Brain and DCNN representational geometries predict variability in conscious access

Daniel Lindh (p.j.d.lindh@uva.nl)[1, 2], Ilja Sligte (i.g.sligte@uva.nl)[1], Kimron Shapiro (k.l.shapiro@bham.ac.uk)[2], Ian Charest (i.charest@bham.ac.uk)[2]

1. Department of Psychology, University of Amsterdam, Amsterdam, Netherlands
2. School of psychology, University of Birmingham, Birmingham, United Kingdom

Abstract

When two targets (T1 and T2) are presented in a rapidly sequentially-presented stream of distractors, subjects often show a clear deficiency to report T2 when presented 200-500 ms after T1. This effect is known as the Attentional Blink (AB). Using the AB as a method to quantify the probability of conscious access, we investigate why some images seem to rise to consciousness more readily. By defining the representational relationships between images using fMRI and CNNs, we show that images that are distinct in high-level representations are more resilient to the AB effect, while low-level similarity to other images increase the probability of conscious access. These results were replicated using representational geometries derived from both functional Magnetic Resonance Imaging (fMRI) and Convolutional Neural Network (CNN). This provides additional parallels between the hierarchical complexity of CNNs trained on object classification and the human visual ventral stream, with CNN and brain representations predicting behaviour in a similar way.

Keywords: Attention; Working memory; CNN; fMRI

Introduction

The attentional blink (AB) (Raymond, Shapiro, & Arnell, 1992) is one of the most studied phenomena in the attention literature. In the AB, two targets (T1 and T2) are embedded in a rapidly sequentially-presented stream of distractors. When T2 is positioned 200-500 ms after T1, subjects often show an impaired ability to report T2 relative to when it is presented with a longer interval. Many of the prominent theoretical frameworks assume a late bottleneck (Dux & Marois, 2009), leaving most of the perceptual processing of T2 intact without conscious access. This makes the AB a paragon task to investigate the events leading up to conscious processing. In a previous study, we showed that there are substantial variability in the degree to which different stimulus categories are affected by AB (Lindh, Sligte, Assecondi, Shapiro, & Charest, 2019). However, more work is needed to understand why certain objects are more affected by the attentional blink window.

One often ignored aspect of AB is the relationship between the targets. For example, how does the particular feature processing of T1 affect the processing of T2? One inherent problem is how you define similarity between two images in a neurally relevant way. Studies on repetition blindness have used the same item but from a different angle (Buffat, Plantier, Roumes, & Lorenceau, 2013) or category (Sy & Giesbrecht, 2009) as a proxy for similarity. Here, we turn to computational models and brain activity patterns using representational similarity analyses to estimate model and brain representational geometries (Kriegeskorte, 2009; Kriegeskorte & Kievit, 8/2013), and extract continuous metrics of similarity between targets at different levels of information processing.

Lindh et al. (2019) computed target-target similarity based on different layers of a deep convolutional neural network (DCNN), and showed facilitation effects. When the two targets were similar in their mid-level visual feature activations, T1 facilitated processing of T2. This is in contrast to earlier studies of repetition blindness (Buffat et al., 2013; Kanwisher, 1987), where task-relevant similarity leads to interference. This suggests that representational geometries at the level of the brain might influence conscious access in object recognition, depending on context or task-relevance. Furthermore, this suggests that an item's representational signature influences it's propensity to be consciously reported. Items that have neurally distinct representations in task-relevant areas of the brain should be better processed than items with overlapping neural representations. Alternatively, one could imagine that items that share low-level physical properties could benefit from similar representations in early visual areas of the brain. In this study, we ask if representational geometry metrics can explain why some objects are more often consciously perceived. We ask two specific related questions: 1) how does target-target similarity in the brain affect conscious access and 2) how does object representations idiosyncrasies predict trial-by-trial and inter-individual variability in conscious access.

Methods

20 participants (mean age = 23, 13 females) participated in the study. Participants completed 4 sessions of the attentional blink task and two sessions of functional magnetic resonance imaging (fMRI). Three participants did not complete all conditions and thus were thus excluded from further data analyses. All participants provided informed consent, and were compensated for their time (at the rate of 10 euros an hour for behavioural and 20 euros an hour for fMRI).The experiment was approved by the ethics committee at the University of Amsterdam.

Stimuli

The visual objects presented in both tasks consisted of forty non-isolated natural scene images (twenty animals, twenty non-animals), with central objects depicting bodies, faces, food, objects and places. All experiments were programmed using Psychtoolbox

Version 3 in MATLAB (The MathWorks, Inc, Natick, Massachusetts, United States).

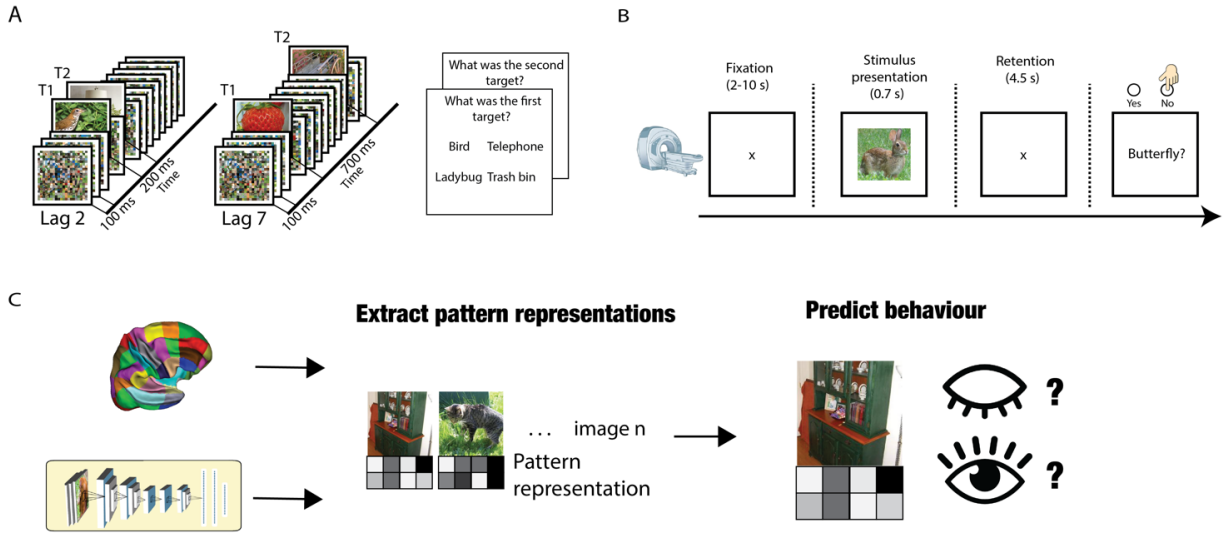

Figure 1: A) Pictorial representation of the Attentional Blink task. B) Working memory task in the scanner. C) Analysis steps.

Attentional Blink task

Participants were comfortably sat in front of a 19" monitor positioned at a distance of 60 cm. Targets and distractors were displayed in the centre of the screen subtending 5 degrees of visual angle on a constant grey background. At the beginning of a trial, participants attended to a white fixation cross for 1.25s. This was followed by a stream of 19 images (17 distractors and 2 targets). Images were shown for 16.7 ms with a stimulus onset asynchrony (SOA) of 100 ms. The first target (T1) was randomly presented at position 4, 5 or 6 in the stream and the second target (T2) was presented either two (lag 2) or seven (lag 7) items further away. The distractors were scrambled image composites randomly created from the stimulus set (similar to (Marois, Yi, & Chun, 2004). After each trial, participants were prompted with a response menu for T1, and asked to choose which of the four possible words corresponded to the first target. Following this, a similar menu was displayed for T2. Attentional Blink Magnitudes were computed as the difference between T2 performances in the lag 2 and lag 7 conditions.

Working memory task

The same natural visual objects were used in the working memory task completed during fMRI scanning. Images were shown with a 5 degree visual angle through a back projected screen visible via a head mounted mirror. Participants were presented with an image for 500 ms, followed by 4000 ms of retention period. Participants were then prompted with a word and asked to respond yes or no, using the corresponding button under left or right index finger if the word matched the semantic content of the image.

fMRI preprocessing

fMRI data was converted to BIDS (Gorgolewski et al., 2016), before being pre-processed using fMRIPrep (Esteban et al., 2019). EPI images were corrected for spatial alignment, and normalised to the Montreal Neurological Institute ICBM template space (Mazziotta et al., 2001). Beta weights for each stimulus condition were obtained using GLMdenoise (Kay, Rokem, Winawer, Dougherty, & Wandell, 2013; Charest, Kriegeskorte, & Kay, 2018) and converted into t-patterns for pattern similarity analyses. Regions of interest (ROI) were defined using the Glasser atlas parcellations (Glasser et al., 2016). Pattern similarity was measured using Pearson's correlation across all pairs of condition t-patterns within each ROI.

Results

The purpose of our experiment was to account for why some objects are more frequently reported correctly in the attentional blink. We did so by calculating each image's similarity to all other images, and correlated that with the attentional blink magnitude (ABM) for each image (see methods). The ABM was calculated by subtracting each image's T2 lag 7 performance from the lag 2 performance.

Behaviour

Subjects showed a higher T2 performance at lag 7 ($M = 0.93$, $SD = 0.068$) in comparison to lag 2 ($M = 0.823$, $SD = 0.05$, $t(15) = -7.79$, $p < 0.001$), indicating the commonly found attentional blink effect.

Similarity and T2 performance

For each trial we calculated the representational pattern similarity between T1 and T2. We then correlated this with T2 performance at lag 2 using Spearman's correlation coefficient (Figure 2). We observed a positive correlation between T2 performance and T1-T2 similarity in V1 ($M = 0.04$, $SD = 0.04$, $t(15) = 3.9$, $p = 0.01$), indicating that when T2 share low-level representational patterns with T1, the processing of T2 is facilitated. Conversely, we observed a negative correlation between T2 performance and T1-T2 similarity

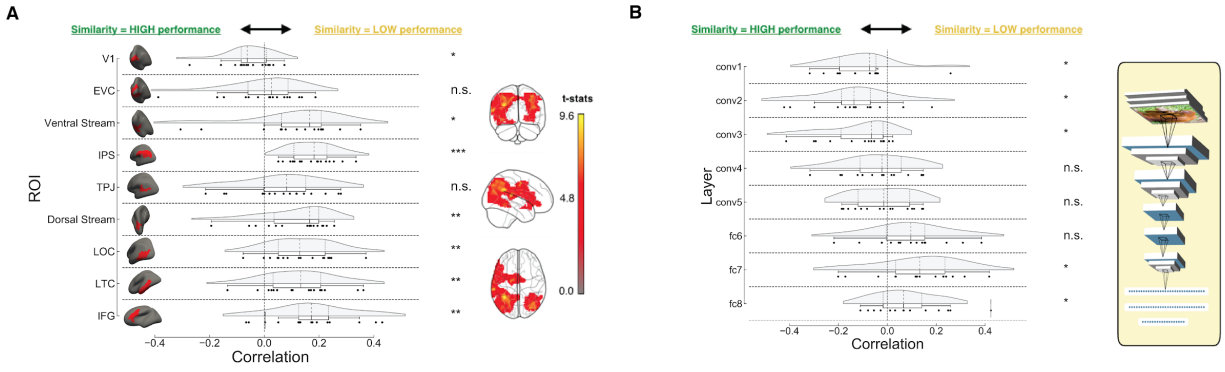

*Figure 3: A) fMRI results. ROIs are ordered to approximate low- to high-level processing. Images that are generally more similar to other images in V1 show a lower attentional blink magnitude (ABM). In contrast, images that share representational patterns with other images in late visual and semantic processing areas are more likely to be blinked. B) DCNN results. Repeating the analysis using DCNN features from different layers, we see the same pattern with negative correlations using low-level visual features and positive correlations using high-level visual features. * = p < 0.05, ** = p < 0.01, *** = p < 0.001.*

in ventral stream (M = -0.08, SD = 0.06, t(15) = -5.19, p = 0.001), IPS (M = -0.08, SD = 0.07, t(15) = -4.76, p = 0.002), TPJ (M = -0.05, SD = 0.05, t(15) = -3.76, p = 0.02), dorsal stream (M = -0.05, SD = 0.05, t(15) = -3.66, p = 0.02), LOC (M = -0.08, SD = 0.06, t(15) = -4.93, p = 0.001) and LTC (M = -0.04, SD = 0.04, t(15) = -3.57, p = 0.03; all p-values are FDR corrected for multiple comparisons). This suggests that high-level representational overlap interferes with processing of T2. Altogether, these findings indicate that T2 processing is affected differently depending on where in the brain it interacts with the processing of T1.

fMRI similarity and ABM

Building on the finding that target-target similarity affects T2 processing, we set out to investigate if the representational distinctiveness of an image can explain why some images are less likely to be blinked. Based on pattern representations for each ROI, we calculated the average similarity (AS) of one image in respect to all other images. This yielded one value per image, indicating how similar this image is overall to the rest of the image set. Using a searchlight procedure, we correlated each image's AS with that image's ABM. We show a large cluster of positive correlations extending from posterior high-level visual areas to left inferior frontal cortex (Fig 3A). Our a priori defined region of interest (ROI) confirmed this finding, showing robust high correlations in ventral stream (M = 0.11, SD = 0.17, t(1615 = 2.54, p = 0.029), inferior parietal cortex (M = 0.17, SD = 0.08, t(15) = 8.42, p < 0.001), dorsal visual stream (M = 0.11, SD = 0.12, t(15) = 3.55, p = 0.005), LOC (M = 0.13, SD = 0.12, t(15) = 4.57, p = 0.0015), lateral temporal cortex (M = 0.12, SD = 0.13, t(15) = 3.8, p = 0.004) and inferior frontal cortex (M = 0.17, SD = 0.15, t(15) = 4.4, p = 0.001). This positive correlation indicates that images that are less distinct in high-level processing areas also are more likely to be blinked. In addition, we also find a negative correlation between overall similarity and ABM in V1 (M = -0.05, SD = 0.08, t(15) = -2.66, p = 0.02; all p-values are corrected for multiple comparisons using FDR). This is in agreement with our earlier finding that low-level, task-irrelevant, similarity facilitates T2 performance.

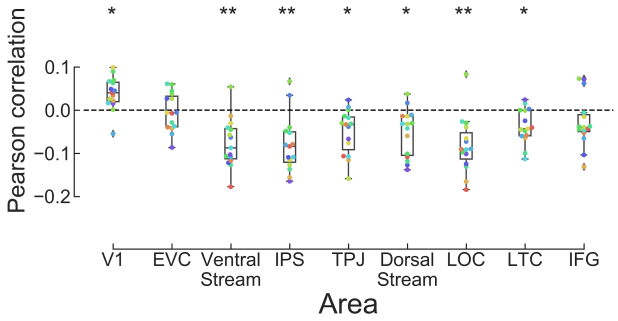

*Figure 2: Correlation between T1-T2 similarity and T2 performance per subject for each ROI. Individual dots indicate subjects. * = p < 0.05, ** = p < 0.01, *** = p < 0.001*

DCNN similarity and ABM

Using the layers of the DCNN to approximate a progression of complexity (Güçlü & van Gerven, 2014), we corroborated the results from the fMRI ROIs. We observed a negative correlation between image overall similarity and ABM in layer conv1 (M = -0.1, SD = 0.13, t(16) = -2.95, p = 0.025), conv2 (M = -0.14, SD = 0.15, t(16) = -3.31, p = 0.022), and conv3 (M = 0.11, SD = 0.13, t(16) = -0.327, p = 0.022). In higher layers of the DCNN, overall similarity correlated positively with ABM, as observed in fc7 (M = 0.13, SD = 0.17, t(16) = 2.86, p = 0.025) and fc8 (M = 0.09, SD = 0.14, t(16) = 2.43, p = 0.046; Fig 2B).

Individual differences

We further developed the idea of representational similarity between objects. In a searchlight procedure, we averaged the similarity of all pairwise comparisons within a subject and correlated that with the subjects' overall attentional blink magnitude. We found that subjects with more similar object representations in right temporo-parietal junction and inferior frontal gyrus are also more vulnerable to the attentional blink (Fig4).

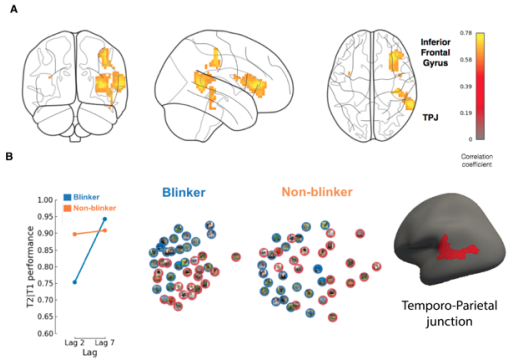

Figure 4: Individual differences. A) A searchlight procedure revealed that subjects with more similar representations in TPJ and IFG are more affected by the AB. B) Example subjects depicting a "blinker" and a "non-blinker". Applying multi-dimensional scaling to the TPJ RDMs, we see a clear modulation of representational richness related to task performance.

Conclusions

In the current study, we measured representational geometries from both fMRI and a DCNN to explain why some objects are more likely to be blinked than others. We show that representational overlap in task-relevant areas (images sharing task-relevant features with other targets) explain substantial trial-by-trial variability in the attentional blink. Moreover, we provide a novel explanation for the variability in conscious processing between different visual objects. Finally, our results suggest that object separation, or representational richness, in the right ventral attentional network (Corbetta, Patel, & Shulman, 2008), is a good predictor of individual differences in the attentional blink.

Altogether our results provide a mechanistic explanation of conscious access in object recognition between trials, images and people.

Acknowledgments

This work was supported by a European Research Council (ERC) Starting Grant ERC-2017-StG 759432 (to I.C.).

References

Buffat, S., Plantier, J., Roumes, C., & Lorenceau, J. (2013). Repetition blindness for natural images of objects with viewpoint changes. *Frontiers in Psychology*, *3*(January), 1–11.

Charest, I., Kriegeskorte, N., & Kay, K. N. (2018). GLMdenoise improves multivariate pattern analysis of fMRI data. *NeuroImage*, *183*, 606–616.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*(2-3), 109–114.

Dux, P. E., & Marois, R. (2009). The attentional blink: a review of data and theory. *Attention, Perception & Psychophysics*, *71*(8), 1683–1700.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 160044.

Güçlü, U., & van Gerven, M. A. J. (2014). *Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Brain's Ventral Visual Pathway*. *35*(27), 10005–10014.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, *27*, 117–143.

Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F., & Wandell, B. A. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Frontiers in Neuroscience*, *7*, 247.

Kriegeskorte, N. (2009). Relating Population-Code Representations between Man, Monkey, and Computational Models. *Frontiers in Neuroscience*, *3*(3), 363–373.

Kriegeskorte, N., & Kievit, R. A. (8/2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412.

Lindh, D., Sligte, I. G., Assecondi, S., Shapiro, K. L., & Charest, I. (2019). *Conscious perception of natural images is constrained by category-related visual features*. https://doi.org/10.1101/509927

Marois, R., Yi, D.-J., & Chun, M. M. (2004). The Neural Fate of Consciously Perceived and Missed Events in the Attentional Blink. *Neuron*, *41*(3), 465–472.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., … Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1412), 1293–1322.

Raymond, J. D., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual procesing in a RSVP task: an attention blink? *Journal of Experimental Psychology*, *18*(3), 849–860.

Sy, J. L., & Giesbrecht, B. (2009). Target-target similarity on the attentional blink: Task-relevance matters! *Visual Cognition*, *17*(3), 1–10.