the proposal that both task-relevant learning and task-irrelevant learning share common mechanisms (reviewed in Seitz & Watanabe, 2005). The two-stage model (Shibata, Sagi, & Watanabe, 2014) of perceptual learning offered a reconciliation to these divergent views by proposing that task-relevant learning involves changes to higher-level areas and task-irrelevant learning reflects changes to lower-level, feature-specific areas. Thus, in this view, learning associated with higher-level, "cognitive" areas would require the trained stimulus to be task relevant. Accordingly, the pre-exposure of task-irrelevant stimuli would be expected to result in transfer of learning if the stimuli were defined by a single feature, as there would be no need to bind distractor features for discriminating the target (e.g., Watanabe et al., 2001). On the other hand, if the stimuli were defined by a conjunction of features, transfer should happen only when the pre-exposed stimuli are task relevant (e.g., Yashar & Carrasco, 2016).

To test whether it is necessary for the pre-exposed distractor to be task relevant (in the sense of its conjunctive features having to be identified to discriminate the target) for this effect to be observed, Experiment 2 had participants perform the search task in parallel search mode. In this case, the search is guided by bottom-up activations due to the perceptual salience of the "pop-out" search target, and it is not necessary to process the conjunctive features of the distractors to locate the target (Theeuwes, 1994; Yantis & Jonides, 1984). The effect of bottom-up activation

peaks at an earlier time-window (~ 120 ms after stimulus onset; Dombrowe et al., 2010) than goal-driven activation (which takes ~ 190 ms to peak; Hickey et al., 2010). Exploiting this difference in time-course, in Experiment 2, we asked whether binding-learning is transferred when the search process is driven bottom-up, that is, by a pop-out target. The experimental design was the same as in Experiment 1, except that a pop-out feature (size) was added to one of the items in the search array. That is, one item in the search display was always presented in a larger size. The pop-out feature coincided with the target in all target-present trials and was the “non-pre-exposed” distractor in target-absent trials, meaning that the pop-out feature never coincided with the pre-exposed item throughout the training phase. Thus, participants could successfully perform the task by selectively focusing on the pop-out feature alone, through bottom-up activation. This renders the pre-exposed distractor fully task-irrelevant. We hypothesized that, if binding-learning were independent of task relevance, then the results observed in Experiment 1 – transfer effects only in the high-variability group – would be replicated. On the other hand, if binding-learning reflected task-based plasticity, as proposed by Yashar and Carrasco (2016), then no transfer of learning would occur in either group.

Method

Participants

Twenty-four students from the Indian Institute of Technology Gandhinagar participated in the experiment (mean age = 21.41, SD = 2.99, three females). There was no overlap in participants between Experiments 1 and 2.

Stimuli, procedure & design

The stimuli, procedure, and design were the same as Experiment 1, except that one of the line segments in the search array was thicker (the pop-out target). The width of the pop-out line was 1.5 times the width of other lines in the display. In target-present trials, the pop-out line was always the target. In target-absent trials, the pop-out line was a distractor other than the test target.

Results

Sensitivity

The Group (high vs. low-variability) by Block (1, 2, 3, 4, 5) ANOVA revealed a significant main effect of Block,

$F(4,88) = 15.14, p < 0.001, \eta_p^2 = 0.41$. Sensitivity improved significantly across the blocks (Fig. 5). There was no significant main effect of Group, $p = 0.59$. The Group by Block interaction was also not significant, $p = 0.84$. A significant reduction in sensitivity was observed in the test block as compared to Block 4 in both the low-variability ($t(11) = 3.70, p = 0.003, r^2 = 0.68$) and the high-variability ($t(11) = 4.35, p = 0.001, r^2 = 0.81$) group. To test whether variability affected search performance in the training blocks, a 2 (high vs. low-variability) by 4 (Blocks: 1, 2, 3, 4) ANOVA was performed. There was a main effect of Block, $F(3,66) = 10.18, p < 0.001, \eta_p^2 = 0.32$. No significant main effect of Group ($p = 0.52$) or Group by Block interaction was observed ($p = 0.90$). However, the linear fit did not reveal a significant change in sensitivity across four training blocks in both low ($p = 0.06$) and high-variability ($p = 0.15$) groups (Fig. 6B).

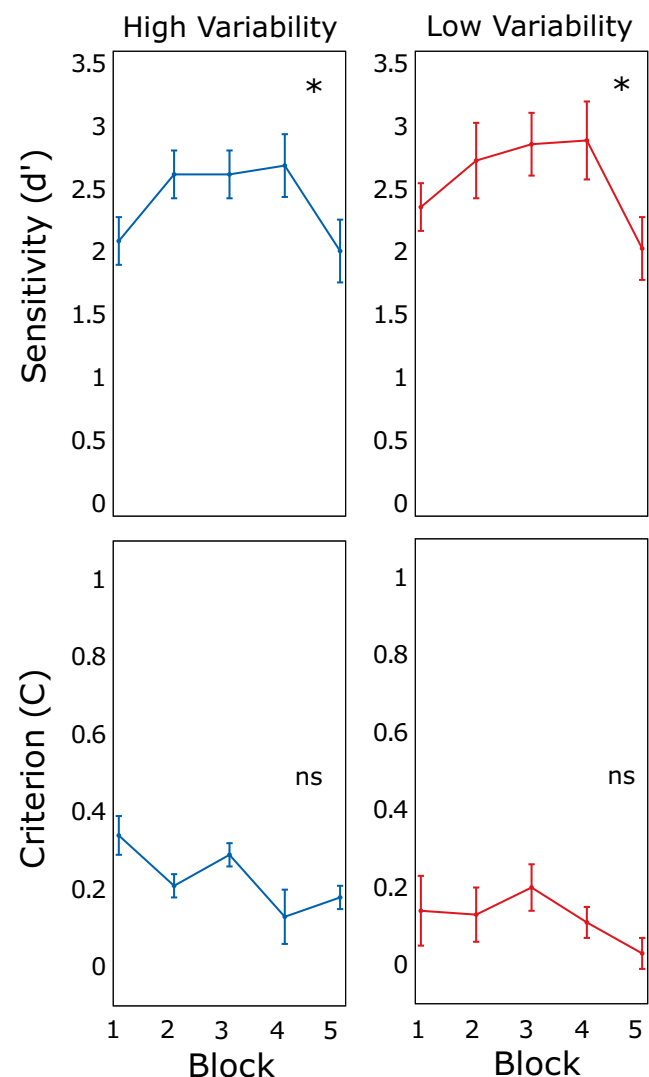


Fig. 5 Results of Experiment 2. Sensitivity and criterion as function of blocks. Error bars show standard error. $*p < 0.05$

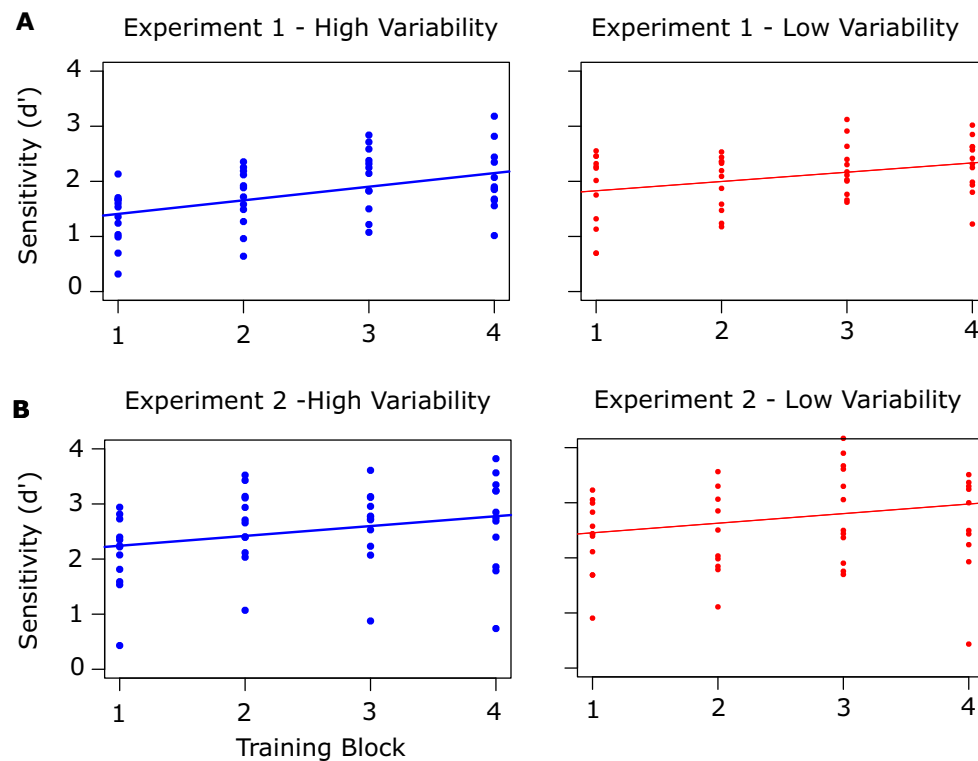


Fig. 6 Regression fit. Linear regression fits between training block and sensitivity for high (blue) and low (red) variability groups. Dots indicate sensitivity of each participant

Criterion

The Group (high vs. low-variability) by Block (1, 2, 3, 4, 5) ANOVA revealed a significant main effect of variability, $F(1,22) = 4.77$, $p = 0.04$, $\eta_p^2 = 0.18$, suggesting that participants in the low-variability group were more liberal in reporting a target to be present compared to the high-variability group. There was also a significant main effect of Block $F(4,88) = 2.74$, $p = 0.03$, $\eta_p^2 = 0.11$ reflecting a general shift to a more liberal criterion as the task progressed. The Group by Block interaction was not significant ($p = 0.51$). No significant correlation was found between criterion and d' , $r(23) = 0.35$, $p = 0.09$ (Fig. 5).

Reaction time

An ANOVA on RT showed that training reduced RT in both groups, as reflected in a main effect of Block, $F(4,88) = 15.93$, $p < 0.001$, $\eta_p^2 = 0.42$. No significant main effect of Group ($p = 0.65$) or Group by Block interaction ($p = 0.54$) was observed.

Transfer index

No significant transfer was observed in either the high-variability ($p = 0.49$) or the low-variability group ($p = 0.49$).

Discussion

Experiment 2 employed a pop-out (parallel) search design to probe whether binding-learning, and its modulation by stimulus variability, could be obtained when the pre-exposed test target is task-irrelevant, that is, when its conjunctive features do not have to be processed at all to discriminate the search target. As in Experiment 1, search performance in the training phase improved across blocks and was not modulated by the stimulus variability manipulation. Crucially, in contrast to Experiment 1, performance in the test phase was significantly lower than the final block of the training phase for both high and high-variability groups. In other words, no transfer was observed in either group. Thus, in a pop-out search, the pre-exposure of the test target does not influence the performance in the test phase, regardless of variability. This result suggests that binding-learning does not occur under conditions where the pre-exposed conjunction is task-irrelevant, and thus presumably unattended. This finding supports the claim that binding-learning reflects task-based plasticity and corroborates the distinction made between task-based versus feature-based plasticity in the two-stage theory of perceptual learning (Shibata, Sagi, & Watanabe, 2014). Specifically, the trained stimuli have to be task relevant for rapid perceptual learning of their feature conjunctions. In

sum, the incidental learning of feature conjunctions requires both task relevance and high stimulus variability, and is absent when the pre-exposed conjunction is task-irrelevant or exposed with little variability.

The failure to observe transfer of learning for task-irrelevant conjunction could also be explained by the difference in the focus of attention between Experiments 1 and 2. In Experiment 1, some attention had to be paid to the distractors in order to identify the target, whereas in Experiment 2, attention was likely exclusively allocated to the pop-out item (target or non-pre-exposed distractor). This weaker attentional exposure of the pre-exposed distractors in Experiment 2 might have contributed to reducing the efficiency of binding during training, as feature binding has been shown to be more efficient when the focus of attention is limited to a single location as opposed to multiple locations (Dowd & Golomb, 2019). However, explicit cueing of attention (as in Dowd & Golomb, 2019) and implicit learning of statistical regularity could have distinct effects on perception. For instance, the transfer of attentional bias to a secondary task and susceptibility to working memory interference are observed for explicitly cued attention but not for implicitly cued attention (Salovich, Remington, & Jiang, 2018; Vickery, Sussman, & Jiang, 2010). Further investigation into the independent or interactive effects of the explicit spatial cueing of attention and learning of statistical regularity is necessary to build an integrative account of feature binding.

The findings of Experiment 2 also suggest that the (non-significant) difference in training phase performance between high and low variability groups is not influencing the pattern of transfer. In both Experiments 1 and 2, there was a numerical performance difference between the groups (high versus low variability) during training, specifically a numerically lower performance for the high variability group compared to low variability group (Experiment 1: mean difference = 0.30; Experiment 2: mean difference = 0.20). In spite of this data pattern, Experiment 1 showed a significant transfer effect in the high-variability group whereas Experiment 2 did not. This suggests that the non-significant differences in training performance are not modulating transfer of learning.

General discussion

Prior research has shown transfer of learning for briefly presented feature conjunctions after a 40-min training session (binding-learning; Yashar & Carrasco, 2016). This training-induced enhancement in feature binding is proposed to be independent of top-down attentional control. First, we tested if stimulus variability, a key determinant of learning, modulates feature binding. Studies on diverse phenomena such

as motor skill learning (Schmidt, 1975; Schmidt & Bjork, 1992) and language acquisition (Banai & Amitay, 2015) have shown that high-variability training enhances learning. Consistent with these findings, we found transfer of learning when the trained feature conjunction had high-variability, whereas learning did not transfer when the trained feature conjunction had low-variability (Experiment 1). Second, we delineated the boundary conditions of binding-learning by showing that a certain degree of task relevance of the trained stimulus is necessary for binding-learning (Experiment 2): transfer occurs when distractors have to be identified to isolate the target (in conjunction search, Experiment 1), but not when the target pops out (Experiment 2), such that distractor features can be fully ignored.

Previous studies supporting attention-independence of feature binding have primarily employed categorization of naturalistic scenes and frequently encountered stimuli, such as faces and vehicles (Reddy, Wilken, & Koch, 2004; Reddy, Reddy, & Koch, 2006; Li, VanRullen, Koch, & Perona, 2002). A major criticism of this line of evidence is that the categorization task on familiar objects/scenes opens up the possibility that participants could perform the categorization by detecting disjunctive sets of features without complete identification (or binding). For instance, participants can detect a tiger by detecting the stripes, which are unique to tigers. Evans and Treisman (2005) tested this possibility by hypothesizing that, if detection of an animal image in a briefly flashed stream of images (RSVP) is driven by the detection of disjunctive sets of animal features, then the presence of a human image in the stream will make the task difficult, as humans and animals share features. The authors reported that a distracting presence of a human image reduced the detection accuracy, suggesting that the rapid categorization of the familiar object is driven by the detection of disjunctive features and not the enhancement of feature conjunction (Evans & Treisman, 2005). However, detecting a disjunctive feature is not sufficient for target detection in the present study, as most distractors in the search display shared a feature (color or orientation) with the target. Instead, the successful detection of the target in our task requires the binding of both features. Thus, the change in detection sensitivity in the present study does not reflect improved detection of a disjunctive feature, but an improvement in the conjunctive representation.

Perceptual learning mechanism underlie feature binding

The hypothesis that both stimulus variability and task relevance are critical to visual perceptual learning was drawn from the reverse hierarchy theory (RHT). Regarding stimulus variability, RHT would predict that high-variability

training enhances perceptual learning in tasks that recruit higher-level, association areas (Clopper & Pisoni, 2004), whereas low-variability training would improve learning when tasks recruit lower-level, feature specific areas (Amitay, Hawkey, & Moore, 2005). Our prediction that binding-learning would be expedited by high-variability training was borne out by reports that conjunctive codes are represented in the higher-level areas in the temporal cortex (e.g., Cowell et al., 2017). Findings from Experiment 1 corroborate this RHT prediction that high-variability training expedites learning, and demonstrate the modulatory effect of stimulus variability on binding-learning. With respect to task relevance, there is a lack of consensus in the literature on whether this is a necessary condition for transfer of (perceptual) learning (see Seitz & Watanabe, 2009, also see Huang et al., 2007). Shibata, Sagi, and Watanabe (2014) attempted to reconcile conflicting proposals about the influence of task relevance in perceptual learning by proposing that learning that involves changes to higher-level areas necessitates task relevance of the trained stimulus, whereas learning that involves low-level areas can progress even when the stimulus is task-irrelevant. Critically, this model would predict that feature binding, which recruits higher-order areas, would necessitate task relevance. Consistent with this proposed dissociation, we show that task relevance of the trained stimuli is necessary for binding-learning. Taken together, our findings corroborate the idea that feature binding underlies integration of features into object representation through the principles of perceptual learning.

Prior studies that demonstrated “ultra-rapid” object categorization in briefly presented stimuli have suggested that attention-independent detection is limited to naturalistic images and familiar real-world objects, and the detection of arbitrary stimuli is impaired under attentionally-demanding dual-task conditions (Li et al., 2002). However, we observed an increase in visual sensitivity to arbitrary (color-orientation) conjunctions over only four blocks of training; thus, learning emerged within a relatively short training session (30 min). Regression analysis showed that this increase in sensitivity was significant in Experiment 1 and was observed in both low and low-variability groups. In Experiment 2, where the target was always defined by a pop-out feature, the improvement in sensitivity during training did not reach significance. However, this is likely due to the fact that search performance reached asymptotic levels very early into the task. Taken together, these results suggest that even arbitrary conjunctions, through sufficient exposure, can be bound without strongly focused attention. This learning-based improvement, however, is not independent of the focus of attention - which enhances feature binding in novel contexts. Furthermore, the failure to observe binding learning in Experiment 2, where there was

a relatively weaker focus of attention on the pre-exposed distractors, suggests that spatial focus of attention interacts with binding learning.

There are two possible alternative explanations for the improvement in sensitivity we observe. First, the improvement could reflect task familiarity or generic practice effects; second, the perceptual learning could be feature-specific rather than binding learning. With respect to general task practice, we believe that this is unlikely to have driven results, because participants performed 100 practice trials where the search items (lines) were presented in black and white. We would thus expect non-feature-specific performance improvement due to practice at line orientation judgments to have saturated by the end of this practice session. The second possibility of whether performance improvements in the present paradigm reflect feature-specific learning was addressed in Yashar and Carrasco (2016). These authors hypothesized that, if learning were feature-specific, then transfer of learning should be observed both when the pre-exposed item coincided with the test target in a single dimension (color or orientation) or both dimension (color and orientation). Alternatively, if learning were conjunction-specific, transfer of learning should be observed only when the pre-exposed item coincided with the test target in both dimension. That study observed transfer of learning only when the pre-exposed item coincided with test target in both dimension and not when it coincided with single feature, corroborating the hypothesis that learning was conjunction-specific (binding learning). In sum, findings of present study and Yashar and Carrasco (2016) study suggests that learning observed in the present paradigm reflects binding learning and not general performance improvement or feature-specific perceptual learning.

The proposal that perceptual learning mechanisms underlie feature integration is consistent with the findings of Humphreys and colleagues (Anderson & Humphreys, 2015; Rappaport, Humphreys, & Riddoch, 2013; Rappaport, Riddoch, Chechlacz, & Humphreys, 2016; Wildegger, Riddoch, & Humphreys, 2015), who in a series of studies documented familiarity-based facilitation in feature binding. These authors showed that detection of feature conjunctions that are more likely to be encountered in real life is faster and more efficient than less probable conjunctions. For instance, object identification is improved when it is presented in its diagnostic color (e.g., yellow corn versus purple corn). Based on these findings, it was proposed that the facilitatory effect of familiarity on feature binding is driven by experience-dependent Bayesian integration of constituent features into conjunctive codes. The present study extends this idea by showing that two key factors that are known to drive perceptual learning, stimulus variability and task relevance, modulate the learning-based improvement in feature binding.

Binding-learning reflects unitization, not re-entrant consolidation

There are two contrasting hypotheses on the mechanism underlying object learning. One view is that training leads to the formation of conjunction/category-specific functional units that are independent of top-down, re-entrant processing (unitization hypothesis; Goldstone, 1998). According to this view, the training-induced plasticity is restricted to category-specific areas, such as the Fusiform Face Area (FFA), and the “non-specific” parietal association cortex does not affect post-training performance. In line with this view, it has been shown that disruption (by TMS) of parietal cortex impairs performance in a conjunction search only when applied during training, and not once perceptual learning had taken place (Lobley & Walsh, 1998). This suggests that although parietal cortex, implicated in top-down attentional feature weighting (Egner et al., 2008), is necessary during the training phase, post-training performance is determined by the training-induced plasticity at category-specific areas. A recent neuroimaging study also showed that familiar conjunctions are unitized in the posterior ventral visual stream (Liang, Erez, Zhang, Cusack, & Barense, 2020).

The alternative view is that training-induced plasticity is specific to early, feature-specific areas, and that the learning-based improvement in a conjunction search task reflects the consolidation by top-down re-entrant processing through the mechanism of feature-based attention (attentional-enhancement hypothesis; Andersen et al., 2008). Su et al. (2014) attempted to pit the unitization account against the attentional-enhancement account by hypothesizing that, if perceptual learning in conjunction search tasks reflects feature-based attentional enhancement, then learning should transfer to a target that shares at least one feature from the trained conjunction (e.g., training target—right-tilted red line; test target—left-tilted red line). On the other hand, if unitization underlies training-induced plasticity, then the learning should not transfer if the target shares only one feature with the trained conjunction, as unitization is specific to the trained conjunction. Consistent with the attentional-enhancement hypothesis, Su et al. (2014) observed transfer only when the target in the test phase matched the training target in at least one dimension (color or orientation), but not when both the features changed. Accordingly, the authors argued that perceptual learning in a conjunction search leads to feature-based attentional enhancement of a specific dimension (color, orientation, shape) rather than of the conjunction/object.

This contrasts with the findings in Experiment 1 (as well as those of Yashar and Carrasco, 2016), where transfer was observed when both features of the conjunction were switched. However, there are two simple but critical differences between Su et al. (2014) and our design. First,

we pre-exposed the test target (as a distractor) during training, whereas Su et al. (2014) did not pre-expose the test target. Second, the search display was presented for 300 ms in Su et al. (2014)’s experiment, which would have allowed feature-based attention to influence performance, whereas search displays in our experiment lasted only 117 ms, thus precluding top-down shifts of attention. Our results suggest that learning can transfer even when both features of the conjunction are changed, as long as the test target is pre-exposed as a distractor, and the search display is short enough to restrict feature-based attentional processes. The difference between the present findings and Su et al. (2014)’s findings can thus be reconciled by a dual-process account of feature binding. That is, the feature-specific transfer observed by Su et al. (2014) could be determined by later-stage binding, as the duration of the search display enabled top-down, re-entrant processing to influence performance. On the other hand, our results are likely determined by an earlier stage that is not dependent on re-entrant processing. Additionally, we also would not expect binding-learning to transfer if the search display is presented for more than 200 ms, as under those conditions the attentional mechanism of distractor suppression might hinder the formation of the conjunctive unit (Andersen & Müller, 2010).

We interpret the improvement in target detection sensitivity in our study to reflect the formation of conjunctive coding that facilitates detection. Furthermore, this fast binding of features is independent of top-down attentional control and could underlie segmentation and integration of features during the initial feed forward sweep (VanRullen, 2007). However, findings from Experiment 2 seem to contradict the claim that binding-learning in brief stimulus displays is independent of top-down attentional processing.

Goal-driven versus habitual attention

The failure to observe transfer of learning when the pre-exposed conjunction was task-irrelevant (i.e., in pop-out search, Experiment 2) could suggest that some degree of top-down (re-entrant) processing is influencing feature binding even during very brief stimulus presentation. This possibility stems from the traditional two-stage conception of information processing, where re-entrant processing is characterized in terms of the single-trial time-course of stimulus processing (Braet & Humphreys, 2009). That is, at longer stimulus durations, re-entrant attentional control is assumed to influence perception, whereas at shorter stimulus duration, it is not. Similar arguments have also been put forward in favor of the dissociation between bottom-up and top-down attentional control (for a review, see Carrasco, 2011). According to this view, re-entrant processing is characterized within the narrow definition of

goal-driven attention. On the other hand, in the perceptual learning literature, re-entrant processing is defined in terms of the task relevance of the trained stimulus (Ahissar & Hochstein, 1993; for an opposing view, see Seitz & Watanabe, 2009), which is mechanistically different from goal-driven attention (Paffen, Gayet, Heilbron, & Van der Stigchel, 2018). Our finding that task relevance is necessary for binding-learning in very brief search displays suggests that attention-dependence cannot be inferred based on differences in the time-course of stimulus presentation. This suggests that a dichotomous classification of feature binding processes into hardwired and on-demand processing as proposed in the dual-process account (VanRullen, 2009, also see Humphreys, 2001) is untenable. However, some recent reports (see below) have provided an alternative account of attention based on studies that show statistical learning effects on attentional tuning.

The alternative characterization of attention, referred to as habitual attention (see Jiang, 2018) or as reflecting effects of “selection history” (Awh, Belopolsky, & Theeuwes, 2012), attempts to move our conceptualization of attention beyond traditional dichotomies (Awh et al., 2012; Jiang, 2018; see also Egner, 2014), such as the one proposed in the dual-process account of binding. Habitual attention is driven by the incidental learning of the probabilistic associations in the environment. Evidence supporting habitual attention have primarily been restricted to mechanisms such distractor suppression (Wang & Theeuwes, 2018a; 2018b), target activation (Geng & Behrmann, 2006) and covert orienting (Jiang, Swallow, & Rosenbaum, 2013). In the present study, we demonstrate that altering the statistical relationship between dimensions influences feature binding—a phenomenon that is traditionally thought to be functionally achieved by focused attention (Treisman, 1998).

Conclusions

In sum, the present study aimed to ground feature binding in the principles of perceptual learning. We demonstrate that feature binding, traditionally thought to require focused attention, can happen with relatively little attention and improve with experience. The formation of this type of habitual feature binding is mediated by a variability-dependent learning mechanism that forms higher-order conjunctive representations, which subsequently aid object detection. Our findings support the claim that some form of feature binding happens during the initial feedforward sweep of sensory processing and is independent of re-entrant processing (VanRullen, 2007). Furthermore, variability-dependent learning might also underlie the perceptual dominance enjoyed by familiar categories such as faces and vehicles. Our finding that feature binding is

determined by principles of perceptual learning calls for future inquiries on how other traditional effects of attention are determined by principles of perceptual learning.

Acknowledgements We thank Meera Mary Sunny for helpful comments on an earlier version of this draft. This research was supported by IIT-GN ORES awarded to NG. We declare no conflicts of interest.

Open practices statement The reported experiments were not preregistered. The data have been uploaded to a repository (https://osf.io/tz8g5/?view_only=ad2fb4abf1d14456a6565f76ee780cfcf).

References

- Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences*, *90*(12), 5718–5722.
- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401.
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse Hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 285–299.
- Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception & Psychophysics*, *67*(4), 691–698.
- Andersen, S., Hillyard, S. A., & Müller, M. M. (2008). Attention facilitates multiple stimulus features in parallel in human visual cortex. *Current Biology*, *18*(13), 1006–1009.
- Andersen, S., & Müller, M. (2010). Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. *Proceedings of the National Academy of Sciences*, *107*(31), 13878–13882.
- Anderson, G. M., & Humphreys, G. W. (2015). Top-down expectancy versus bottom-up guidance in search for known color-form conjunctions. *Attention, Perception, & Psychophysics*, *77*(8), 2622–2639.
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443.
- Banai, K., & Amitay, S. (2015). The effects of stimulus variability on the perceptual learning of speech and non-speech stimuli. *PLoS One*, *10*(2), e0118465.
- Braet, W., & Humphreys, G. W. (2009). The role of reentrant processes in feature binding: Evidence from neuropsychology and TMS on late onset illusory conjunctions. *Visual Cognition*, *17*(1-2), 25–47.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525.
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., & et al (2018). Package ‘pwr’. R package version, 1–2.
- Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*(3), 207–238.
- Cowell, R. A., Leger, K. R., & Serences, J. T. (2017). Feature-coding transitions to conjunction-coding with progression through human visual cortex. *Journal of Neurophysiology*, *118*(6), 3194–3214.
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms face detection in just 100 ms. *Journal of Vision*, *10*(4), 16–16.
- Dombrowe, I. C., Olivers, C. N., & Donk, M. (2010). The time course of color-and luminance-based salience effects. *Frontiers in Psychology*, *1*, 189.

- Dowd, E. W., & Golomb, J. D. (2019). Object-feature binding survives dynamic shifts of spatial attention. *Psychological Science*, *30*(3), 343–361.
- Eglin, M., Robertson, L. C., & Knight, R. T. (1989). Visual search performance in the neglect syndrome. *Journal of Cognitive Neuroscience*, *1*(4), 372–385.
- Egner, T. (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, *5*, 1247.
- Egner, T., Monti, J. M., Trittschuh, E. H., Wieneke, C. A., Hirsch, J., & Mesulam, M. M. (2008). Neural integration of top-down spatial and feature-based information in visual search. *Journal of Neuroscience*, *28*(24), 6141–6151.
- Esterman, M. (2000). Preattentive and attentive visual search in individuals with hemispatial neglect. *Neuropsychology*, *14*(4), 599.
- Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free?. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1476.
- Frank, S. M., Reavis, E. A., Tse, P. U., & Greenlee, M. W. (2014). Neural mechanisms of feature conjunction learning: enduring changes in occipital cortex after a week of training. *Human Brain Mapping*, *35*(4), 1201–1211.
- Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology*, *14*(7), 573–578.
- Geng, J. J., & Behrmann, M. (2006). Competition between simultaneous stimuli modulated by location probability in hemispatial neglect. *Neuropsychologia*, *44*(7), 1050–1060.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612.
- Hickey, C., van Zoest, W., & Theeuwes, J. (2010). The time course of exogenous and endogenous control of covert attention. *Experimental Brain Research*, *201*(4), 789–796.
- Huang, X., Lu, H., Tjan, B. S., Zhou, Y., & Liu, Z. (2007). Motion perceptual learning: when only task-relevant information is learned. *Journal of Vision*, *7*(10), 14–14.
- Humphreys, G. W. (2001). A multi-stage account of binding in vision: Neuropsychological evidence. *Visual Cognition*, *8*(3-5), 381–410.
- Hung, S. C., & Seitz, A. (2014). Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *Journal of Neuroscience*, *34*(25), 8423–8431.
- Jeter, P. E., Doshier, B. A., Petrov, A., & Lu, Z. L. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision*, *9*(3), 1–1.
- Jiang (2018). Habitual versus goal-driven attention. *Cortex*, *102*, 107–120.
- Jiang, Swallow, K. M., & Rosenbaum, G. M. (2013). Guidance of spatial attention by incidental learning and endogenous cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 285.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601.
- Liang, J. C., Erez, J., Zhang, F., Cusack, R., & Barense, M. D. (2020). Experience transforms conjunctive object representations: Neural evidence for unitization after visual expertise. *Cerebral Cortex*, *30*(5), 2721–2739.
- Lobley, K., & Walsh, V. (1998). Perceptual learning in visual conjunction search. *Perception*, *27*(10), 1245–1255.
- Lu, Z. L., Chu, W., Doshier, B. A., & Lee, S. (2005). Independent perceptual learning in monocular and binocular motion systems. *Proceedings of the National Academy of Sciences*, *102*(15), 5624–5629.
- Paffen, C. L., Gayet, S., Heilbron, M., & Van der Stigchel, S. (2018). Attention-based perceptual learning does not affect access to awareness. *Journal of Vision*, *18*(3), 7–7.
- Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97.
- Rappaport, S. J., Humphreys, G. W., & Riddoch, M. J. (2013). The attraction of yellow corn: Reduced attentional constraints on coding learned conjunctive relations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(4), 1016.
- Rappaport, S. J., Riddoch, M. J., Chechlacz, M., & Humphreys, G. W. (2016). Unconscious familiarity-based color–form binding: Evidence from visual extinction. *Journal of Cognitive Neuroscience*, *28*(3), 501–516.
- Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Research*, *46*(15), 2336–2343.
- Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, *4*(2), 4–4.
- Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecy, M. (1997). The interaction of spatial and object pathways: Evidence from Balint’s syndrome. *Journal of Cognitive Neuroscience*, *9*(3), 295–317.
- Robertson, L., Treisman, A., et al. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science*, *269*(5225), 853–855.
- Salovich, N. A., Remington, R. W., & Jiang, Y. V. (2018). Acquisition of habitual visual attention and transfer to related tasks. *Psychonomic Bulletin & Review*, *25*(3), 1052–1058.
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*(4), 207–218.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological review*, *84*(1), 1.
- Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences*, *99*(26), 17137–17142.
- Seitz, A., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in Cognitive Sciences*, *9*(7), 329–334.
- Seitz, A., & Watanabe, T. (2009). The phenomenon of task-irrelevant perceptual learning. *Vision Research*, *49*(21), 2604–2610.
- Shibata, K., Sagi, D., & Watanabe, T. (2014). Two-stage model in perceptual learning: toward a unified theory. *Annals of the New York Academy of Sciences*, *1316a*, 18.
- Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, *46*(5), 823–835.
- Sowden, P. T., Davies, I. R., & Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults’ visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 379.
- Su, Y., Lai, Y., Huang, W., Tan, W., Qu, Z., & Ding, Y. (2014). Short-term perceptual learning in visual conjunction search. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1415.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 799.

- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.
- Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1295–1306.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.
- VanRullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology*, *3*(1-2), 167.
- VanRullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition*, *17*(1-2), 103–119.
- Vickery, T. J., Sussman, R. S., & Jiang, Y. V. (2010). Spatial context learning survives interference from working memory load. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1358.
- Walsh, V., Ashbridge, E., & Cowey, A. (1998). Cortical plasticity in perceptual learning demonstrated by transcranial magnetic stimulation. *Neuropsychologia*, *36*(1), 45–49.
- Wang, B., & Theeuwes, J. (2018a). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 13.
- Wang, B., & Theeuwes, J. (2018b). Statistical regularities modulate attentional capture independent of search strategy. *Attention, Perception, & Psychophysics*, *80*(7), 1763–1774.
- Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*(6858), 844.
- Wildegger, T., Riddoch, J., & Humphreys, G. W. (2015). Stored color–form knowledge modulates perceptual sensitivity in search. *Attention, Perception, & Psychophysics*, *77*(4), 1223–1238.
- Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 601.
- Yashar, A., & Carrasco, M. (2016). Rapid and long-lasting learning of feature binding. *Cognition*, *154*, 130–138.
- Yashar, A., Wu, X., Chen, J., & Carrasco, M. (2019). Crowding and binding: Not all feature dimensions behave in the same way. *Psychological Science*, *30*(10), 1533–1546.
- Zhang, J. Y., Zhang, G. L., Xiao, L. Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *Journal of Neuroscience*, *30*(37), 12323–12328.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.