Attention, Perception, & Psychophysics (2018) 80:1390–1408
https://doi.org/10.3758/s13414-018-1504-0

Sensitivity to stimulus similarity is associated with greater sustained attention ability

David Rothlein[1,2] · Joseph DeGutis[1,2,3,4] · Laura Germine[4,5] · Jeremy Wilmer[6] · Regina McGlinchey[2,3,4] · Michael Esterman[1,2,3,7]

Published online: 12 April 2018

Abstract

Sustained attention is critical for tasks where perceptual information must be continuously processed, like reading or driving; however, the cognitive processes underlying sustained attention remain incompletely characterized. In the experiments that follow, we explore the relationship between sustaining attention and the contents and maintenance of task-relevant features in an attentional template. Specifically, we administered the gradual onset continuous performance task (gradCPT), a sensitive measure of sustained attention, to a large web-based sample (N>20,000) and a smaller laboratory sample for validation and extension. The gradCPT requires participants to respond to most stimuli (city scenes – 90 %) and withhold to rare target images (mountain scenes – 10 %). By using stimulus similarity to probe the representational content of task-relevant features—assuming either exemplar- or category-based feature matching—we predicted that RTs for city stimuli that were more "mountain-like" would be slower and "city-like" mountain stimuli would elicit more erroneous presses. We found that exemplar-based target-nontarget (T-N) similarity predicted both RTs and erroneous button presses, suggesting a stimulus-specific feature matching process was adopted. Importantly, individual differences in the degree of sensitivity to these similarity measures correlated with conventional measures of attentional ability on the gradCPT as well as another CPT that is perceptually less demanding. In other words, individuals with greater sustained attention ability (assessed by two tasks) were more likely to be influenced by stimulus similarity on the gradCPT. These results suggest that sustained attention facilitates the construction and maintenance of an attentional template that is optimal for a given task.

Keywords Sustained attention · Attentional templates · Visual similarity · Individual differences

Introduction

To sustain visually-directed attention in real-world contexts (e.g., driving at night), a complicated stream of incoming visual information must be continuously monitored for task- or goal-relevant features, locations, or objects (e.g., deer). An important component of this process is constructing and maintaining a task set whereby select features stored in an attentional template are mapped to the appropriate behavioral responses. For example, when driving on a highway through the forest at dusk, the driver should maintain a task set that potentiates deer or deer-like features and link these features with an avoidance response (e.g., break, accelerate, swerve). If a driver fails to develop an efficient attentional template—potentiating features that are either too specific (e.g., Bambi) or too broad (e.g., brown)—it will hinder the driver's ability to rapidly identify and avoid deer. Even if an optimal attentional template is constructed, it must be maintained, and failure to do so may result in a sub-optimal attentional state so that the

✉ David Rothlein
david.rothlein@gmail.com

[1] Boston Attention and Learning Laboratory, VA Boston Healthcare System, Boston, MA 02130, USA

[2] Translational Research Center for TBI and Stress Disorders (TRACTS), VA RR&D TBI Center of Excellence, VA Boston Healthcare System, Boston, MA 02130, USA

[3] Geriatric Research Education and Clinical Center (GRECC), Boston Division VA Healthcare System, Boston, MA 02130, USA

[4] Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA

[5] Institute for Technology in Psychiatry, McLean Hospital, Belmont, MA 02478, USA

[6] Department of Psychology, Wellesley College, Wellesley, MA 02481, USA

[7] Department of Psychiatry, Boston University School of Medicine, Boston, MA 02118, USA

presence of deer-like features would fail to elicit a rapid avoidance response. In the sustained attention literature, such failures to respond to target features are labeled attentional lapses and, as described above, the underlying causes of such lapses can be variable and multifaceted. Individual differences in the ability to sustain attention are stable (trait-like) and the factors driving these individual differences are likely multifaceted for similar reasons. In the present paper, we seek to better characterize the relationship between individual differences in sustained attention ability and the content and maintenance of an attentional template during a sustained attention task.

Probing attentional templates using T-N similarity

In order to assess the strength and content of individual attentional templates we rely on the notion that the similarity of a target stimulus to its nontarget distractors (called T-N similarity by Duncan & Humphreys, 1989) and vice versa can influence performance. This influence has been thoroughly explored and validated in static visual search paradigms (for a recent review see Hout et al., 2015), where participants must give a response indicating whether a target stimulus exists amongst a complex scene of simultaneously presented nontarget stimuli. Specifically, as visual T-N similarity increases (the targets and nontargets are more similar to one another), participants become slower and more error prone (Duncan & Humphreys, 1989). Additionally, RSVP and masked priming experiments have demonstrated that T-N similarity can reduce the likelihood of target detection when nontargets are temporally (within 400 ms) adjacent (Raymond, Shapiro, & Arnell, 1995).

Because T-N similarity has a consistent effect on visual search, T-N similarity effects have been used to probe the content and dynamics of attentional templates during visual search tasks (Geng, DiQuattro, & Helm, 2017; Lee & Geng, 2017). It is assumed that target detection is accomplished by matching an incoming stimulus representation with the features stored in the attentional template (Carlisle, Arita, Pardo, & Woodman, 2011; Desimone & Duncan, 1995; Hout & Goldinger, 2015; Lee & Geng, 2017). Therefore, differences in the influence of T-N similarity on performance could reflect differences in the attentional template. In one study, Lee and Geng (2017) demonstrated that individual differences in the representational similarity for face-morph stimuli predicted behavioral differences in an identity search task. Specifically, individuals who found a given face-pair to be more similar were more likely to be distracted when one item in the pair was a target and the other was a distractor. The differences in the perceptual similarity space of faces were shown to influence individual attentional templates and thus modulate the degree of behavioral interference due to T-N similarity on the visual search task. Aside from this study (see also Charest, Kievit, Schmitz, Deca, & Kriegeskorte, 2014), little work has been done exploring individual differences in the representation of stimulus similarity and how this relates to attentional templates and performance on a given task. Accordingly, we sought to determine the relationship between individual differences in sensitivity to T-N similarity and more standard measures of performance during a sustained attention task. On the one hand, heightened sensitivity to T-N similarity could be a sign of poor filtering/greater interference and thus would lead to conflicting or incorrect signals sent from perceptual and decision processes to motor responses. On the other hand, it could be that those with more robust and efficient attentional templates are more accurate overall but—because both target and non-target features are strongly potentiated—more susceptible to stimulus-specific differences in T-N similarity.

Quantifying T-N similarity in naturalistic scenes

Experimentally measuring the influence of T-N similarity on behavior depends on correctly identifying (or at least approximating) the behaviorally-relevant features involved in the matching process. In a visual search task where a single target is precisely cued, exemplar-specific features are used to identify the target from the distractors. However, as the precision of the cue deviates from the target (e.g., the cue deviates by a size or orientation transformation, or the cue is a category label), the features used to discriminate the target from nontargets become increasingly abstract (Hout & Goldinger, 2015; Maxfield & Zelinsky, 2012; Peelen & Kastner, 2011; Schmidt & Zelinsky, 2009a; Vickery, King, & Jiang, 2005; Yang & Zelinsky, 2009). In the experiments presented below, participants needed to discriminate photographs of city scenes from photographs of mountain scenes. While the discrimination falls across a basic category boundary, there were only ten exemplars in each category and each exemplar was repeated many times. This experimental design allowed for both category-level and exemplar-level features to mediate the target-matching process. Therefore, we quantified the strength and content of attentional templates as the degree of observed T-N similarity and either exemplar or category-level feature spaces were considered when computing T-N similarity. As a consequence of making explicit the feature spaces underlying T-N similarity, we also explored how task demands shape the representational content of attentional templates. It could be that category-level features were optimal as they mediate our ability to rapidly identify novel scenes more generally; however, because this task enabled the use of exemplar-level features, leveraging such features may be optimal to the matching process.

Relating sustained attention to attentional templates

By using T-N similarity to measure the contents and strength of an individual's attentional template during a sustained attention task, we sought to determine if individual differences in the influence of T-N similarity on performance were associated with more standard measures of sustained attention ability. This required a sustained attention task that uses a complex enough stimulus set and target/non-target category boundary to allow for measurable individual differences in attentional templates while providing a reliable measure of sustained attention ability. Conventionally, studies of sustained attention use variants of go/no-go tasks called continuous performance tasks (CPTs) where stimuli are shown in succession and participants are instructed to respond to non-target stimuli while withholding their response to target stimuli (Conners, 2000; Esterman, Noonan, Rosenberg, & Degutis, 2013; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997; Rosenberg, Noonan, Degutis, & Esterman, 2013). Performance is generally evaluated using response errors (commission errors for target stimuli and omission errors for nontargets) and variability in response times (RTs) with greater variability indicative of worse attentional ability. When these tasks are used to measure dynamic changes in attentional states (e.g., Cheyne, Carriere, & Smilek, 2009; Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Esterman et al., 2013; Smallwood et al., 2004; Weissman et al., 2006), it is generally assumed that the fluctuations in performance are driven by internal cognitive processes falling under the broad category of cognitive control. On the other hand, stimulus-driven properties such as stimulus quality, complexity or T-N similarity are either rarely considered or controlled for with the use of very simple stimuli (but see Head & Helton, 2012; Helton & Russell, 2011; Parasuraman et al., 2009). However, given a sufficiently complex stimulus set, the stimulus-driven influences on performance are those that would be most likely to be modulated by the strength and efficiency of a given attentional template. Accordingly, we tested participants on the gradual onset continuous performance task (gradCPT), a CPT variant that removes phasic onsets/offsets between stimuli (Esterman et al., 2013; Fortenbaugh et al., 2015; Rosenberg et al., 2013). Participants were shown a series of images of natural scenes selected from ten photographs of city scenes (nontargets, 90 % of trials) and ten photographs of mountain scenes (targets, 10 % of trials) and made a category decision on each stimulus image. Critically, because the stimulus set involved complex natural images with immense numbers of visual features in each stimulus, the diagnostic features selected in the attentional template, and the degree to which these features were potentiated and maintained, could vary considerably across participants.

In the current study, we report two experiments that examined the relationship between T-N similarity and sustained attention performance using the gradCPT. First, using data collected from a massive online dataset (TestMyBrain.org, N > 20,000), we established that T-N similarity influenced performance (Experiment 1)—evident in both RTs for go trials (nontarget city scenes) and the probability of an erroneous button press (commission errors) on no-go trials (target mountain scenes). Critically, we examined the degree to which individual differences in sensitivity to visual T-N similarity predicted overall performance (as measured by accuracy and RT variability) on the task, showing greater sensitivity was associated with better performance. In Experiment 2, we replicated the primary findings from Experiment 1in a lab-based sample and found that individual differences in sensitivity to similarity derived from performance on the gradCPT predicted performance on a clinically used CPT that used simple stimuli instead of complex scenes.

Experiment 1: Examining the effect of stimulus similarity on gradCPT performance using a large online dataset

In Experiment 1, we set out to determine if and how visual T-N similarity influenced performance on the gradCPT. Data in this dataset have been published in previous studies (Fortenbaugh et al., 2015; Riley et al., 2016; Riley, Esterman, Fortenbaugh, & DeGutis, 2017). Thus, we include a very similar description of the methods below with the addition of new analytic procedures.

Methods

Participants We included 21,406 unpaid volunteers between the ages of 10 and 70 years in the analyses. A total of 25,519 participants performed the gradCPT by visiting TestMyBrain.org, a website containing various cognitive tests, but 3,739 participants were removed due to participant exclusion criterion reported in Fortenbaugh (2015). An additional 374 participants were removed because either their accuracy (d' or criterion), RT variance, or mean RT fell outside 3 standard deviations of the sample mean. There was a nearly equal ratio of males and females (46 % female).

Procedure and stimuli Participants visited TestMyBrain.org—a website that allows users to participate in experimental tasks and provides feedback on performance relative to other users. The gradCPT was presented on its home page as a "Continuous Concentration Task." TestMyBrain.org receives traffic mostly from social-networking sites and search engines (Germine et al., 2012). The gradCPT was one test out of several on TestMyBrain.org (others included face recognition and working memory). Participants were free to complete one or more of these tasks; for some participants, the gradCPT was the first or only test they completed, while others may have completed other tasks first. Single experiment studies on TestMyBrain.org are kept brief (< 10 min) in order to maintain a balance between task completion, participation, and test length. Given the demanding nature of continuous performance tasks, the web-based gradCPT was a shortened version (4 min) of the continuous go/no-go task originally reported in Esterman et al. (2013). This test length was chosen because it was sufficiently short that participant attrition rates would be comparable with those of other experiments on the site. The concern with participants dropping out during testing regards a selection bias in which individuals with poorer sustained-attention ability would be less likely to complete the experiment. This led to a total experiment time of approximately 7 min from consent to debriefing.

Before starting the gradCPT, participants gave informed consent according to the guidelines set by the Committee on the Use of Human Participants at Harvard University and the Wellesley College Institutional Review Board. Participants were then asked to complete a voluntary demographic survey asking about their age, gender, native language, and ethnicity. Comprehensive instructions as well as three practice sessions (30 s each) were then provided to familiarize participants with the scenes and the task so that it could be completed without assistance from an experimenter. Across these practice sessions, each stimulus exemplar was presented at least twice. After completing 4 min of the gradCPT, participants were asked if they had cheated or if any problems occurred during the task and then they were provided with performance feedback. This feedback consisted of a personal score, which was the percentage of correct withholdings to mountains, as well as how this score compared with the average participant.

The current version of the gradCPT contained 20 round, gray-scale photographs; half depicting mountain scenes and half depicting city scenes (Esterman et al., 2013). The stimulus images were resized to a 256 × 256 pixel image and then cropped to appear within a circular frame (radius ~ 126 pixels). In the gradCPT, the scene images were presented in a pseudo random order with 10 % of trials displaying mountain scenes and 90 % city scenes without allowing an identical scene to repeat on consecutive trials. Scene images gradually transitioned from one to the next, in a linear pixel-by-pixel interpolation, with each transition occurring over 800 ms. The new image increased in clarity at the same rate as the old image decreased in clarity. Participants were instructed to press the space bar for city scenes (267 trials) and withhold a response for mountain scenes (32 trials). Trial order was fixed across participants to ensure that stimulus-driven difficulty was matched across participants.

Computing stimulus similarity In order to quantify T-N similarity, we considered the degree to which the task-relevant features were *exemplar-based* or *category-based*. According to the *exemplar-based*[1] hypothesis, participants memorized information about each of the ten city and ten mountain image stimuli and these exemplar-specific features were used to match an incoming stimulus to one of the exemplars stored in memory. Responses could be based entirely on exemplar image identification instead of a broader category label assignment or the exemplar-specific features could facilitate the proper categorization. Alternatively, the *category-based* hypothesis predicted that participants used the same general features that enable them to categorize any photograph upon its first viewing. In assessing T-N similarity, the *exemplar-based* hypothesis predicted that the relevant similarity would be between a given city image and the most similar (nearest neighbor) mountain image in the sample. Specifically, if a given city exemplar was visually similar to a mountain exemplar, even if the city exemplar was not particularly "mountain-like," the stored visual representation of that mountain exemplar would interfere with the response to the city, resulting in a slower response. Alternatively, if T-N similarity was category-based, we predicted the relevant similarity would be between a given city image and the set of stored mountain features that were abstract enough to classify scene images despite the large degree of possible physical image variability within its basic category. As such, cities that shared more "mountain-like" attributes would elicit slower responses.

Along with considering how abstract the task-relevant features were, we also examined how to sensibly characterize the features that composed the scene representations at an exemplar and basic-category level. The exemplar-based hypothesis would rely on image features that could reliably distinguish the exemplars within a category. Here we considered two types of features—pixel intensity and the GIST descriptor. Pixel-based feature matching assumed the stimulus images were represented as a set of position-specific pixel intensity values. Indeed, pixel intensity correlations have been shown to be a neurally-relevant (but not necessarily behaviorally relevant) measure of image similarity for objects (e.g., Peelen & Caramazza, 2012) and scenes (Walther, Caddigan, Fei-Fei, & Beck, 2009). Additionally, pixel-based similarity measures can be used to approximate visual similarity judgments better than chance (Sinha & Russell, 2011). It is worth noting that we do not suggest that pixel intensity correlations were the best approximation of exemplar-level perceptual or representational similarity. However, pixel intensity correlations were easy to compute and sensitive to position-specific, high-frequency commonalities and variations across image pairs. Further, pixel intensity features provided a computational maximum for the fidelity of exemplar representations as the stimulus images were composed of gray-scaled pixel intensity values so identical exemplars were necessarily identical in this feature space and different exemplars were necessarily different. Like any image feature descriptor, the success of pixel intensity values for this purpose rested on the assumption that these features bear a second-order isomorphic relationship to the representations of the actual task-relevant features. The second type of

[1] We wish to differentiate this exemplar-based hypothesis with exemplar theories recognition more generally. In our task, using exemplar features could be beneficial because each exemplar was repeated multiple times. Therefore, memorizing and leveraging exemplar-specific features could provide a fast and accurate way to identify an incoming stimulus by matching it to the stored exemplar features of that stimulus. However, due to the task-specific nature of this exemplar-based hypothesis, results reported here should not be taken as evidence for or against more general theories of exemplar recognition.

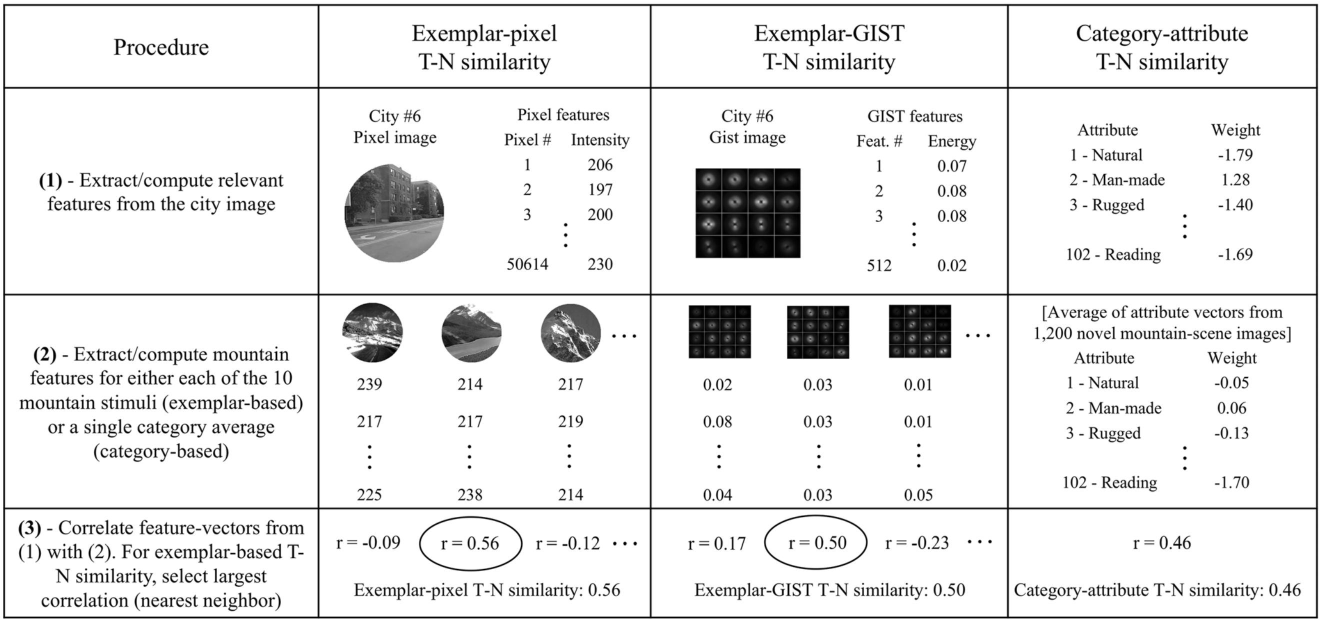
image features consisted of the GIST descriptor values for each image (Oliva & Torralba, 2001). The GIST descriptor is a feature vector that is thought to provide an important summary of the spatial layout of a scene that helps mediate the rapid identification of scene images (Oliva & Torralba, 2001, 2006). This measure is less sensitive to high frequency local variations and more sensitive to spectral features distributed throughout the image.

Accordingly, we computed two exemplar-based T-N similarity values for each stimulus, assuming either pixel image features or GIST descriptor feature vectors. The procedures are depicted in Fig. 1. To construct the exemplar-pixel similarity values, each of the 20 stimulus images were converted into vectors of pixel intensity values. To quantify the influence of exemplar-specific similarity, we reasoned that the most similar cross-category stimulus would lead to the greatest interference and consequently would best reflect the influence of exemplar-specific feature similarity. Therefore, for each stimulus, we computed ten pairwise Pearson correlation values by correlating the pixel image vector of a given stimulus with the pixel vectors of the ten cross-category stimuli. We then selected the largest r (the nearest cross-category neighbor) to be the exemplar-pixel T-N similarity value for that stimulus. For example, the pixel vector for a given city stimulus was correlated with each of the ten mountain pixel vectors and the largest correlation coefficient value was selected. To construct the exemplar-GIST similarity values, the same procedure was employed but instead of using pixel intensity values, each stimulus image was represented by a GIST descriptor feature vector with a length of 512. To construct the GIST feature vector, a 4 × 4 grid was used to divide each stimulus into 16 equal subsets. Within each image subset, Gabor-like filters tuned to 8 different orientations and four different sizes were applied resulting in 32 values for each of the 16 subsets that combined to 512 values in total (Oliva & Torralba, 2001).

In order to quantify each stimulus image's T-N similarity assuming more abstract, category-level features, we used a pre-trained support vector regression model to estimate parameter weights for 102 scene attributes (Patterson & Hays, 2012; Patterson, Xu, Su, & Hays, 2014; Xiao et al., 2013). Briefly, Patterson and Hays (2012) used Mechanical Turk to obtain binary scene attribute judgments for 102 attributes and 14,340 scene images taken from the Sun database (Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). A kernel-based support vector regression model was trained to predict the attribute judgments using as input a set of four image descriptors: the GIST descriptor, Histogram of Gradients (HOG) 2×2, self-similarity, and color histograms (for more details about each descriptor, see Xiao et al., 2010). This pre-trained model allowed us to take any scene image and compute a vector of 102 scene attribute values for that image. Using images from the Places database (Zhou, Lapedriza, Xiao, Torralba, & Oliva, 2014), these scene-attribute vectors were computed for a sample of 1,200 city images taken at random from the *street* and the *downtown* categories as well as 1,200 mountain images taken from the *mountains* and the *snowy mountains* categories. Images were resized and gray-scaled in order to match the gradCPT stimuli. For each category, the 1,200 attribute vectors were averaged to compute a template attribute vector. The resulting mountain and city template attribute vectors were tested for accuracy using a novel sample of 1,200 images from each category. Finally, attribute vectors were computed for each of the gradCPT stimuli and attribute-based T-N similarity was quantified as the Pearson correlation between a given stimulus's attribute vector and the template attribute vector from the opposing category.

When evaluating the feasibility of our category-attribute feature templates on the selected 2,400 novel images (1,200 from each category) we found that a simple correlation classifier achieved 93 % accuracy for both the city and mountain images (chance = 50 %) suggesting that the templates contained information about city and mountain scenes that generalized to novel images. The templates also correctly classified all ten of the gradCPT city stimuli and nine out of the ten mountain stimuli (Mountain 2, shown on Table 2, was incorrectly classified as a city). To further help motivate the use of these attribute vectors, we determined what the most and least informative attributes were. By taking the absolute difference between each attribute from the two category template vectors, we found the ten most informative attributes (greatest absolute difference) were *natural*, *man-made*, *rugged scene*, *mostly vertical components*, *far-away horizon*, *hiking*, *climbing*, *snow*, *camping*, and *enclosed area*. Further, the ten least informative attributes were *cleaning*, *rubber/plastic*, *carpet*, *stressful*, *shingles*, *smoke*, *exercise*, *playing*, *spectating/being in an audience*, and *reading*. That the most and least informative attributes made intuitive sense as being features that should (or should not) distinguish images of city and mountain scenes suggested the templates were successfully approximating components of this abstract feature space.

Reaction time Throughout the task, response times (RTs) were recorded. RTs were calculated relative to the beginning of each image transition such that an RT of 800 ms indicated a button press at the moment the image was 100 % coherent and not mixed with other images. A shorter RT indicated that the current scene was still in the process of transitioning from the previous scene, and a longer RT indicated that the current scene was in the process of transitioning to the subsequent scene. For example, an RT of 720 ms would indicate a button press at the moment of 90 % image n and 10 % image n-1, and so forth. On rare trials with highly deviant RTs (before 70 % coherence of image n and after 40 % coherence of image n-1) or multiple button presses, an iterative algorithm maximized correct responses. The algorithm first assigned



Procedure	Exemplar-pixel T-N similarity	Exemplar-GIST T-N similarity	Category-attribute T-N similarity
(1) - Extract/compute relevant features from the city image	City #6 Pixel image Pixel features Pixel # Intensity 1 206 2 197 3 200 ⋮ 50614 230	City #6 Gist image GIST features Feat. # Energy 1 0.07 2 0.08 3 0.08 ⋮ 512 0.02	Attribute Weight 1 - Natural -1.79 2 - Man-made 1.28 3 - Rugged -1.40 ⋮ 102 - Reading -1.69
(2) - Extract/compute mountain features for either each of the 10 mountain stimuli (exemplar-based) or a single category average (category-based)	239 214 217 … 217 217 219 ⋮ ⋮ ⋮ 225 238 214	0.02 0.03 0.01 … 0.08 0.03 0.01 ⋮ ⋮ ⋮ 0.04 0.03 0.05	[Average of attribute vectors from 1,200 novel mountain-scene images] Attribute Weight 1 - Natural -0.05 2 - Man-made 0.06 3 - Rugged -0.13 ⋮ 102 - Reading -1.70
(3) - Correlate feature-vectors from (1) with (2). For exemplar-based T-N similarity, select largest correlation (nearest neighbor)	r = -0.09 r = 0.56 r = -0.12 … Exemplar-pixel T-N similarity: 0.56	r = 0.17 r = 0.50 r = -0.23 … Exemplar-GIST T-N similarity: 0.50	r = 0.46 Category-attribute T-N similarity: 0.46

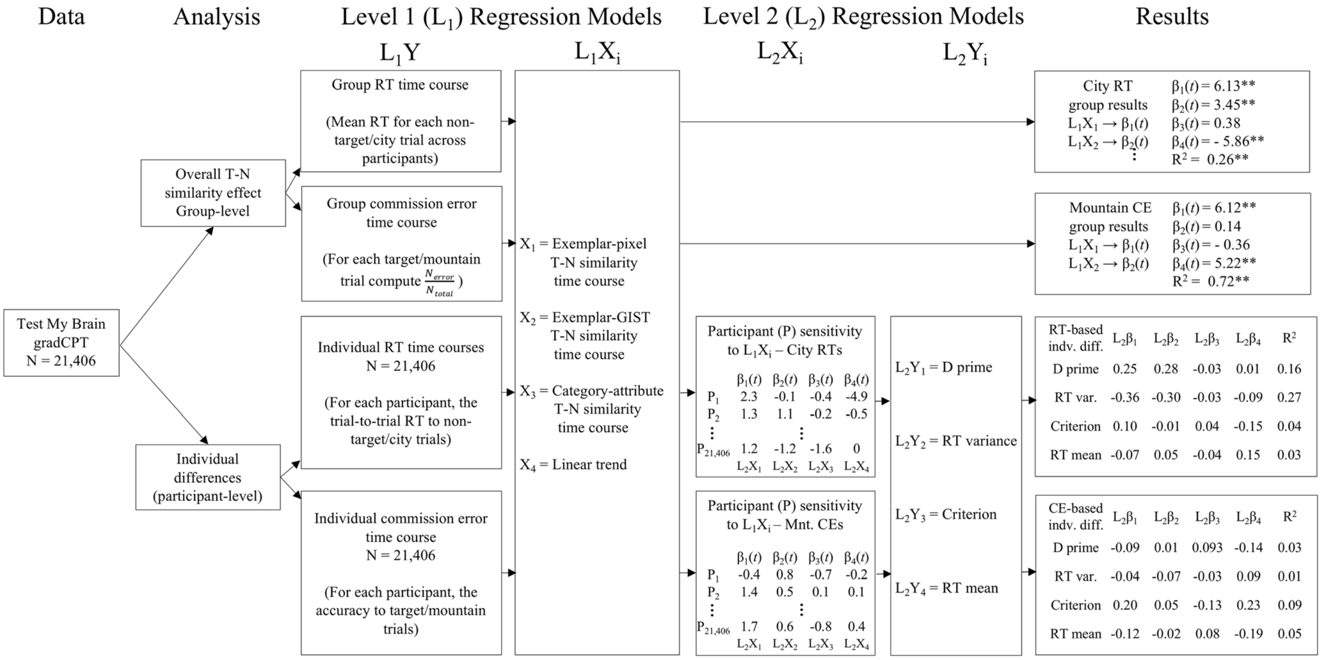
Figure 1. The three types of T-N similarity. The figure depicts how each type of T-N similarity was computed for city exemplar # 6 (see Table 1). Each column indicates a different type of T-N similarity and each row indicates a step in the procedure used to compute T-N similarity

unambiguous correct responses, leaving few ambiguous button presses. Ambiguous presses were then assigned to an adjacent trial if one of the two trials had no response. If both adjacent trials had no response, the press was assigned to the closest trial unless one trial was a no-go target, in which case participants were given the benefit of the doubt that they correctly omitted. If there were multiple presses that could be assigned to any one trial, the fastest response was selected. Slight variations to this algorithm yielded highly similar results, as most button presses showed a 1:1 correspondence with presented images.

Measures of gradCPT performance Taking a signal detection approach, we defined hits as correct omissions to mountains and false alarms as incorrect omissions to cities. Following these definitions, we computed discrimination ability (d') and criterion (a measure of strategy or bias to respond in the case of uncertainty). We used standard procedures to correct for cases in which hit rates were 100 % or false alarm rates were 0 %, with one-half error deducted or added on the basis of the number of target or nontarget trials presented, respectively. In addition to d prime and criterion, we computed, for each participant, RT variance or more precisely the coefficient of variation of their RTs (CV), which is a measure of ability with lower CV corresponding to greater ability. Finally, we computed the mean of their RTs, which is a measure of strategy with slower RTs corresponding to a more conservative strategy. This ability/strategy distinction stems from Fortenbaugh et al. (2015), who found, through a factor analysis, that the ability (d prime and CV) and strategy (criterion and mean RT) factors dissociate across the lifespan.

Quantifying the influence of T-N similarity on RTs Following Duncan and Humphreys (1989), we predicted that city stimuli that had a greater degree of T-N similarity would result in slower RTs. To test this (see Fig. 2), an estimate of the population RT time course for the 267 city trials was constructed by averaging together, at each trial, the RTs across the 21,406 participants. The trial order was identical across participants so averaging together the RTs maintained trial-to-trial variance in RT due to stimulus-driven factors like T-N similarity. Furthermore, the N was large enough for us to treat the mean RT as an approximate to the population values.[2] Because we wanted to see which type of T-N similarity best explained trial-to-trial variance in RTs, the RT time course was tested for the influence of T-N similarity by running a simultaneous regression analysis including four predictors—the pixel-based exemplar T-N similarity measure, the GIST-based exemplar T-N similarity measure, the attribute-based T-N similarity measure, and a linear trend. The linear trend predicted RT slowing over the course of the experiment. Each of the 3 T-N similarity predictors assigned the T-N similarity value corresponding to the city stimulus shown on that trial for each of the 267 trials. The R^2 computed from this model reflected the amount of trial-to-trial RT variance the similarity measures and the linear trend explained and the standardized beta coefficients reflected the unique relative contributions of each of the T-N similarity measures. In order to compute standardized beta coefficients, a constant was added to each model and all

[2] We computed 95 % confidence intervals for the average RTs from each trial in the 267 trial time course. On average, the 95 % CIs spanned 3.7 ms with the largest CI spanning 4.7 ms.


Data
Analysis
Level 1 (L$_1$) Regression Models
L$_1$Y
L$_1$X$_i$
Level 2 (L$_2$) Regression Models
L$_2$X$_i$
L$_2$Y$_i$
Results
Test My Brain gradCPT N = 21,406
Overall T-N similarity effect Group-level
Individual differences (participant-level)
Group RT time course (Mean RT for each non-target/city trial across participants)
Group commission error time course (For each target/mountain trial compute $\frac{N_{error}}{N_{total}}$)
Individual RT time courses N = 21,406 (For each participant, the trial-to-trial RT to non-target/city trials)
Individual commission error time course N = 21,406 (For each participant, the accuracy to target/mountain trials)
X$_1$ = Exemplar-pixel T-N similarity time course
X$_2$ = Exemplar-GIST T-N similarity time course
X$_3$ = Category-attribute T-N similarity time course
X$_4$ = Linear trend
City RT group results
L$_1$X$_1$ → β$_1$(t)
L$_1$X$_2$ → β$_2$(t)
⋮
β$_1$(t) = 6.13**
β$_2$(t) = 3.45**
β$_3$(t) = 0.38
β$_4$(t) = - 5.86**
R^2 = 0.26**
Mountain CE group results
L$_1$X$_1$ → β$_1$(t)
L$_1$X$_2$ → β$_2$(t)
β$_1$(t) = 6.12**
β$_2$(t) = 0.14
β$_3$(t) = - 0.36
β$_4$(t) = 5.22**
R^2 = 0.72**
Participant (P) sensitivity to L$_1$X$_i$ – City RTs
Participant (P) sensitivity to L$_1$X$_i$ – Mnt. CEs
L$_2$Y$_1$ = D prime
L$_2$Y$_2$ = RT variance
L$_2$Y$_3$ = Criterion
L$_2$Y$_4$ = RT mean

	$\beta_1(t)$	$\beta_2(t)$	$\beta_3(t)$	$\beta_4(t)$
P_1	2.3	-0.1	-0.4	-4.9
P_2	1.3	1.1	-0.2	-0.5
⋮		⋮		
$P_{21,406}$	1.2	-1.2	-1.6	0
	L_2X_1	L_2X_2	L_2X_3	L_2X_4

	$\beta_1(t)$	$\beta_2(t)$	$\beta_3(t)$	$\beta_4(t)$
P_1	-0.4	0.8	-0.7	-0.2
P_2	1.4	0.5	0.1	0.1
⋮		⋮		
$P_{21,406}$	1.7	0.6	-0.8	0.4
	L_2X_1	L_2X_2	L_2X_3	L_2X_4

RT-based indv. diff.	$L_2\beta_1$	$L_2\beta_2$	$L_2\beta_3$	$L_2\beta_4$	R^2
D prime	0.25	0.28	-0.03	0.01	0.16
RT var.	-0.36	-0.30	-0.03	-0.09	0.27
Criterion	0.10	-0.01	0.04	-0.15	0.04
RT mean	-0.07	0.05	-0.04	0.15	0.03

CE-based indv. diff.	$L_2\beta_1$	$L_2\beta_2$	$L_2\beta_3$	$L_2\beta_4$	R^2
D prime	-0.09	0.01	0.093	-0.14	0.03
RT var.	-0.04	-0.07	-0.03	0.09	0.01
Criterion	0.20	0.05	-0.13	0.23	0.09
RT mean	-0.12	-0.02	0.08	-0.19	0.05

Figure 2. Summary of the analyses and results from Experiment 1. The level one (L1) regression analyses fit to the group average data estimated the overall effect of T-N similarity while the same regression analyses fit to each participants data estimated that participants sensitivity to T-N similarity and the coefficient *t*-values were used as input to the level 2 regression analyses that fit these measures of sensitivity to the different types of T-N similarity to measures of individual performance on the gradCPT

variables were normalized to have a mean of 0 and a standard deviation of 1.

Quantifying the influence of T-N similarity on commission errors We additionally predicted that greater T-N similarity for mountain stimuli (i.e., mountain images that looked more "city-like") should result in more errors. To test this, we ran the same type of group regression analysis as described in the previous section but adapted them for mountain trials. First, for each of the 32 mountain trials, the rate of erroneous button presses or commission error rate (CE rate) was computed by dividing the number of participants that made an error at a given trial by the total number of participants. This CE rate time-course was the dependent variable for the regression analysis that included the three T-N similarity predictors for the mountain stimuli and a linear trend that predicted an increase in error likelihood over the experimental time course.

Individual differences To examine whether individual differences in sensitivity to similarity were related to performance on the gradCPT, we computed individual sensitivity to T-N similarity by running the regressions described above on each participant's data separately. Therefore, for each participant, we ran two regression analyses, one predicting mountain-trial accuracy and one predicting the city-trial RTs. Each model produced four beta coefficients and sensitivity to T-N similarity was quantified as the *t*-values from the beta coefficients. The set of beta *t*-values from each participant resulted in four individual-difference indices—three quantifying individual differences in sensitivity to T-N similarity (pixel-based, GIST-based and attribute-based) and one for linear trend. First, each of the *t*-values were correlated (Pearson) with the four different measures of gradCPT performance (d prime, criterion, RT variance, or mean RT). Next, to determine the unique contribution of the sensitivity to similarity measures, they were entered into second-level regression models—one for each of the measures of gradCPT performance.

Results

We set out to determine (1) if T-N similarity influenced performance on the gradCPT and (2) if individual differences in sensitivity to T-N similarity were related to more conventional measures of task performance. To do this, we computed three measures of T-N similarity—one assuming a pixel-based exemplar matching procedure, another assuming a GIST-based exemplar matching procedure, and a third assuming an attribute-based category template matching procedure. Tables 1 and 2 show each of the stimuli along with their three T-N similarity values as well as mean RT for city stimuli (Table 1) and CE rate for mountain stimuli (Table 2). Tables 3 and 4 show the degree to which these measures covary.

The influence of T-N similarity on city RTs To determine if (and if so, which measure of) T-N similarity influenced performance on the gradCPT, a regression analysis was performed to simultaneously test for the influence of exemplar-pixel, exemplar-GIST, and category-attribute T-N similarity (as well as a linear trend) on the trial-to-trial group average RTs in response to city stimuli. As reported in Fig. 3, the overall model was significant ($R^2 = 0.26$, $F(4, 262) = 22.78$, $p < .001$) with the coefficients for the exemplar-pixel ($\beta = 0.34$, $t(262) = 6.13$, $p < .001$) and exemplar-GIST ($\beta = 0.21$, $t(262) = 3.45$, $p < .001$) T-N similarity measures explaining unique variance while the attribute-based measure did not ($\beta = 0.02$, $t(262) = 0.38$, $p = .70$). The significant linear trend ($\beta = -0.31$, $t(262) = -5.86$, $p < .001$) indicated that participants sped up over time.

The influence of T-N similarity on mountain CEs We additionally ran an analogous regression analysis predicting the trial-to-trial commission error rates for mountain trials (see Figs. 3 and 4). This model was quite successful explaining nearly 75 % of the variance in trial to trial error rates ($R^2 = 0.72$, $F(4, 27) = 17.48$, $p < .001$). Like the model predicting RTs to city stimuli, the exemplar-pixel measure of T-N similarity and the linear trend predictors had significant coefficients while the category-attribute predictor did not (exemplar-pixel: $\beta = 0.69$, $t(27) = 6.12$, $p < .001$; linear trend: $\beta = 0.57$, $t(27) = 5.22$, $p < .001$; category-attribute $\beta = -0.05$, $t(27) = -0.36$, $p = .72$). Unlike the city RT model, the exemplar-GIST predictor was not significant ($\beta = 0.02$, $t(27) = 0.14$, $p = .89$). Overall, the exemplar-pixel and, to a lesser degree, the exemplar-GIST measures of T-N similarity could explain some of the trial-to-trial variance in performance such that more "mountain-like" city stimuli and more "city-like" mountain stimuli according to these measures led to slower and more error-prone responses, respectively, while the category-attribute was not significant for either. The significant linear trends revealed that, while accounting for T-N similarity, participants responded more quickly and made more commission errors as the experiment progressed.

Relating sensitivity to T-N similarity to individual differences in gradCPT performance We ran a set of eight second-level regression analyses in order to evaluate how individual differences in the sensitivity to each of the three measures T-N similarity as well linear trend predicted gradCPT ability (d prime and CV) and measures of strategy (criterion and mean RT). Because the N was so large, even a model with a small R^2 was significant. We therefore followed Cohen's (1988) convention that states that R^2 cutoffs for small, medium, and large effects are $R^2 = 0.02$, 0.13, and 0.26, respectively. We found that sensitivity to similarity as measured by the influence of T-N similarity on RTs to city stimuli could predict both measures of gradCPT ability—specifically, discrimination ability/accuracy (d prime) with a medium effect ($R^2 = 0.16$) and RT variability (CV) with a large effect ($R^2 = 0.27$). Examination of the coefficients revealed that the success of the models was largely driven by sensitivity to the exemplar-pixel (d prime: $\beta = 0.25$; CV: $\beta = -0.36$) and exemplar-GIST (d prime: $\beta = 0.28$; CV: $\beta = -0.30$) measures of T-N similarity (see Figure 2). Importantly, the signs of these coefficients revealed that individuals who were more sensitive to the exemplar-based measures of T-N similarity were more accurate (greater d prime) and had more consistent RTs (lower RT variability). Neither sensitivity to the category-attribute measure T-N similarity (d prime: $\beta = -0.03$; CV: $\beta = -0.03$) nor linear trend (d prime: $\beta = 0.01$; CV: $\beta = -0.09$) substantially contributed to either model. These measures did not predict either measure of strategy beyond a small effect (criterion: $R^2 = 0.04$; RT mean: $R^2 = 0.03$).[3] Furthermore, none of the models using sensitivity to T-N similarity derived from accuracy to mountain stimuli predicted any of the four measures of gradCPT performance beyond a small effect.

Experiment 1 discussion

Experiment 1 provided, to our knowledge, the first analysis of how visual image similarity impacts sustained attention and how individual differences in sensitivity to visual similarity are related to sustained attention ability. In particular, using a sample size of more than 20,000 participants, we found compelling evidence that increased T-N similarity was related to performance on the gradCPT—both in terms of slowing RTs for city trials and increasing the likelihood of commission errors for mountain trials—and that individual differences in sensitivity to certain measures of T-N similarity was associated with more optimal performance on the task.

First, we showed that, at the group level, the exemplar-pixel measure of T-N similarity explained variance in both the average RT time course for non-target city stimuli as well as the error-rate time course for the target mountain stimuli while the exemplar-GIST measure only explained variance in the RT time course. The category-attribute measure, however, did not significantly explain variance in either. As expected, stimuli having greater T-N similarity showed slower RTs and greater error rates. We posit that the cognitive processes underlying the success of the exemplar-based T-N similarity measures are computationally analogous to how the T-N similarity measures were constructed—namely, that the slower

[3] The overall mean RT from the online version of the gradCPT reported in Experiment 1 was an average of ~80 ms slower than lab-based versions of the gradCPT. We believe this difference was likely due to hardware issues intrinsic to online data collection and therefore the overall mean RT would not be a reliability indicator of the state of the stimulus when the response was made (see Fortenbaugh et al., 2015 for a more extensive discussion). With this caveat in mind, the correlation between mean RT and sensitivity to T-N similarity was small, meaning participants with greater sensitivity to similarity were not faster overall. The data showed that participants with greater sensitivity to T-N similarity were instead more accurate and more stable with RTs which could mean they were identifying the scenes at a faster rate but instead of responding more quickly, allowing more evidence to accrue before responding. This could explain how these participants had more accurate responses while keeping up with the fast pace of the task.

Table 1 The three predicted T-N similarity measures and the mean RT in ms (from Exp. 1) for the ten non-target (city) stimuli. Intervals reported below the mean RTs are their 95% confidence intervals

City variables	C. 1 	C. 2 	C. 3 	C. 4 	C. 5 	C. 6 	C. 7 	C. 8 	C. 9 	C. 10
Pixel	0.24	0.21	0.21	0.30	0.14	0.56	0.51	0.42	0.40	0.41
GIST	0.31	0.49	0.35	0.34	0.42	0.50	0.33	0.32	0.33	0.27
Attribute	0.42	0.25	0.44	0.40	0.40	0.46	0.40	0.32	0.52	0.57
Mean RT	858 ±0.93	864 ±0.99	858 ±0.94	873 ±0.99	861 ±0.96	899 ±0.97	884 ±0.95	867 ±0.95	883 ±1	861 ±0.91

C city, *RT* response time

RTs were a consequence of interference from image features from the most confusable mountain exemplar (nearest cross-category neighbor) stored or potentiated in the attentional template, sending a conflicting response signal (withhold in the case of city stimuli and press in the case of mountain stimuli). The success of the exemplar-based measures of T-N similarity further suggests that participants executed the task, at least in part, by adopting an exemplar-based matching procedure whereby the perceptual decision (city vs. mountain) was accomplished by storing stimulus-specific image features from each stimulus and leveraging all of these features when making the perceptual decision. The failure of the attribute-based category T-N similarity measure could mean that exemplar features were used at the exclusion of the attribute-based T-N similarity. However, this interpretation presupposes that the attribute-based T-N similarity sufficiently approximates the appropriate category-level features used to classify novel scenes. Since this was not independently verified, any interpretation of the null effect here is ambiguous.

Inclusion of the linear trend in the regression models revealed a vigilance decrement by showing an increase in error rate throughout the time course of the experiment as well as RT speeding, which has been linked with more automated (less vigilant) attentional states (deBettencourt, Norman, & Turk-Browne, 2017). While not the focus of this study, Fig. 4 (plot on bottom right) illustrates that using T-N similarity can help reveal performance variation due to intrinsic factors like executive or cognitive control by controlling for trial-to-trial variance due to stimulus driven factors.

We further found that both the pixel-based and the GIST-based image features contributed unique variance to explaining reaction times. This suggests that the exemplar representations stored in memory were representing multiple types of image properties and that similarity, as derived through pixel intensity and the GIST descriptor features, approximated the similarity structure at separate levels of representation involved in the categorization of scene images. Alternatively, the pixel and GIST measures could capture different facets of the similarity structure at a single level of representation.

Finally, individuals who were more sensitive to T-N similarity as indexed by both the pixel and GIST exemplar T-N similarity measures to city RTs were better at the task overall, with greater discrimination ability (higher d prime) and less variable RTs (CV). This result is novel, even in the broader context of visual similarity across the visual search and RSVP literature. While other studies have found that individual differences in the perceived similarity of target and non-target stimuli could predict individual differences in sensitivity to T-N similarity (Lee & Geng, 2017), sensitivity to T-N similarity has not (to our knowledge) been used as an index to predict overall performance on a given task.

One potential concern is that since sensitivity to T-N similarity was based on city RTs, perhaps it was really an indirect measure of RT variability and therefore the individual difference correlations between sensitivity to T-N similarity and CV were statistically circular. However, the results were not consistent with this interpretation. Increased sensitivity to T-N similarity was measured from an increase of a particular form of RT variability, namely slower responses to cities that have a similar mountain exemplar. Therefore, if everything else was equal, increased sensitivity to T-N similarity would lead to an increase in RT variance. In fact, the opposite was true—

Table 2 The three predicted T-N similarity measures and the commission error rate (from Exp. 1) for the ten target (mountain) stimuli

Mountain variables	M. 1 	M. 2 	M. 3 	M. 4 	M. 5 	M. 6 	M. 7 	M. 8 	M. 9 	M. 10
Pixel	0.39	0.10	0.19	0.14	0.06	0.41	0.32	0.21	0.56	0.21
GIST	0.22	0.31	0.21	0.49	0.27	0.16	0.26	0.24	0.50	0.47
Attribute	0.59	0.61	0.58	0.43	0.45	0.51	0.44	0.48	0.41	0.59
Error rate	0.23	0.10	0.16	0.21	0.15	0.26	0.34	0.23	0.33	0.38

Table 3 Correlation (Pearson) matrix of item-level (ten cities) variance in predicted T-N similarity measures and mean RT for the non-target (city) stimuli

City variables	Pixel	GIST	Attribute	Mean RT
Pixel	-	-0.08	0.33	0.80*
GIST	-	-	-0.46	0.35
Attribute	-	-	-	0.19
Mean RT	-	-	-	-

*Indicates significance at p < .05

RT response time

greater sensitivity to T-N similarity was associated with lower overall RT variability. This suggests that greater sensitivity to T-N similarity was associated with a large enough reduction of other causes of RT variability (e.g., ability to sustain attention) to counteract the increased RT variability that was directly a result of increased sensitivity to T-N similarity.

While sensitivity to T-N similarity was associated with overall performance on the task, the nature of this association was unclear. Specifically, because complex stimuli were used, it was unclear whether the individual differences in both sensitivity to T-N similarity and gradCPT performance were due underlying differences in perceptual or attentional abilities. In Experiment 2, in addition to replicating the findings of Experiment 1 in a diverse, lab-based sample of Veterans, we explored the underlying causes of the individual differences in sensitivity to exemplar T-N similarity. In particular, we examined if the individual differences in T-N similarity were driven by differences in perceptual processing ability such that the perceptual decision component of the gradCPT was easier for some participants. Alternatively, such differences could have been driven by differences in the ability to sustain attention more generally. By testing participants on both the gradCPT and the TOVA—a CPT that uses simple visual stimuli—we assessed attentional ability on a perceptually simple task where individual differences in perceptual processing ability had a minimal effect and saw if it related to sensitivity to T-N similarity as measured on the gradCPT. Any relationship would probably not have been due to differences in perceptual processing ability, leaving sustained attention ability as the likely cause.

Table 4 Correlation (Pearson) matrix of item-level (ten mountains) variance in predicted T-N similarity measures and the item level commission error rate for the target (mountain) stimuli

Mountain variables	Pixel	GIST	Attribute	Error rate
Pixel	-	0.05	-0.23	0.61
GIST	-	-	-0.31	0.40
Attribute	-	-	-	-0.28
Error rate	-	-	-	-

Experiment 2: Examining how gradCPT sensitivity to similarity relates to performance on a neuropsychological test of sustained attention

In Experiment 1 we demonstrated that the exemplar-based T-N similarity influenced RTs for the city (non-target) trials and the likelihood of a CE error for mountain (target) trials. Critically, increased individual sensitivity to similarity, when assessed by the influence of exemplar-based T-N similarity on RTs, was associated with increased overall ability on the task. The meaning of this association, however, is open to interpretation. One possibility is that sensitivity to image similarity solely reflects enhanced bottom-up perceptual processing or visual memory independent of an individual's ability to sustain attention. Alternatively, sensitivity to similarity could reflect individual differences in the ability to maintain the task set. In other words, those who are more capable at sustaining attention more generally are consequently more capable at selecting and maintaining task-relevant features and responding appropriately. To choose between these possible interpretations, we analyzed data from participants who performed the gradCPT along with a battery of surveys and other tests. One such test was the Test of Variables of Attention (TOVA), a CPT that is widely used to help diagnose attention-deficit hyperactivity disorder (Leark, Greenberg, Kindschi, Dupuy, & Hughes, 2007). The TOVA was the only other CPT presented in the battery and was selected for this experiment because it assessed sustained attention in a similar manner to the gradCPT but without the same perceptual demands. No other task in the battery was examined for the purposes of this study. For the target/non-target category distinction, it employs a highly discriminable position location (top vs. bottom) of a simple square. If increased sensitivity to similarity arises from enhanced bottom-up perceptual processing or visual memory capacity, the TOVA—using simple stimuli with low perceptual discrimination demands—would largely eliminate any advantage individuals with enhanced perceptual processing would have. Therefore, we would predict no association between sensitivity to T-N similarity on the gradCPT and attentional ability on the TOVA. If the latter hypothesis is true, we would predict that individual differences in sensitivity to T-N similarity on the gradCPT should predict performance on the TOVA as well because the attentional demands across the two tasks are similar and the sustained attention ability that determines the degree of sensitivity to T-N similarity will be measured during the TOVA.

Methods

Participants The participant sample included 152 Veterans (male: 141; age: 21–56 years) from Operation Enduring Freedom/Operation Iraqi Freedom/Operation New Dawn (OEF/OIF/OND) who were recruited from the Translational

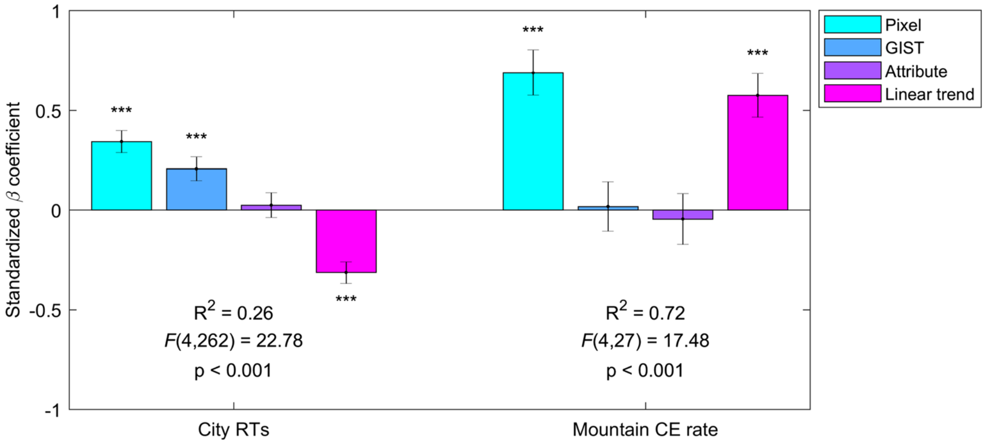


Figure 3. Bar graph depicting the standardized beta coefficients for the regression models predicting city RTs and mountain CE rates. Error bars represent the standard errors of the beta coefficients. Statistics for each model are placed below the set of four bars associated with that model. * indicates p < .05, ** indicates p < .005 and *** indicates p < .001

Research Center for Traumatic Brain Injury and Stress Disorders (TRACTS) at the Veterans Affairs Boston Healthcare System. Details regarding the larger TRACTS cohort has been described in detail elsewhere (Lippa et al., 2015). Participants in the present sample completed the current tasks as part of a larger battery of tests that included clinical interviews and neuropsychological testing that took place over one or more sessions. This Veteran sample was used because it had available data from participants who did both the gradCPT and the TOVA. The Institutional Review Board of Human Studies Research at the VA Boston Healthcare System approved all research procedures. All participants provided informed consent and were reimbursed for their time and travel expenses.

Initially 169 participants completed the gradCPT but some participants (N = 17) were removed due to exclusion criteria described below. Overall exclusion criteria for TRACTS includes: (a) history of neurological illness (other than traumatic brain injury (TBI)); (b) history of seizures; (c) current diagnosis of schizophrenia spectrum or other psychotic disorders (not

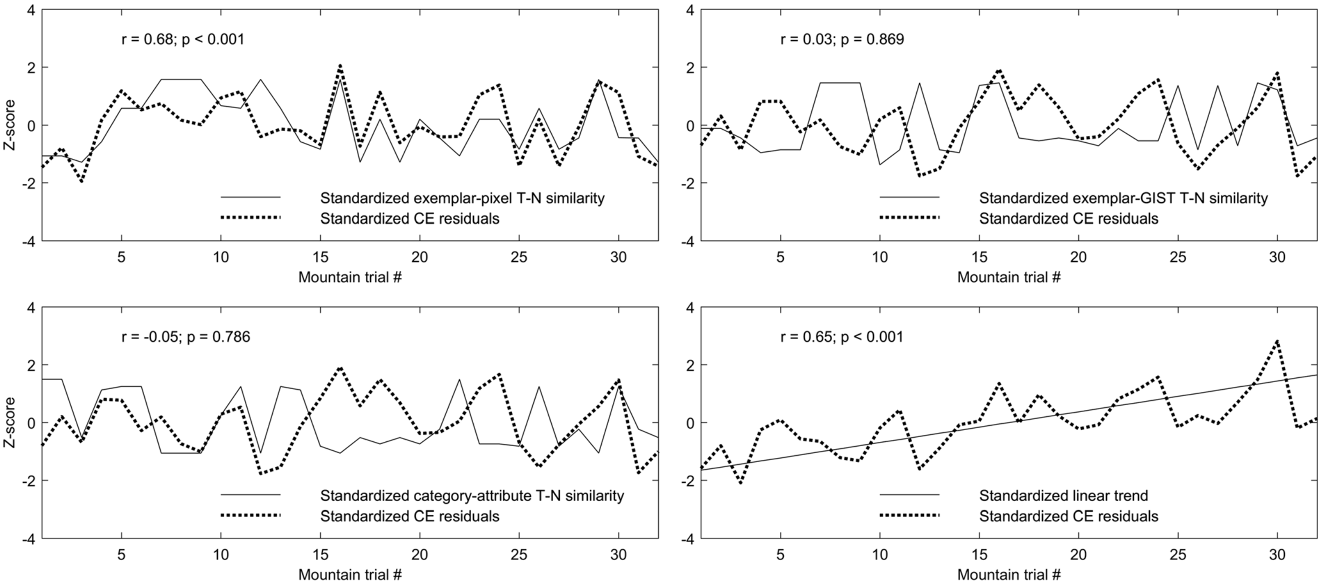


Figure 4. Plots of the unique contribution between the group commission error rate (CE rate) and each of the 4 predictors included in the regression analysis (the three types of T-N similarity and a linear trend). In order to visualize the unique contribution of a given predictor to the CE rate, that predictor is fit against standardized CE residuals from a regression analysis with the other three predictors. The similarity time courses consist of the similarity value assigned to the corresponding mountain image at each trial. The similarity time courses were *Z*-normalized for visualization purposes. Of particular interest is the graph depicting the CE residuals fit to the linear trend (bottom-right). By controlling for the influence of T-N similarity on the CE time-course, a strong linear trend is revealed. Controlling for T-N similarity could reveal effects that otherwise may have been missed

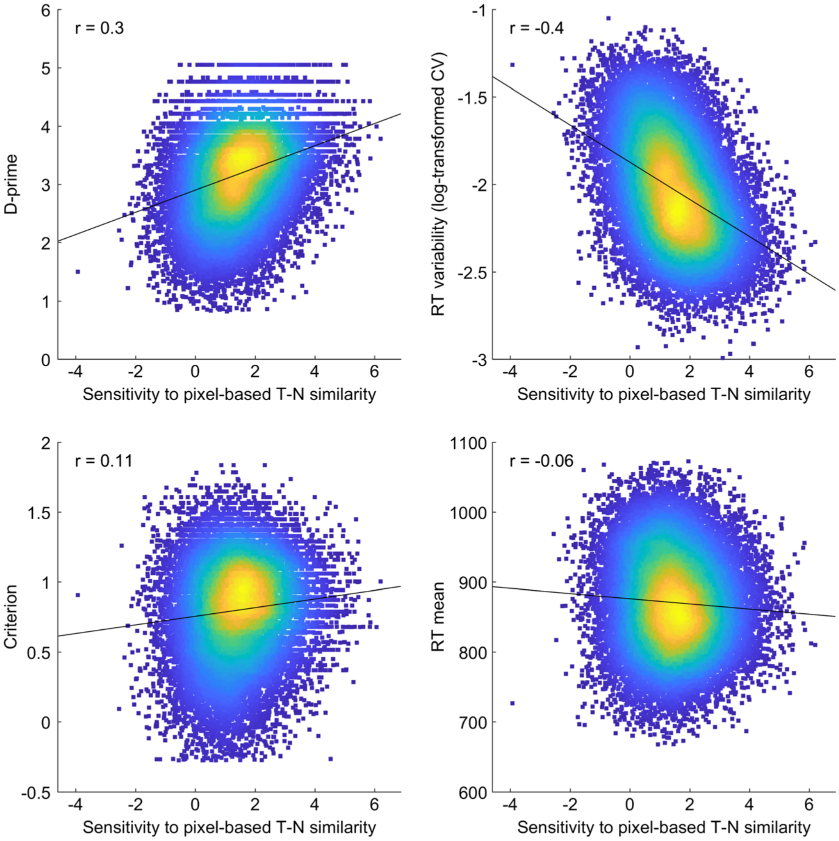


Figure 5. Scatterplots between the sensitivity to the pixel-based T-N similarity and measures of gradCPT performance. The color of each data point tracks the surrounding data density with blue, indicating a data point in a low-density region and yellow indicating a high-density region. Sensitivity to the pixel-based T-N similarity was estimated using city RTs. Pearson correlation *r* values are reported. Because the N is large, correlations are assessed by Cohen's "rule of thumb" with *r* > 0.1, 0.3 and 0.5 being the cutoffs for small, medium, and large correlation effect sizes. The top row shows the scatterplot between the sensitivity to exemplar-pixel T-N similarity and measures of ability on the gradCPT whereas the bottom row shows the relationship with measures of strategy. The correlations are much stronger with the measures of ability than the measures of strategy

related to PTSD); (d) current active suicidal and/or homicidal ideation, intent, or plan requiring crisis intervention; or (e) cognitive disorder due to general medical condition other than TBI (exclusions based on these criteria: N = 2). For the present study, veterans with a history of moderate or severe TBI or participants who failed the Green's Medical Symptom Validity Test were excluded (N = 13). For gradCPT-specific exclusion criteria, as in the previous experiments, we excluded participants who showed significant periods (30 s or greater) with no response (N = 2).

Participants completed both the gradCPT and the TOVA along with a battery of additional tasks. A majority of participants returned for an additional testing session (N = 94) after an average of 762 days (range 363–1,743 days). For the participants who did the gradCPT in both sessions (N = 10) only the gradCPT results from the first session were analyzed. Additionally, of the 152 participants who did the gradCPT, 136 performed the TOVA—133 participants have data from the gradCPT and TOVA collected during the same session and 90 participants have data collected from different sessions.

Stimuli and procedure

GradCPT The stimuli and procedure were identical to Experiment 1 with a few exceptions. First, the participants performed that gradCPT during an fMRI scan; however, the fMRI data are out of the scope of this paper and not discussed further. The experimental task was projected onto a wall to the rear of the scanner and participants viewed the task through a mirror. Next, this version of the gradCPT was longer, consisting of 600 trials (540 city non-targets and 60 mountain targets) and lasting 8 min. Finally, unlike Experiment 1, in which the trial order was fixed across participants, each participant had a unique randomized trial order.

Tests of Variables of Attention (TOVA) The TOVA consisted of a simple continuous performance task—widely used to diagnose ADHD—where participants were instructed to respond to the target stimulus and ignore the non-target stimulus (Leark et al., 2007). Both the target and non-target stimuli were composed of a white square with a smaller black square horizontally centered within the white square. For the target stimulus, the black square was positioned near the top of the white square while for the non-target stimulus the black square was positioned near the bottom. Responses were made by pressing a button on a switch box. A trial had a duration of 2,000 ms and consisted of a stimulus presented for 100 ms and a 1,900-ms response period. In total, the TOVA consisted of 648 trials, with 50 % of trials being target trials. However, in the first half of the TOVA target trials were infrequent (22.5 % of first-half trials) while in the second half target trials were more frequent (77.5% of second-half trials). Similar to the gradCPT, we obtained each participant's RT variability (in this case RT standard deviation or RT SD), mean RT, d prime, and criterion.

Data analysis and results

Replicating Experiment 1 First we set out to replicate the association between individual differences in sensitivity to T-N similarity and variables that measure strategy (criterion and mean RT) and ability (d prime and CV) on the gradCPT reported in Experiment 1. Using the same analytical methods described in Experiment 1, we ran a regression model for each participant fitting their RT time course to non-target city stimuli with exemplar-pixel, exemplar-GIST, and category-attribute measures of T-N similarity as well as a linear trend as predictors. The resulting *t*-values derived from the four predictor coefficients (a set of four *t*-values for each participant) quantified each participant's sensitivity to the different measures of T-N similarity and linear trend and entered as predictors into four second-level regression analyses predicting each participant's d prime, RT variance (CV), criterion, and RT mean on the gradCPT. Consistent with Experiment 1, we found that sensitivity to T-N similarity could predict discrimination ability/accuracy (d prime) with a medium effect ($R^2 = 0.20$) and RT variance (CV) with a medium effect ($R^2 = 0.24$). Moreover, examination of the coefficients revealed that the success of the models was largely driven by sensitivity to the exemplar-pixel (d prime: $\beta = 0.18$, $t(147) = 2.53$, p = .013; CV: $\beta = -0.22$, $t(147) = -3.22$, p = .002) and exemplar-GIST (d prime: $\beta = 0.25$, $t(147) = 3.22$, p = .002; CV: $\beta = -0.28$, $t(147) = -3.76$, p < .001) measures of T-N similarity. The coefficients for sensitivity to category-attribute T-N similarity were not significant (d prime: $\beta = -0.09$, $t(147) = -1.04$, p = .30; CV: $\beta = 0.12$, $t(147) = 1.38$, p = .17). However, unlike Experiment 1, sensitivity to similarity also predicted measures of strategy with medium effect sizes (criterion: $R^2 = 0.11$; RT mean: $R^2 = 0.18$) such that higher T-N similarity was associated with a more liberal response strategy (greater bias to respond) and faster RTs.

Sensitivity to similarity index and the TOVA Given that individual differences in sensitivity to exemplar-based T-N similarity predicted with measures of ability on the gradCPT, the primary goal of this experiment was to see if sensitivity to exemplar-based T-N similarity index could predict performance on an attention task with minimal perceptual processing demands as well. Much like the gradCPT, measures of task ability on the TOVA (d prime and RT SD) were computed for each participant. First, for participants who completed the gradCPT and the TOVA during the same testing session (N = 138), two second-level regression analyses were run with gradCPT-derived sensitivity to exemplar-pixel and exemplar-GIST T-N similarity as regressors predicting either TOVA derived d prime or RT variance (RT SD). Both of the models were significant, albeit with small effect sizes (d prime: $R^2 = 0.07$, $F(2, 135) = 5.39$, p = .006; RT variance: $R^2 = 0.07$, $F(2, 135) = 4.75$, p = .01). In other words, individuals who were more sensitive to exemplar-based similarity were less variable in response times and more accurate overall on the TOVA. To determine if this association was attributable to a stable cognitive characteristic or a more temporary cognitive state, we identified participants who had been given the TOVA on a different day from the gradCPT. If the association was the result of a more consistent trait, we would predict the association should remain when the data were collected on different days. For these participants (n = 90) the association was nearly identical with medium effect sizes for the model predicting d prime ($R^2 = 0.05$, $F(2, 88) = 2.49$, p = .088) as well as the model predicting RT variance ($R^2 = 0.09$, $F(2, 88) = 4.26$, p = .017). Finally, to determine if the measures of gradCPT ability related to the analogous measures in the TOVA, each participant's d prime and RT SD from the gradCPT were correlated with the corresponding d' and RT variance measured from the TOVA. This revealed a strong correlation between these analogous task measures (same session d prime: r = 0.42, p < .001; cross-session d prime: r = 0.29, p < .005; same-session RT variance: r = 0.53, p < .001; cross-session RT variance: r = 0.42, p < .001).

Experiment 2 discussion

In Experiment 4, the gradCPT was given to participants as part of a large battery of other tasks, including the TOVA—a clinically-used CPT that uses simple visual stimuli. This experiment furthered the exploration of the relationship between sensitivity to T-N similarity and sustained attention by examining whether high sensitivity to exemplar-based similarity reflected (1) enhanced, attention-independent, perceptual processing ability that consequently reduced the attentional demands of the gradCPT or (2) enhanced perceptual processing and encoding that resulted from maintaining an optimal attentional template for tasks with and without complex perceptual demands. To this end, we fit sensitivity to T-N similarity derived from the gradCPT with measures of sustained attention performance observed from the same participants tested on a perceptually simple continuous performance task (the TOVA) and found that individuals who were more sensitive to exemplar-based T-N similarity were better at both the gradCPT and the TOVA with larger d primes and lower RT variability (CV for the gradCPT and RT SD for the TOVA). Because the TOVA, unlike the gradCPT, used two simple stimuli, the demands on perceptual processing and visual working memory were minimal. Therefore, any attention-independent perceptual processing advantage that a given participant may have had on the gradCPT would be minimized when the perceptual demands were trivial. Because bottom-up perceptual processing ability was ruled out as the sole source of the relationship between sensitivity to exemplar-based similarity on the gradCPT and performance on the TOVA, the underlying cognitive processes driving this relationship are likely to be, at least in part, attentional in nature.

Additionally, we set out to replicate the previously observed relationship between individual differences in sensitivity to T-N similarity and more standard measures of gradCPT performance in this lab-based study[4] with a substantial sample size and cohort. Like Experiment 1, we found that participants who were more sensitive to exemplar-based T-N similarity were more accurate (higher d prime) and had less variable RTs (lower CV). Replicating this association expands upon Experiment 1 in two important ways. First, these participants performed the gradCPT in the lab providing a replication of the results in a more controlled setting. Next, because the participants were selected from the veteran population, we demonstrated that this effect generalizes to different populations. The association between sensitivity to similarity and measures of gradCPT strategy, however, is novel and may be related to sample-specific attributes (e.g., greater prevalence of PTSD or mild TBI).

Finally, the TOVA, while simple, has been assessed for reliability and validity and has been shown to be strongly predictive of attentional dysfunction (e.g., ADHD). Therefore, the fact that variables on the gradCPT are predictive of results on the TOVA further validates the link between measures of gradCPT performance and the ability to sustain attention more generally.

General discussion

In two experiments we explored the relationship between individual differences in sustained attention ability and the construction and maintenance of an attentional template during a sustained attention task. To do this, we used T-N similarity to probe both the content and the strength of attentional templates by examining the effects of T-N similarity on reaction times and accuracy on the gradCPT—a go/no-go task where participants must decide whether a presented image is a city or mountain. Specifically, we established (1) that at least some of the trial-to-trial variance in RTs and errors could be attributable to behavioral interference from T-N similarity; (2) that the exemplar-based measures of T-N similarity—based on pixel intensity and to a lesser degree GIST descriptor features—outperformed the measure of T-N similarity assuming more abstract category-based attributes; (3) that individual differences in the sensitivity to the exemplar-based measures of T-N similarity were associated with individual differences in measures of attentional ability (d prime and RT variance) such that those with greater sensitivity to T-N similarity displayed greater attentional ability; and (4) that this relationship cannot be entirely explained by individual differences in perceptual processing ability.

The contents of the attentional template We distinguished between three types of image features used to compute T-N similarity—exemplar-pixel, exemplar-GIST, and category-attribute. We hypothesized that the exemplar-based measures were consistent with the attentional template containing exemplar-specific image features that were rapidly learned and stored in memory and participants classified incoming images based on these features. The category-based measure was consistent with an attentional template containing more abstract previously-learned features used to classify any city or mountain image upon first viewing. We found that the exemplar-based similarity measures (using either pixel intensities and to a lesser extent GIST descriptors) best predicted the RTs to city stimuli suggesting that the recently learned exemplar-specific features were influencing gradCPT performance.

Importantly, participants who were more sensitive to these exemplar-based T-N similarity measures had greater overall ability at the task as measured by greater discrimination ability (d prime) and lower RT variance. Participants who were more sensitive to the category-based T-N similarity measure showed no such relationship. Crucially, this suggests that the exemplar-based feature matching procedure—given the particular task and stimulus sets examined here—was optimal and associated with better task performance.

Given the clear distinction between how the exemplar and category-based feature matching processes related to performance, one may wonder what the source of individual

[4] One notable difference between the TestMyBrain version of the gradCPT used in Experiment 1 and the lab-based version of the gradCPT used in Experiment 2 is the length—specifically, the lab-based version has roughly twice as many trials as the TestMyBrain version. As a result, participants had greater exposure to each exemplar in the lab-based experiment. To see if this greater exposure influenced the strength or nature of T-N similarity we computed the average sensitivity to T-N similarity from the first half of the experiment (roughly matching the magnitude of exemplar exposure from the TestMyBrain gradCPT) and compared it to the average computed from the second half. We found that the sensitivity to T-N similarity for all three measures were remarkably consistent across halves with the average β coefficients for each half for the exemplar-pixel measure being (1st half: $\beta = 0.048$; 2nd half: $\beta = 0.042$), exemplar-GIST (1st half: $\beta = 0.038$; 2nd half: $\beta = 0.047$), and category-attribute (1st half: $\beta = -0.006$; 2nd half: $\beta = 0.004$). This suggests that individuals who develop attentional templates that leverage exemplar-specific features do so with relatively few exposures. This is consistent with studies of statistical learning for visual objects that show such learning can occur with few exposures (Turk-Browne, Scholl, Chun, & Johnson, 2009).

differences along these dimensions were. Why, for example, would one rely on category-level features if exemplar-level features were clearly superior? One possible source of such variation may be due to some form of enhanced visual-perceptual processing or visual memory that enables certain participants to make better use of the exemplar-specific features. Conflicting with this notion, however, was the finding that participants who were the most capable at using exemplar-based features on the gradCPT (as shown by increased sensitivity to exemplar-based T-N similarity) also displayed superior performance on the TOVA—a CPT involving two simple stimuli that left little room for a perceptual advantage to make a difference. To resolve this apparent conflict, we posit that the stable characteristic was the ability to sustain attention such that the participants who sustained an optimal level of attention to the gradCPT were superior at maintaining optimal attentional templates during each of these tasks.

The notion that an optimal attentional template is associated with increased interference from exemplar-based T-N similarity (i.e., sensitivity to T-N similarity) could be considered counterintuitive as one might expect that greater interference from T-N similarity would lead to poorer performance overall for two reasons. First, on a narrow scale, increased sensitivity to T-N similarity is observed in virtue of relative increases in RT variance and errors. Second, there is compelling evidence that elevated interference from T-N similarity is an indicator of a sub-optimal attentional template and optimizing an attentional template through training sharpens and separates the target features from the non-target features (Geng et al., 2017), effectively decreasing the degree of T-N similarity. To reconcile our results with the issues outlined above, we posit the following explanation. When performing a task that requires identifying or categorizing stimuli (like the city/mountain categorization in the gradCPT), it is beneficial to maximize the number of category discriminating features stored in an attentional template. In this version of the gradCPT, there were multiple, but repeating, exemplars within each category. Consequently, both category and exemplar features could be leveraged to make the perceptual decision. Since only sensitivity to exemplar-based measures of T-N similarity were associated with task ability, we propose that this sensitivity indexed the degree to which a given participant was leveraging exemplar features towards making the city/mountain decision. Participants who were more capable at leveraging exemplar features were better at the perceptual discrimination but also more susceptible to interference from exemplar-based T-N similarity. As Experiment 2 demonstrated, sustained attention ability and not perceptual processing ability was the primary determinant of the degree to which a participant could leverage these exemplar features.

Finally, the influence of exemplar-based T-N similarity has further implications for the capacity of visual memory for photographs of scenes. Given that participants were sensitive to similarity using pixel-based features of particular scene exemplars, the scene images must be encoded in memory with enough local detail so that these pixel-based features approximate properties of the stored features used for matching and identifying the incoming stimulus. Participants, therefore, had to maintain local image details for 20 images (ten city and ten mountain scenes) in working memory in order to use this information during the task. While surprising, other studies have shown that exemplar-specific visual information can be encoded in visual memory with remarkable speed and accuracy (Konkle, Brady, Alvarez, & Oliva, 2010a, 2010b; Standing, 1973). Further studies could take advantage of this T-N similarity effect and use increasing numbers of scenes to test the capacity and even time course of influence for each scene photograph.

The content of scene representations Our results point to the exemplar-based feature matching as being an important process in meeting the task demands of the gradCPT. Because the stimulus exemplars were stored (at least temporarily) in memory, how we characterized the image features in computing the exemplar similarity measures served as a window into the features that composed these newly learned, exemplar-level representations. We found that both pixel-based and GIST descriptor-based image features produced similarity values that influenced performance. These results are consistent with two possibilities. First, GIST and pixel similarity reflect image representations from separate levels of scene processing and features from both of these representations are used to classify incoming stimuli. This is consistent with numerous accounts of visual scene processing that posit that gist representations are extracted early on in the processing stream and more local scene details are processed later (Hochstein & Ahissar, 2002; Oliva & Torralba, 2001; but see Goldzieher, Andrews, & Harris, 2016). Alternatively, GIST and pixel similarity could capture unique aspects of the same representation of scenes such that the feature sets that compose exemplar-level scene representations reflect some combination of image properties that both pixel and GIST similarity partially capture. Further studies could use fMRI to analyze patterns of activation in response to scenes presented in the gradCPT to determine if pixel and GIST-based representations share the same neural substrate.

Regardless of whether pixel and GIST similarity reflect one, two, or more types of image representations, one would need to do further research to see whether or not pixel intensity values or the GIST descriptor are the best characterizations of the levels of representations in scene processing or whether they merely capture components of the true underlying mental representation. For example, one could explicitly test properties of scene structure like layout, clutter, expanse, and navigability to determine if the influence of either GIST or pixel-similarity described in these experiments could be better characterized in terms of more interpretable dimensions of similarity (e.g., Choo & Walther, 2016; Harel, Kravitz, & Baker, 2013; Kravitz, Peng, & Baker, 2011; Park, Brady, Greene, & Oliva, 2011; Walther et al., 2009; Watson, Hartley, & Andrews, 2014).

Note that the relative importance of pixel-like, GIST-like, or category-based features likely depends on the task at hand. In the experiments presented in this paper, only 20 exemplars were shown and each exemplar was presented multiple times. This means that while exemplar-level information could be exploited in order to complete this task, the importance of exemplar-level information may not generalize to the underlying representations that are used to categorize novel scenes. Indeed, we would predict that a version of the gradCPT where stimuli do not repeat would greatly diminish the importance of exemplar-level similarity and may boost the importance of category-based features as those would be the features that likely mediate the categorization of unseen scenes. Given this caveat, T-N similarity would still provide a window into underlying scene representations, but one would need to stipulate how context influences the underlying representation. For example, the fact that the gradCPT has repeating exemplars that gradually fade in and out may encourage the use of exemplar-based pixel features because such features are available and reliable. If the scene is presented for a short duration without any fading, GIST features—which have been shown to be extracted rapidly and exploit information distributed across the whole image (e.g., Greene & Oliva, 2009)—may become more important. Overall, future studies could use this method to probe the scene representations that enable categorization under different conditions and at different levels of specificity. This is one of the primary goals within the visual scene recognition literature (for review see Malcolm, Groen, & Baker, 2016).

Finally, it would be interesting to know whether the results regarding pixel and GIST image features and exemplar-based T-N similarity are domain specific to scene processing. Would other categories of objects (e.g., faces) result in the similarity structure of both pixel and GIST image features predicting performance with a ratio similar to the scene stimuli or would the relative contribution of each change as a function of object domain? Furthermore, perhaps scene images are encoded in memory in a way that is fundamentally different from other domains of objects like letters or digits. While the goal of letter and digit processing is identifying a particular stimulus despite the many possible font changes that will alter the exemplar-specific features, with scenes, often it is the exemplar features—not the category (basic level) features—that will be useful at any given moment. To illustrate, only exemplar-specific information indicates whether you are looking at a city street in New York City versus Boston or whether a given forest has a navigable path or not. Therefore, while the visual processing goal of certain domains of objects (e.g., letters and numbers) may be to abstract away from the visual details in service of the flexible processing of object identity, exemplar-specific visual details of scenes provide critical landmark and navigability information and therefore scene processing may default to exemplar-level encoding even when it is not necessary.

Implications for tasks measuring sustained attention One practical implication of these results is that fluctuations in performance that were once assumed to be the result of intrinsically-driven phenomena (e.g., sustained attention, vigilance, inhibitory control) may instead be due, in part, to extrinsically driven phenomenon like T-N similarity. This demonstrates that when making inferences about the moment-to-moment fluctuations in attentional states it is important to factor in possible stimulus-driven factors that could make a given trial more difficult than the next regardless of one's attentional state. One way to correct for such trial-to-trial inequalities in difficulty would be to derive, based on T-N similarity and other task-driven phenomena like linear trend (Esterman, Noonan, Rosenberg, & Degutis, 2013; Rosenberg, Noonan, Degutis, & Esterman, 2013) or trial order effects, a difficulty time course for each run of the gradCPT. This difficulty time course would predict, based on stimulus or trial order information alone, the relative difficulty of any given trial. Including this measure in any measure that predicts the trial-to-trial fluctuations in attention would allow the isolation of performance fluctuations that are truly internally-driven by the participants' state of mind, like mind-wandering, not externally driven by the inequality of stimulus-driven difficulty across trials. Furthermore, such a measure of task difficulty could be applied to studies that have tracked neural activation fluctuations during different states of sustained attention (Esterman et al., 2013; Esterman, Poole, Liu, & DeGutis, 2016). Previously, such neural fluctuations have been assumed to be driven by intrinsic attentional states, but including this task-difficulty measure could identify the extrinsic, stimulus-driven contribution to these activation fluctuations.

Finally, it is worth highlighting the fact that these results were only tested in go/no-go tasks where the stimuli gradually fade from one to the next. As a result, information about the identity of an upcoming, yet-to-be-classified image accrues gradually instead of appearing all at once. How this interacts with T-N similarity remains to be seen but it could be that the gradual nature of the task facilitates the exemplar-based matching process that appears to be happening in the present experiments. While stimulus image similarity clearly makes an important contribution to rapid sequentially presented stimuli in RSVP and priming paradigms (Kinoshita & Kaplan, 2008; Visser, Davis, & Ohan, 2009; Ward, Duncan, & Shapiro, 1997), it would be important to compare the findings from this study with results from an "abrupt" onset CPT, like the SART (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997).

Conclusion

The current study presented two experiments that described a novel effect in which exemplar-based T-N similarity influenced RTs and error rates in a sustained attention task (the gradCPT) across thousands of participants. Individual differences in the sensitivity to this exemplar-based T-N similarity were most strongly and consistently associated with measures of discrimination ability and RT variance. This pattern supports a model of cognitive processing during the gradCPT whereby participants store features of stimulus exemplars and these exemplar representations play an important role in facilitating the recognition of incoming stimuli. This suggests that individuals who are more capable of representing, leveraging, and maintaining exemplar-based features perform better on the task. Further, these individuals are better at sustaining attention more generally since sensitivity to exemplar-based similarity on the gradCPT correlated with performance on the TOVA—a CPT with only trivial perceptual elements. Sensitivity to category-based T-N similarity, however, was unrelated to task performance.

This study also can serve as the foundation for a number of future studies. Questions about the nature of visual representations stored in memory and how they interact with image domains (e.g., faces vs. scenes) can be explored. Additionally, T-N similarity can be used to dissociate internal and external (stimulus driven) contributions to fluctuations in sustained attention. Finally, the cognitive and neural mechanisms that account for the individual differences in sensitivity to similarity need to be further clarified and explored as they can be potentially useful in characterizing intrinsic attentional or perceptual ability as well as the precision of visual memory.

Acknowledgements This work was supported by the US Department of Veteran Affairs through a Clinical Science Research & Development Career Development Award (grant number 1IK2CX000706-01A2) to M.E and the Translational Research Center for TBI and Stress Disorders (TRACTS), a VA Rehabilitation Research and Development Traumatic Brain Injury Center of Excellence (B9254-C).

References

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience*, *31*(25), 9315–9322. https://doi.org/10.1523/JNEUROSCI.1097-11.2011

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570. https://doi.org/10.1073/pnas.1402594111

Cheyne, J. A., Carriere, J. S. A., & Smilek, D. (2009). Absent minds and absent agents: Attention-lapse induced alienation of agency. *Consciousness and Cognition*, *18*(2), 481–493. https://doi.org/10.1016/j.concog.2009.01.005

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8719–24. https://doi.org/10.1073/pnas.0900234106

Choo, H., & Walther, D. B. (2016). Contour junctions underlie neural representations of scene categories in high-level human visual cortex: Contour junctions underlie neural code of scenes. *NeuroImage*, *135*, 32–44. https://doi.org/10.1016/j.neuroimage.2016.04.021

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hilsdale, NJ: Lawrence Earlbaum Associates.

Conners, C. K. (2000). Continuous performance test II: Technical guide and software manual. Toronto, ON: Multi-Health Systems.

deBettencourt, M. T., Norman, K. A., & Turk-Browne, N. B. (2017). Forgetting from lapses of sustained attention. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1309-5

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Esterman, M., Noonan, S. K., Rosenberg, M., & Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*(11), 2712–2723. https://doi.org/10.1093/cercor/bhs261

Esterman, M., Poole, V., Liu, G., & DeGutis, J. (2016). Modulating Reward Induces Differential Neurocognitive Approaches to Sustained Attention. *Cerebral Cortex*, 1–11. https://doi.org/10.1093/cercor/bhw214

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained Attention Across the Life Span in a Sample of 10,000: Dissociating Ability and Strategy. *Psychological Science*. https://doi.org/10.1177/0956797615594896

Geng, J. J., DiQuattro, N. E., & Helm, J. (2017). Distractor Probability Changes the Shape of the Attentional Template. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000430

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9

Goldzieher, M. J., Andrews, S., & Harris, I. M. (2016). Two scenes or not two scenes: The effects of stimulus repetition and view-similarity on scene categorization from brief displays. *Memory & Cognition*, 1–14. https://doi.org/10.3758/s13421-016-0640-9

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001

Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing Visual Scenes in Cortex: Gradients of Object and Spatial Layout Information. *Cerebral Cortex*, *23*(4), 947–957. https://doi.org/10.1093/cercor/bhs091

Head, J., & Helton, W. S. (2012). Natural scene stimuli and lapses of sustained attention. *Consciousness and Cognition, 21*(4), 1617–1625. https://doi.org/10.1016/j.concog.2012.08.009

Helton, W. S., & Russell, P. N. (2011). Feature absence-presence and two theories of lapses of sustained attention. *Psychological Research, 75*(5), 384–392. https://doi.org/10.1007/s00426-010-0316-1

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Hout, M. C., Godwin, H. J., Fitzsimmons, G., Robbins, A., Menneer, T., & Goldinger, S. D. (2015). Using multidimensional scaling to quantify similarity in visual search and beyond. *Attention, Perception, & Psychophysics*, 3–20. https://doi.org/10.3758/s13414-015-1010-6

Hout, M. C., & Goldinger, S. D. (2015). Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, *77*(1), 128–149. https://doi.org/10.3758/s13414-014-0764-6

Kinoshita, S., & Kaplan, L. (2008). Priming of abstract letter identities in the letter match task. *Quarterly Journal of Experimental Psychology (2006)*, *61*(12), 1873–85. https://doi.org/10.1080/17470210701781114

Konkle, T., Brady, T. F., Alvarez, G. A, & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology. General*, *139*(3), 558–78. https://doi.org/10.1037/a0019165

Konkle, T., Brady, T. F., Alvarez, G. A, & Oliva, A. (2010b). Scene Memory Is More Detailed Than You Think. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places. *Journal of Neuroscience*, *31*(20), 7322–7333. https://doi.org/10.1523/JNEUROSCI.4588-10.2011

Leark, R. A., Greenberg, L. K., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2007). Test of Variables of Attention: Professional Manual. Los Alamitos: The TOVA Company.

Lee, J., & Geng, J. J. (2017). Idiosyncratic Patterns of Representational Similarity in Prefrontal Cortex Predict Attentional Performance. *The Journal of Neuroscience*, *37*(5), 1257–1268. https://doi.org/10.1523/JNEUROSCI.1407-16.2016

Lippa, S. M., Fonda, J. R., Fortier, C. B., Amick, M. A., Kenna, A., Milberg, W. P., & McGlinchey, R. E. (2015). Deployment-Related Psychiatric and Behavioral Conditions and Their Association with Functional Disability in OEF/OIF/OND Veterans. *Journal of Traumatic Stress, 28*(1), 25–33. https://doi.org/10.1002/jts.21979

Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making Sense of Real-World Scenes. *Trends in Cognitive Sciences*, *20*(11), 843–856. https://doi.org/10.1016/j.tics.2016.09.003

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching Through the Hierarchy: How Level of Target Categorization Affects Visual Search. *Visual Cognition*, *20*(10), 1153–1163. https://doi.org/10.1080/13506285.2012.735718

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175. https://doi.org/10.1023/A:1011139631724

Oliva, A., & Torralba, A. (2006). Chapter 2 Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36. https://doi.org/10.1016/S0079-6123(06)55002-2

Parasuraman, R., de Visser, E., Clarke, E., McGarry, W. R., Hussey, E., Shaw, T., & Thompson, J. C. (2009). Detecting threat-related intentional actions of others: Effects of image quality, response mode, and target cuing on vigilance. *Journal of Experimental Psychology: Applied, 15*(4), 275–290. https://doi.org/10.1037/a0017132

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *The Journal of Neuroscience*, *31*(4), 1333–40. https://doi.org/10.1523/JNEUROSCI.3885-10.2011

Patterson, G., & Hays, J. (2012). SUN attribute database: Discovering, annotating, and recognizing scene attributes. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2751–2758. https://doi.org/10.1109/CVPR.2012.6247998

Patterson, G., Xu, C., Su, H., & Hays, J. (2014). The SUN attribute database: Beyond categories for deeper scene understanding. *International Journal of Computer Vision*, *108*(1–2), 59–81. https://doi.org/10.1007/s11263-013-0695-z

Peelen, M. V., & Caramazza, A. (2012). Conceptual Object Representations in Human Anterior Temporal Cortex. *Journal of Neuroscience*, *32*(45), 15728–15736. https://doi.org/10.1523/JNEUROSCI.1953-12.2012

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 12125–30. https://doi.org/10.1073/pnas.1101042108

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1995). Similarity determines the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 653–662. https://doi.org/10.1037/0096-1523.21.3.653

Riley, E., Esterman, M., Fortenbaugh, F. C., & DeGutis, J. (2017). Time-of-day variation in sustained attentional control. *Chronobiology International*, 1–9. https://doi.org/10.1080/07420528.2017.1308951

Riley, E., Okabe, H., Germine, L., Wilmer, J., Esterman, M., & DeGutis, J. (2016). Gender Differences in Sustained Attentional Control Relate to Gender Inequality across Countries. *Plos One*, *11*(11), e0165100. https://doi.org/10.1371/journal.pone.0165100

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rosenberg, M., Noonan, S., Degutis, J., & Esterman, M. (2013). Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. *Attention, Perception & Psychophysics*, *75*(3), 426–39. https://doi.org/10.3758/s13414-012-0413-x

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, *62*(10), 1904–1914. https://doi.org/10.1080/17470210902853530

Sinha, P., & Russell, R. (2011). A perceptually based comparison of image similarity metrics. *Perception*, *40*(11), 1269–1281. https://doi.org/10.1068/p7063

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition, 13*(4), 657–690. https://doi.org/10.1016/j.concog.2004.06.003

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. https://doi.org/10.1080/14640747308400340

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–45. https://doi.org/10.1162/jocn.2009.21131

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81–92. https://doi.org/10.1167/5.1.8

Visser, T. A. W., Davis, C., & Ohan, J. L. (2009). When similarity leads to sparing: Probing mechanisms underlying the attentional blink. *Psychological Research*, *73*(3), 327–335. https://doi.org/10.1007/s00426-008-0155-5

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience*, *29*(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009

Ward, R., Duncan, J., & Shapiro, K. (1997). Effects of similarity, difficulty, and nontarget presentation on the time course of visual attention. *Perception & Psychophysics*, *59*(4), 593–600. https://doi.org/10.3758/BF03211867

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410. https://doi.org/10.1016/j.neuroimage.2014.05.045

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9*(7), 971–978. https://doi.org/10.1038/nn1727

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970

Xiao, J., Hays, J., Russell, B. C., Patterson, G., Ehinger, K. A, Torralba, A., & Oliva, A. (2013). Basic level scene understanding: categories, attributes and structures. *Frontiers in Psychology*, *4*(August), 506. https://doi.org/10.3389/fpsyg.2013.00506

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*(16), 2095–2103. https://doi.org/10.1016/j.visres.2009.05.017

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning Deep Features for Scene Recognition using Places Database. *Advances in Neural Information Processing Systems 27 NIPS*, *14*(6), 487–495. https://doi.org/10.1162/153244303322533223