Investigating a Neural Correlate of Attention as a Metric of Learning

Lauren Hansen-Manguikian

Major Thesis MSc Behavioral and Cognitive Neuroscience

Faculty of Science and Engineering

University of Groningen

Date: 12-08-2022

Author Note

First supervisor: prof. dr. Elkan Akyürek

Second supervisor: prof. dr. Hedderik van Rijn

Daily Supervisor: Güven Kandemir, MSc

This thesis was originally supervised by dr. Berry van den Berg

Disclosures and Acknowledgments: I would like to thank Berry van den Berg for his kind supervision and his great idea in starting this project, as well as express wishes for his speedy recovery. I would like to thank Elkan Akyürek and Güven Kandemir for taking on this project even though it was not theirs. Lastly, I would like to thank all the friends that supported me and gave me feedback on the final drafts: Niklas Erdmann, Felix Schröder, Lars Cordes, Kimberly Wickborn, and Hao Lou.

Contact information: lhansen9393@gmail.com (s3938174)

Abstract

Forming associations between two stimuli, such as a word and its definition, make up a large part of the human learning experience. Our current methods for assessing learning are primarily behavioral (e.g., tests). Here we investigate attention as a possible metric for assessing learning. Research has shown that our attention is easily biased towards the current content of our thoughts, working memory, and anything closely related to that content. It follows that thinking about one object should bias attention towards a related object if an association has been formed between the two. Here we measure attentional bias over the course of an associative learning task. This study aims to use EEG to answer the question: can we use neural correlates of attention to probe the strength of associations stored in memory? Attention was assessed using the N2pc, an event related potential that has been found to reflect the amount and time course of visual attention on an object. Participants engaged in an hour-long learning task in which they form associations between three letter airport codes (e.g., AMS) and their corresponding runway layouts. Neural data of twelve participants were measured using a 64-electrode equidistant EEG system. Results show an increase in the N2pc amplitude as the association between the two stimuli strengthens over the course of the experiment. The onset of the N2pc did not show a significant trend with increased learning. Future applications of this research could be to improve measurements of learning by using neural markers of learning alongside behavioral metrics.

Keywords: N2pc, Attention, Learning, Association learning

Investigating a Neural Correlate of Attention as a Metric of Learning

From a newborn baby trying to find patterns in the world, to students cramming for an exam in school, learning is an essential part of our biological and cultural existence. In a nutshell, learning is the process by which we form an internal model of the outside world (Dehaene, 2020). The brain performs this fundamental act with two essential tools: attention, which allows us to select and amplify relevant information, and memory, which allows us to store that information for later use. The perception of a relevant stimulus is represented in working memory and from there, if it is deemed important enough, it is stored in long term memory. When this information becomes relevant again, usually due to perception of an associated stimulus, the act of recalling the information brings it from long-term memory back into working memory (James, 1890; Sternberg & Sternberg, 2015).

It has been robustly shown that items in working memory bias attention (Hollingworth et al., 2013; Huang & Pashler, 2007). For example, maintaining the color green in working memory will bias one’s gaze to green objects in an unrelated task (Van Moorselaar et al., 2014). Outside of specific lab conditions, the bias is usually adaptive. The contents of working memory are often what is most important at the current moment, thus attention should be drawn to related stimuli. This mechanism becomes relevant when we perform a visual search. With the object of interest in mind, we are more likely to find the object in a distractor-filled environment if our attention is naturally biased towards the content of our thoughts.

Our current understanding is that the act of recalling information brings it from long-term memory to working memory, and the contents of working memory bias our attention. With this reasoning, attention should be more biased towards stimuli that can be recalled, compared to stimuli that were never learnt. In the same vein, the degree of bias should be contingent upon the degree to which we can recall the stimulus. We sought to investigate this hypothesis by measuring a neural correlate for attention (the N2pc) within a learning and recall task. The N2pc is associated with the orienting and time-course of covert attention towards a stimulus (Luck & Kappenman, 2012). To do this we created a lateralized visual search task with one target and three distractors. The target shape can only be identified if it is sufficiently recalled based on a cue. Each trial is, in essence, a multiple-choice question. If the association has been learned between the cue and target shape, the cue will trigger the target shape to be brought from long-term memory into working memory. While the target is on screen in a visual search array, we hypothesized an attentional bias towards the target as measured by the N2pc. We also predicted a continuous relationship: as the association becomes well-learned, the degree of attentional bias towards the target should increase.

Why the N2pc

The N2pc is particularly well-suited as a metric of attentional bias in this task because it has been extensively studied in the context of visual search tasks (Luck et. al, 1993; Luck & Kappenman, 2014; Woodman & Luck, 2003), and more recently in a few learning paradigms (e.g., An et al., 2012; Qu et al., 2017; van den Berg et al., 2019; Oemisch et al., 2017). Here we will shortly introduce visual search processes and where the N2pc fits in.

Visual attentional bias is usually measured in two ways: covert and overt attention. Overt attention is when one’s eyes contain the object in the foveal region, the central area of the retina. Covert attention is attending to an object that is not foveated. While attentional bias can be measured by overt attention (e.g., Van Moorselaar et al., 2014), covert attention reflects the earlier attentional bias to object features that may subsequently draw one’s gaze.

When searching for an item amongst other items, like searching for one’s pencil on a cluttered desk, we use both covert and overt attention. In a natural setting, the act of searching for an item and attending to it can be broken down into a series of steps described by Luck and Kappenman (2012): We first establish a goal for our visual search, for instance, the yellow pencil somewhere on the desk. This is known as the search template, and it is maintained in working memory (Duncan & Humphreys, 1989). The search template guides the search by increasing sensitivity to the features related to the search template (yellow, pencil-shape; Wolfe et al., 1989). Any object in the visual field with these features can trigger a shift in covert attention in the object's direction (Carrasco et al., 2000). The N2pc is most closely associated with this shift of covert attention.

Would a purposeful shift in attention, as indexed by the N2pc, lead to an increase in subsequent memory for that object? Geib et al. (2021) tested this by measuring N2pc in a attention and memory task. In their task, two circles (one pink, one yellow) were shown on the screen, each containing an object. Participants were instructed to attend to the object within the pink (or yellow, counterbalanced between participants) circle. In a second phase of the experiment, they showed participants objects one by one, some of which were shown in the previous task and some were novel to the participants. They asked participants to determine whether they had seen the object earlier in the experiment. They found that the N2pc was larger for subsequently remembered objects than for forgotten ones, indicating that the top-down application of attention to an object increases likelihood of subsequent recognition. This may be due to the fact that shifts in covert attention are associated with increased sensory processing at that location (Luck et al., 1993). This also shows that telling participants to attend to an item on each trial does not lead to uniform attentional shifts toward the target item on each trial, and that the N2pc amplitude can reflect those attentional differences likely to lead to differential memory storage. In essence, the question is not just 'is there attention here or not?' but 'how much attention is being allocated here?'.

This naturally leads to the research question at hand: if the N2pc amplitude can predict how well items will be encoded for the future, can it also predict how well items have been encoded in the past? There are some precedents for using the N2pc to indicate learning. In An et al. (2012) and Qu et al. (2017) participants were trained to identify a single type of stimulus orientation or shape, respectively. When comparing a trained group and an untrained group in the same visual search task, the trained group had a larger N2pc for the trained stimulus type. This promising result indicates the N2pc can reflect training on a specific perceptual stimulus. Other experiments provide a precedent for tracking N2pc changes over course of the learning process. The N2pc has been shown to be strongly modulated by reward (Failing & Theeuwes, 2018), so a number of studies have successfully tracked the N2pc over the course of reward learning. These studies have shown that the N2pc amplitude does increase continuously with learning which stimulus is most likely to produce a reward over the course of a set of trials (van den Berg et al., 2019; Oemisch et al., 2017). This provides evidence that the N2pc amplitude can be a continuous measure of reward learning, but it remains to be seen whether this will also be the case for the learning of new facts.

All this taken together indicates that attention comes into play in two ways with regards to learning: first during the encoding phase, then secondarily during the recall phase as the mechanism by which we recognize relevant stimuli in our environment. Greater N2pc has been successfully linked to better encoding by Geib et al. (2021). This study aims to do the same for the recall phase, by attempting to determine if the N2pc, a well-studied neural correlate of attention, can track the strength of associations stored in memory.

Stimulus Validation Pre-Study

Based on this reasoning, we wanted to create a new set of learning stimuli that would be both novel to participants (unlearned) and ecologically valid (learning real associations). One common type of association learning that humans perform is matching a name to a visual stimulus in our environment. With this logic, airport runway layouts were chosen as novel learning stimuli for multiple reasons: i) airports are named with a 3-letter code, making the names visually similar to control for visually-evoked neural changes and also easy to read quickly; ii) each airport has a unique configuration of runways that can be distinguished from one another with learning; iii) most people have never seen or memorized runway layouts except pilots and airport staff; iv) there exist many airports, giving a large pool of potential stimuli and lastly; v) the N2pc is often elicited with stimuli that can be quickly identified from a visual search array, such as a shape, color, or orientation (Luck et. al, 1993; Luck & Kappenman, 2014; Woodman & Luck, 2003). Since these stimuli had never been used in an experiment before, we performed a behavioral pre-study in order to validate the stimuli and task design prior to the main EEG study.

Method

Ethics Statement

Ethical approval for this study was granted by the Ethical Committee Psychology (ECP) affiliated with the University of Groningen, the Netherlands. The study was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants included in the study.

Participants

Twenty-six psychology students [(18-28 years, $M = 20$, $SD = 2$), 18 female, 24 right-handed] participated in the pre-study. Participants were recruited through the University of

Groningen undergraduate participant pool and were compensated with course credit. Inclusion criteria were that participants must be between 18-32 years old and have normal or corrected-to normal vision. Participants were required to speak English well enough to communicate with the researcher and read the instructions for the task. Because this was a pre-study to determine the validity of the stimuli and task design, no participants were excluded.

Stimuli selection

Airport layouts were drawn from NOMO Design, a company that makes simple artistic posters of various themes such as cars, airports, and racetracks. Permission to use the runway art was granted by the Creative Director of NOMO Design. Seventy-six airport runway designs were drawn from the website, in order to determine a smaller subset to be utilized for the main study. All airport images taken from the website can be found in Appendix A. See Figure 1 for an example of one of the 76 images pre-editing. Runway images were edited to remove the text, have a whiter color, and contain a transparent background using Imagemagick (Imagemagick Development Team, 2021).

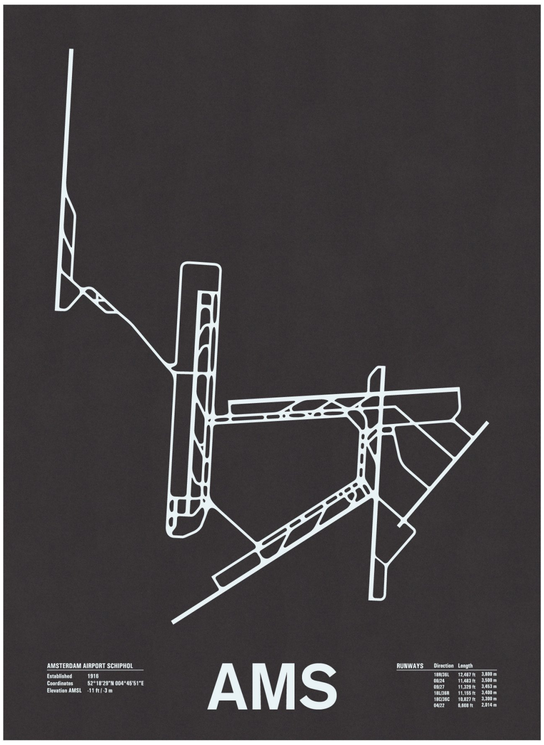


Figure 1. Example of an unedited image from shop.nomodesign.com. AMS stands for Amsterdam.

Runway layouts were categorized by the number of runways they contained. This was done according to official airport metrics. As an additional measure of size, researchers visually estimated the size of the layouts on a scale of 1-3. A table of all airport codes, their corresponding number of runways, and size can be found in Appendix B.

Task design

The computerized learning task was designed and presented using OpenSesame (Mathôt et al., 2012). Before starting the task participants saw a series of instructions that included one example of a trial. The exact instructions used in the task can be found in Appendix C. After reading the instructions but before starting the task, the researcher present asked the participant to make mental notes of the strategies they used during the task as well as any comments they had for improvement. It was indicated that they would be asked about their experience at the end of the experiment.

Three different versions of the task were made based on the results and feedback of the participants. The different versions and their respective changes can be seen in Appendix D.

In general, each trial can be considered a test as to whether the participant has learned the given association between a three-letter airport name and the corresponding airport layout, as well as learning material. At the beginning of each trial, participants were shown a three-letter airport code (i.e., AMS, representing the Amsterdam airport) for 500 ms (see Figure 2). After a fixation interval of 500-600 ms, four airport layouts were presented on the lower half of the screen, one of which was always the correct layout. During this screen, participants had 1300-2000 ms (depending on task version) to respond using the key corresponding to the target location. They then received feedback (correct or incorrect) followed by the cue-layout pair in order to learn the association. The cue-layout pair was presented on each trial regardless of the participant's response, so that each trial could be considered equal learning exposure to the material.

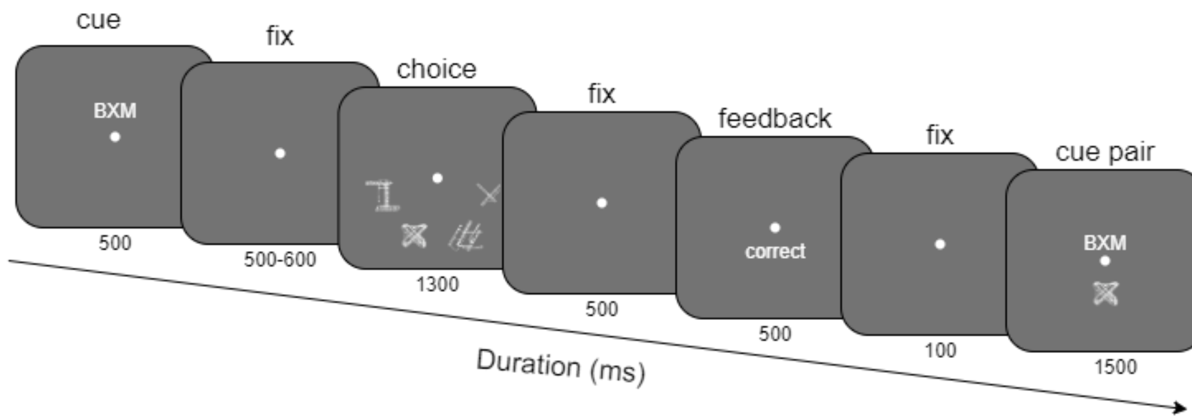


Figure 2. Diagram of a trial. Choice screen duration is reflective of the final version.

Four airports were displayed rather than two in order to minimize the use of strategies other than bringing up the template from long-term memory. With just two options, participants are much more likely to use strategies such as elimination: knowing the identity of the distractor. Or, if the distractor is a new item, by recognizing that the target has been seen before but not what it is. By having four items that have all been seen before, with the exception of the first few trials, the likelihood of using these strategies can be minimized.

Blocking design

As we aimed to measure learning by how many times a particular airport had been seen, we created the variable 'repetition'. Each time an airport has been displayed as the target for one trial, this counted as one repetition of that airport. In this blocking design, each airport has 12 repetitions, but not all repetitions are equally spaced from each other. The blocking design was created to allow for spaced out learning: some repetitions of the items were very close, for easy learning, while others were farther, to add difficulty. To do this, the entire task consists of three blocks of equal length (see Figure 3). Each block consists of eight sets of five airports. To begin a block, one set of five airports is seen in a randomized order. This set is repeated three more times for a total of 20 trials, shuffling the order the five are presented for each repetition of the set. The distractors for these trials are all pulled from the same set of five, to decrease use of general familiarity to find the target. Within a block, this process repeats for all of the sets, until all 40 items have been seen four times (160 trials). At this point the next block starts with the same 40 items, same sets, and same set order, while the order of item presentation within a set is still randomized. After each set of airports (20 trials), an optional break was provided with the duration regulated by the participant. This blocking design results in 4-8 trials in between most repetitions (for easier learning) but 140-144 trials between repetitions 4-5 and repetitions 8-9 (adding difficulty).

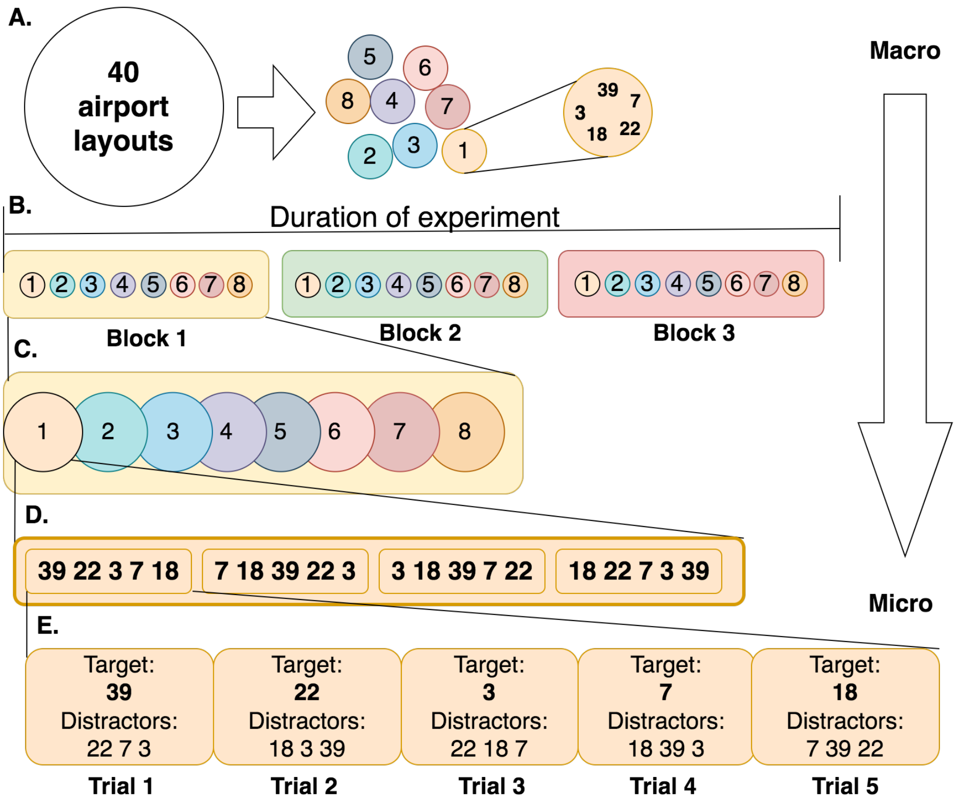


Figure 3. Diagram of blocking design.
A. Forty items are first randomly shuffled into eight sets of five. These sets are maintained for the duration of the experiment. **B.** Each block maintains the same order of the sets. **C.** Each block contains eight sets. **D.** Each set is seen four times in a row. The items within a set are shuffled for each repetition. **E.** Each item in the set is seen once as the target. Semi-random design allowed for sufficient randomization to minimize order effects, while still forcing items to be close enough in time such that learning can occur. This design leads to a rapid increase in learning within blocks, and a forgetting effect between blocks.

Pre-study goals

Four main goals were laid out for the pre-study. Goal 1: Confirm participants are learning the material with a steady increase in accuracy from the first to the last repetition of the items. Goal 2: Optimize the possibility that the participant must bring up the template from working memory in order to succeed at the task. This entailed determining if participants are using the strategy of elimination and if necessary, decreasing time to respond in order to reduce use of this strategy. Goal 3: Instead of 40 airports randomly picked out of 75, decrease the pool to 40 so that all participants learn the same set of airports. Goal 4: Confirm the task is doable while fixating gaze on the center of the screen so that the results can be reasonably extrapolated to an EEG task setting. The three versions of this task were created sequentially, as expectations for behavioral results were met and the task was improved accordingly.

Results

Goal 1: With the behavioral data we confirmed that learning was occurring. As can be seen in Figure 4, average accuracy increases from the first to the last repetition of the items. The sudden decreases in accuracy at repetitions 5 and 9 were expected, as these were the first repetitions at the start of a new block (at which point items may have been forgotten). Graphs displaying the average accuracy, response time, and number of missed responses for all three versions can be found in Appendix E.

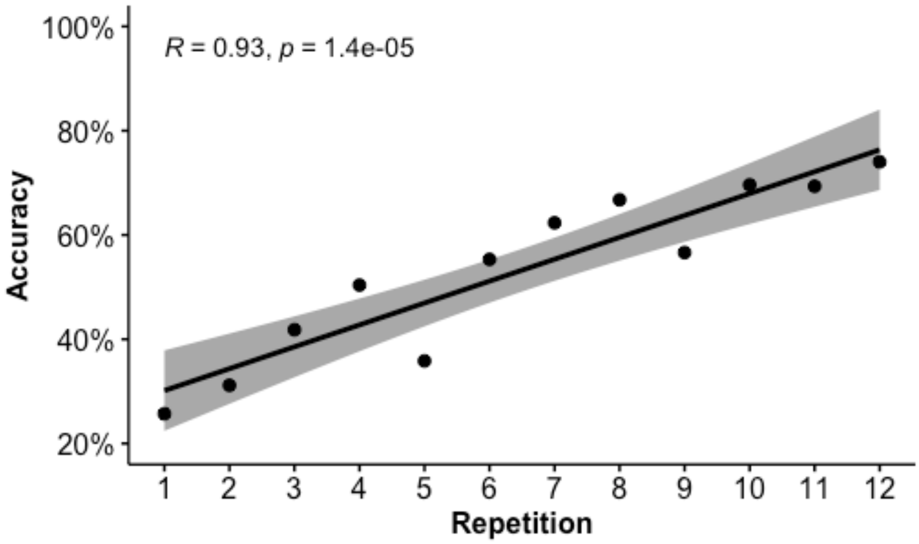


Figure 4. Accuracy per repetition with regression line fitted and including a significant ($p < .001$) positive correlation. These data only reflect version 3 of the experiment ($N = 11$). For data from all three versions see Appendix E.

Goal 2: To determine whether participants were using the strategy of elimination, increasing knowledge of the distractors was assessed for an effect on accuracy. The effect of increasing knowledge of distractors was examined by conducting a repeated-measures ANOVA on the effect of number of recognizable distractors on accuracy. Results showed that decreasing the time to respond from 2000 ms to 1300 ms successfully reduced the use of the strategy of elimination. The effect of increasing knowledge of distractors was significant, F (4, 10) = 1.235, p = .025, when the time given to respond was 2000 ms. However, in the final version of the task, with 1300 ms to respond, there is no effect of increasing knowledge of the distractors, F (4, 40) = 1.845, p = .139. These results suggest that the strategy of elimination could not be utilized in the shorter (1300 ms) time frame. All three versions of the task and their quantitative and qualitative results in regards to the elimination strategy can be found in Appendix F.

Goal 3: The pool of stimuli was reduced from 75 to 51 by eliminating airports with less than 2 runways and more than 4, as well as those which looked very similar. An RM-ANOVA was performed to compare accuracy between the 3 remaining number of runways, finding an effect of number of runways on accuracy, F (2, 50) = 5.798, p = .005. Post hoc tests found the airports with four runways showed greater accuracy compared to the airports with three runways, t (25) = 3.23, p = .01. This resulted in elimination of 11 more airports with four runways, for a final total of 40 stimuli to learn. A detailed explanation of how the pool was reduced can be found in Appendix G.

Goal 4: With no way to enforce fixation, qualitative data was assessed to give an indication that the task was doable while fixating one’s gaze on the center of the screen. When explicitly asked, some participants reported difficulty with fixating at the beginning of the experiment (12/22), but found it possible (14/22). Two participants even noted that fixating was useful because there was not enough time to look at each option.

Main Experiment

Method

Ethics Statement

Ethical approval for this study was granted by the Ethical Committee Psychology (ECP) affiliated with the University of Groningen, the Netherlands. The study was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants included in the study.

Participants

Twelve psychology students [(18-22 years, *M* = 19.1, *SD* = 0.9), all female, 8 right-handed] participated in the main experiment. Participants were recruited through the University of Groningen undergraduate participant pool and were compensated with course credit. Inclusion criteria were that participants must be between 18-32 years old, have normal or corrected-to normal vision, and had not participated in the pre-study. Participants were required to speak English well enough to communicate with the researcher and read the instructions for the task. Data from one participant with low accuracy (2 *SD* below the mean in the second half of the experiment) was excluded, leaving 11 participants for the final analysis.

Task and stimuli

The computerized learning task was the same as the final version of the pre-study task (Figure 2) with the exception of the stimuli pool. The pool of airports used for the EEG study was narrowed down to 40 so that all participants saw the same airports. The final 40 chosen stimuli can be found in Appendix B. The task was presented on a 22-inch CRT monitor with a 60 Hz refresh rate. Participants were comfortably seated approximately 60 cm from the screen. Prior to recording data, participants were given one set of 20 practice trials with airport layouts that were not used in the main task. They were told to blink freely but to perform no eye movements during the task. Eye-blinks were permitted throughout the task because blinks can be reliably identified and removed in pre-processing (Dowding et al., 2015). If they felt comfortable with the task and showed no saccadic eye movements by the end of the practice trials, they started the main experiment. Otherwise, they repeated the practice.

EEG recording and pre-processing

EEG was recorded using a 64-channel, Ag/AgCl electrode, Duke-layout, equidistant, extended-coverage Waveguard cap (ANT Neuro, Netherlands). The ground electrode was placed on the left clavicle, and all electrodes were online referenced to 5Z. Continuous EEG signals were amplified using the battery-powered eego mylab amplifier (ANT Neuro, Netherlands), and recorded at a sampling rate of 500 Hz while impedances were kept below 5 kΩ.

For each participant, the following pre-processing procedure was followed: first, data was offline re-referenced to the mastoids (M1/M2). Raw data were cleaned by hand to remove muscle artifacts. Eye-blinks artifacts were identified and removed from the data using independent component analysis (ICA) implemented by EEGlab's extended infomax algorithm [EEGlab software package version 2021.1 (Delorme & Makeig, 2004)]. To achieve a higher-quality ICA decomposition, data were high-pass filtered using a 1 Hz causal finite impulse response (FIR) filter prior to running ICA (Dowding et al., 2015). The resulting ICA weights were transferred to the raw data of the same participant, and eye-blinks were then removed. The data was subsequently band-pass filtered with a Hamming windowed FIR filter consisting of a 0.05 Hz high-pass cut-off and a 30 Hz low-pass cut-off. This maintains the frequency range in which the N2pc is typically found (Marturano et al., 2020), and filters out most drift effects and muscle artifacts (Luck et al., 2005) providing a good signal-to-noise ratio. EEG data was epoched from 200 ms pre-choice screen onset until 1200 ms post-onset. Baseline correction was performed -200 ms prior to the onset of the choice screen. Epochs containing remaining artifacts such as horizontal eye-movements were identified and excluded from the final averages using a 40 µV step function (from -200 to 1200 ms) provided within ERPlab (Lopez-Calderon & Luck, 2014).

Data binning and averaging

For each participant, trials were binned together as a function of how many times a participant had seen an item (i.e., all trials where the participant saw an item for the first time were binned together, items seen for the second time binned together, and so on). Target-evoked activity was averaged separately for each target location (upper left, lower left, lower right, and upper right). For each participant, this resulted in 48 bins with an average of 100 trials (*SD* = 22.6) per bin after rejection of epochs containing noise or horizontal eye movements. After merging across top and bottom and left and right in order to create contra-ipsi difference waves, this resulted in an average of 399 epochs (*SD* = 17.8) going into each grand average repetition in the final analysis.

Statistical analysis

The attentional bias (as reflected by the N2pc) calculated by measuring total activity at the posterior occipital electrode channels (PO7/PO8) typically associated with the maximal N2pc activity (Marturano et al., 2020) between 200-300 ms after onset of the choice screen. The classic N2pc difference wave (Luck & Kappenman, 2012) was calculated by subtracting activity ipsilateral to the target from the activity contralateral to the target on the same level of the screen, then collapsing over the right and left sides. That is, activity from upper right targets was paired with activity from upper left targets for analysis. This was done in order to account for changes in N2pc that may occur from vertical differences in target location.

Amplitude values were calculated by taking the mean amplitude of the contralateral-ipsilateral difference wave between 200-300 ms after choice screen onset. Onset latency values were calculated using the jackknife approach (Ulrich & Miller, 2001) and the fractional area technique (Hansen & Hillyard, 1980; Luck, 2014) with a parameter of 50%, as recommended for measuring N2pc onset by Keisel et al. (2008). This is done by measuring the first time point between 200-300 ms at which a negative contra-ipsi difference wave reached 50% of its total area for each participant per repetition.

In order to test how the behavioral measures and neural attentional bias changed as a function of learning, we implemented a mixed modeling approach utilizing JASP (JASP Team, 2022). Mixed models were chosen because of the continuous nature of the independent variable 'repetition.' While the differences in repetitions 5-6 and 8-9 due to block reset make the variable not purely continuous, it is still approximately continuous. A repeated-measures ANOVA treats the repetitions as categorical, increasing the likelihood that one repetition would be considered significantly different than another, and inflating the *p*-value. Additionally, the need for post-hoc tests to determine the source of the effect between 12 levels of a categorical variable would lead to a large increase in the experiment-wise error rate. For these reasons, as well as the flexibility of mixed models to include or not include the variation per participant in the model, a mixed modeling approach was considered the best choice for these data.

Using how many times a participant has seen an item (repetition) as an indication of 'time spent learning'; we modeled how accuracy, response time, and N2pc amplitude changes were associated with repetition. To account for between-subjects differences, a random effects grouping factor of participant was included in the model. Linear mixed effects models (LMEM) were used for all dependent variables except accuracy. The Satterthwaite formula was used to approximate the degrees of freedom of these models in order to test statistical significance. Accuracy was treated differently because it is binary (0 or 1 on each trial). For this reason, a generalized linear mixed model (GLMM) with a Binomial Logit Link function was used to analyze accuracy, with Likelihood ratio tests to determine statistical significance.

Results

Behavioral Results

Behavioral results (Table 2 and Figure 5A) show that participants were able to learn the name-layout associations over the course of the experiment. By the final repetition, average accuracy was 83% (*SD* = 9%). The GLMM estimates that with each increase in repetition, we expect to see a 19% increase in the odds of being correct on the next repetition ($p < .001$).

As can be seen from Figure 5B, participants were able to respond faster with repeated testing and exposure. LMEM results show that response time decreased by an average of 9.750 ms with each repetition between repetitions 2-12 ($p < .001$) (See Table 2 note for justification on removal of repetition 1 from the model). By the last repetition, participants were responding on average in just 762 ms (*SD* = 79 ms), despite having 1300 ms to respond.

Attentional Bias Results

As can be seen from Figures 6A and 7, participants showed no attentional bias during the first three exposures to the objects and their associated names. Then, as the association between airport name (cue) and airport layout is strengthened through repetition, the attentional bias towards the side of the screen with the correct runway layout also increases. More specifically, the LMEM predicts that N2pc amplitude relative to the target increases by -0.111 uV ($p = .007$) with each repetition of the cue-layout pair (see Table 2). Despite a visual trend in the average onset latency (Figure 6B), repetition was not significantly associated with a change in N2pc onset latency in the linear mixed model ($p = .339$).

Table 2

Model Equations and Fixed Effects Estimates

	Fixed effect estimate					
Model	Term	Estimate	*SE*	*df*	*t*	*p*
Acc ~ rep + (1|subject)*	Intercept	-.783	.129		-6.052	<.001
	Rep	.191	.009		21.080	<.001
RT ~ rep + (1|subject)**	Intercept	879.536	19.997	13.424	44.028	<.001
	Rep	-9.750	1.056	4828	-9.232	<.001
Amp ~ rep + (1|subject)	Intercept	0.506	.322	72.019	1.574	.120
	Rep	-0.111	0.041	120	-2.721	.007
Latency ~ rep + (1|subject)	Intercept	275.714	5.422	106	5.085	<.001
	Rep	-2.842	0.680	106	.4177	.339

Note: *GLMM using Binomial Logit Link function. **Repetition 1 was removed from the LMEM of response time, because response times on repetitions 2-12 are approximately representing the time taken to try to respond correctly, while response time on the first exposure to an item is reflective of the time taken to make a guess at random (participants are told to guess the first time they see an item, rather than fail to respond).

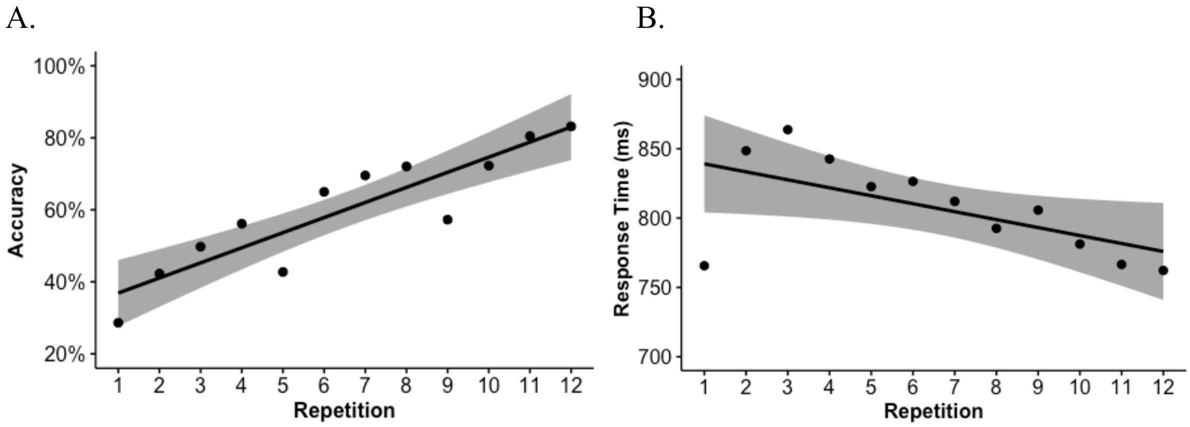


Figure 5. Behavioral metrics of learning A) Accuracy increases as a function of repetition. Participants start by guessing at 25% chance level. Accuracy dips at repetitions 5 and 9 because of block reset. B) Response time decreases as a function of repetition. Response time starts out low because guessing requires little deliberation. Gray shaded areas represent 95% confidence intervals.

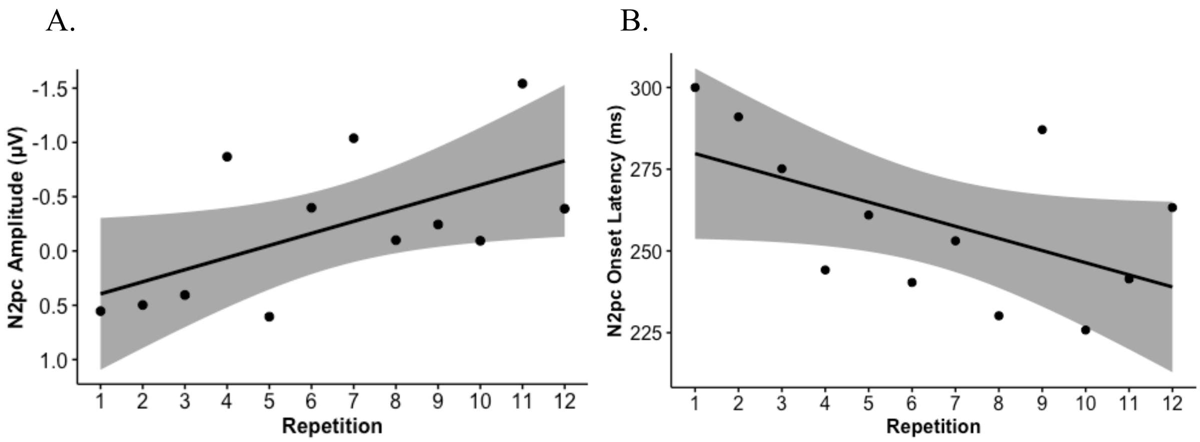
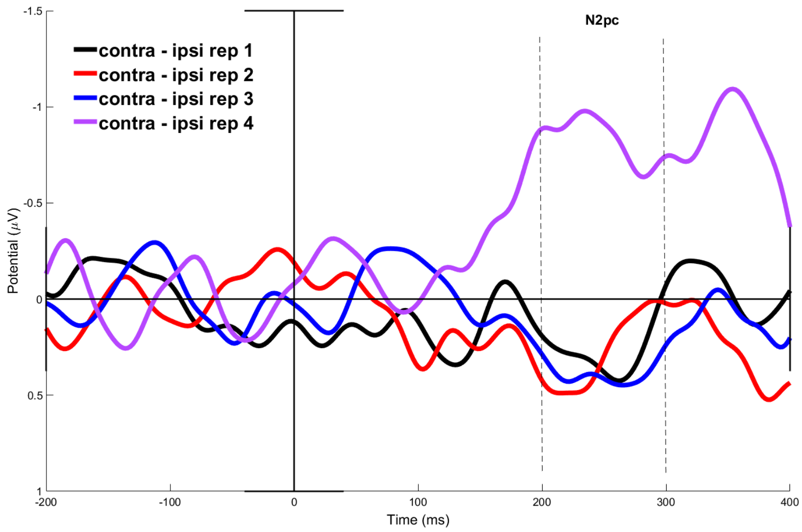
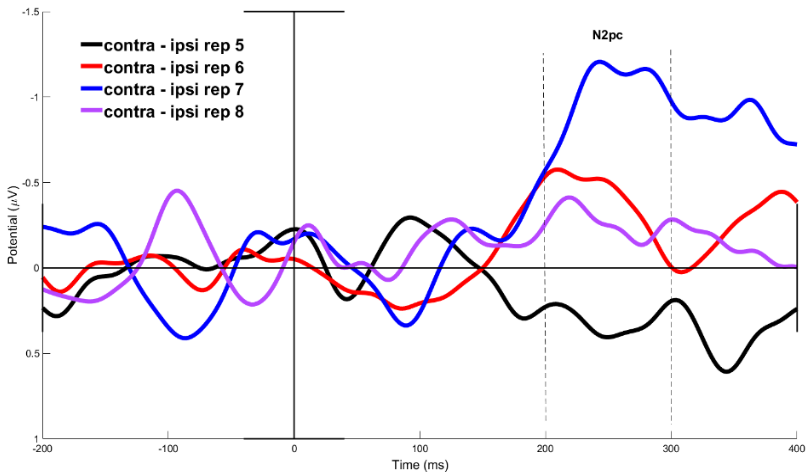


Figure 6. Neural metrics of attention. A) Learning was reflected by an increase in the negative difference wave elicited contralateral to the correct airport layout, which is associated with attention to the correct side. This negative deflection increased in size as a function of learning over 12 repetitions of each item. B) Latency appears to decrease as a function of repetition, but this trend is not significant. Gray shaded areas represent 95% confidence intervals.

A.



B.



C.

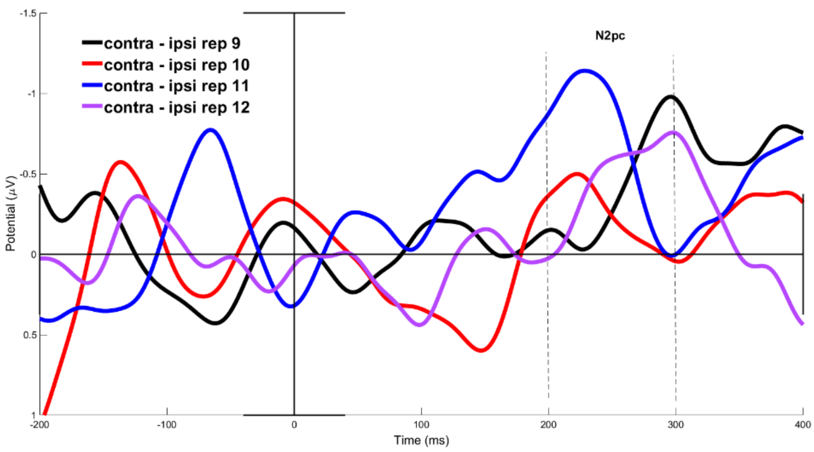


Figure 7. Grand average contra-ipsi waveforms of repetitions 1-12, split into three plots for clarity. ERPs filtered for plotting with a 10 Hz low-pass and a .2 Hz high-pass IIR butterworth second order filter. A) Repetitions 1-4. Plot shows no negative deflection is visible from repetitions 1-3, while the fourth presentation of the associated stimulus (shown in purple) shows a 1uV negative deflection between 200-300 ms contralateral to the airport layout associated with the cued name. B) Repetitions 5-8. No negative deflection can be seen at repetition 5, while repetitions 6-8 show an N2pc. Repetition 7 shows the greatest size. C) Repetitions 9-12. All show an N2pc with the greatest negative amplitude at repetition 11.

Discussion

This study investigated whether learning could be indexed with a neural measure of attentional bias. We hypothesized that recalling learned information into working memory would cause an attentional bias towards the target in a search array. We tracked attentional bias using the N2pc, an event-related potential associated with the orienting and time course of attention to a lateralized target (Luck et al., 2012). By measuring N2pc over the course of a 12-repetition learning task with a set of 40 novel stimuli, we aimed to track whether an increase in attentional bias occurred as a result of learning.

We found that participants used the 12 repetitions to successfully learn the material (Figure 5A), as in the pre-study (Figure 4). The learning led to an increase in N2pc amplitude over the course of the 12 repetitions (Figures 6A & 7). Response time also decreased as a function of repetition (Figure 5B), but the onset of the attentional bias did not show a similarly significant decrease (Figure 6B). These results support the hypothesis that before an item has been learned, there is no attentional bias towards the target side, and participants perform at guessing rate. Then, as the item is more likely to be recalled correctly, the attentional bias occurs 200-300 ms after onset of the choice screen towards the correct choice.

While many studies have examined the N2pc in both encoding and responding to learned information, this is, to the best of our knowledge, the first study that attempts to use the N2pc to measure how well a set of facts have been encoded. Most studies investigating the role of the N2pc in learning have focused on its role in enhancing sensory processing (Luck et al., 1993) and encoding information (Gieb et al., 2021). The studies that have used N2pc during the testing phase of learning have primarily focused on the learning of rewards (Oemisch et al., 2017; Wei & Ji, 2021; Pütz et al., 2022), or the learning of a shape or orientation of the correct target in a search array (An et al., 2012; Qu et al., 2017). This means that the learning task usually involves the consistent (trial-to-trial) repetition of the same target to be found in a search array, rather than both the location and the target shape being different on each trial as in the current study. Trial-to-trial repetition of the same target does not require recall, but maintenance of the item in working memory, and it has even been shown that after a while the maintenance is no longer needed to succeed (Geoffrey et al., 2007). In order to succeed in the current study, participants needed both knowledge of the right target (successfully bring the search template from long term memory to working memory) and also a timed visual search for said target. Due to these differences, it was initially unclear whether the learning process for learning word-shape associations would produce similar results as studies looking at the N2pc and reward learning.

The results of this study that the N2pc amplitude increases with repeated chances to learn the material are in line with previous studies that have looked at the N2pc as a function of reward learning (Oemisch et al., 2017; Wei & Ji, 2021; Pütz et al., 2022). Here we show that the N2pc is similarly affected by learning a set of novel associations as when learning to associate a stimulus with reward.

The results for the onset of the attentional bias are more difficult to place in the context of previous literature. We examined the onset of the N2pc because reaction time is often correlated with the onset of the N2pc (van den Berg et al., 2016), and the response times in this task decreased with repetition (Figure 5B). If response time can provide information about how well something has been learned above and beyond accuracy, we hypothesized that the onset of the attentional bias may be able to do the same for the measure of amplitude. However, none of the studies mentioned so far have measured onset latency of the N2pc, despite measuring amplitude. This might be because it is considered difficult to measure onset latency of ERPs due to its sensitivity to trial-to trial variability and noise (Keisel et al., 2008). Here, the result that latency is not predicted by repetition may be because of low sample size and/or noisy trial-to-trial data. In addition, our task allowed 1300 ms to respond, but N2pc onset latency is measured between 200-300 ms. Due to variable search time it is possible that the attentional bias occurred after the 200-300 ms window in which we measured latency. Whereas is it possible that we did not find a significant effect of repetition on N2pc onset latency because latency does not reflect learning in the way amplitude does, the above mentioned alternative explanations cannot be ruled out.

While the result for amplitude is in line with our hypothesis that attentional bias can track the learning of facts, it is worth discussing whether the effect is driven solely by an increasing strength of the representation in working memory, or whether additional mechanisms of attentional capture may be at play.

The idea that attention might have different roles in learning tasks is a topic of much debated research (Le Pelley, 2004). The Pearce and Hall (1980) model of learning posits the idea that attention could reflect effortful learning. This model predicts that attention is driven by surprising outcomes: when the feedback is negative and uncertainty is introduced on the next trial, attention will increase. This theory contradicts the model posited by Mackintosh (1975), who argues that attention is biased towards stimuli that have the highest probability of reward. Efforts to reconcile these two theories have suggested that both aspects of attention are at play during learning tasks: one aspect of attention is concerned with action: Choose the most rewarding stimulus, and the other is concerned with learning: attend to the uncertain stimuli when adjusting for previous errors (Holland & Gallagher, 1999; Hogarth et al., 2010). While this study does not use reward, it has been demonstrated that being correct is rewarding in itself (Satterthwaite et al., 2012). It is not possible in this study to disentangle to what extent attention is being driven by the reward of being correct, the level of uncertainty, or the strength of their target representation in working memory. However, this is not necessary to place these results in the context of these two theories.

If both aspects of attention are at play in this task, this could explain why attentional bias reflected by the N2pc amplitude seems to drop unexpectedly at the end of the second and third blocks. This can be seen at repetitions 8 and 12 (Figure 6A). Compared to the graph of accuracy (Figure 5A), which shows the highest accuracy at the repetition at the end of each block, the graph of N2pc amplitude shows this only in block 1 (Figure 7A). Blocks 2 and 3 show the highest amplitude in the repetition preceding the last (Figure 7B & C). With repetitions 8 and 12 being the last of a set of four, they should be well-learned and show a strong N2pc, but this does not seem to be the case. While the idea that N2pc is being driven by how well a participant has learned an item does not explain these drops, the idea that attention is being driven by reward (Mackintosh, 1975) and also by uncertainty (Pearce & Hall, 1980) may provide an answer. Airports at the end of a block may be well-learned compared to previous repetitions but additionally have low uncertainty. The low uncertainty may make the possibility of being correct less exciting, and thus less rewarding. In other words, participants may be rewarded by being correct when they are uncertain about their choice, but not when they are certain. This would be in line with the body of literature supporting the prediction-error framework (Fiorillo et al., 2003; Dreher et al., 2005; Satterthwaite et al., 2007). Higher difficulty induces more uncertainty that one will get correct feedback, causing correct feedback to induce a higher reinforcement response (Satterthwaite et al., 2012). Simply put, the last repetitions of the second and third blocks may be too easy to induce large attentional bias, whereas the third repetition of these blocks finds a sweet spot between being sufficiently learned and also sufficiently difficult. Further studies could try to disentangle these mechanisms by controlling for uncertainty. This could be done by making all trials equally uncertain with a dynamic task that updates based on its prediction of how well an item has been learned (van Rijn et al., 2009), or by making certainty very high with a high number of repetitions, allowing accuracy to plateau. Any effect of repetition on attentional bias in these task designs could indicate an effect of learning alone rather than the effect of varying levels of uncertainty.

The conflict between N2pc being modulated by both learning and uncertainty may relate to why latency does not show a linear trend with learning. Latency may be more reflective of the effortful attention used to decrease uncertainty as described by the Pearce and Hall (1980) model. In this task, the onset of the attentional bias may have decreased during phases of active learning, such as the first few repetitions, or the first half of the experiment. Then, as uncertainty lowers and accuracy becomes easier to achieve with little effort, the attentional bias becomes slower to occur once again. This pattern would not be picked up by the linear mixed effects model because it is no longer a linear relationship, and would have to be accounted for with a more complicated model. Further studies could investigate this by testing whether adding a polynomial term to account for a non-linear relationship increases the fit of the model.

To conclude, we found that the N2pc amplitude can be used to track increases in learning of a novel set of stimuli. We suggest further research is needed to disentangle the multiple aspects of learning that attention could be reflecting, and to investigate whether the onset of this bias may be indicative of learning as well. Future studies could benefit from a greater sample size, as well as checking a longer window of latency measurement to accommodate the variable search time. Implications of these results could be wide ranging in the field of learning research. Results support the hypothesis that attentional bias could be a metric for how well facts have been learned. Whether this metric is beneficial above and beyond the metric of accuracy has yet to be explicitly determined. Future research could expand on this finding by including a measure of attentional bias in learning models. Models that predict learning are an increasing area of research because of their potential to improve rates of learning and actively adjust for individual differences in learning (van Rijn et al., 2009). While EEG measures cannot be implemented on a wide scale to improve learning, it would still benefit the field of learning research to know how accurate these models can become with increasing predictors, including neural measures such as the N2pc.

References

An, A., Sun, M., Wang, Y., Wang, F., Ding, Y., Song, Y., & Herzog, M. H. (2012). The n2pc is increased by perceptual learning but is unnecessary for the transfer of learning. *Plos One*, 7(4). https://doi.org/10.1371/journal.pone.0034826

Beninger, R. J. (2018). *Life's rewards: linking dopamine, incentive learning, schizophrenia, and the mind.* Oxford University Press.

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research*, 40, 1203–1215.

Dahaene, S. (2020). *How We Learn: Why Brains Learn Better Than Any Machine… for Now.* Penguin Random House.

Delorme A. & Makeig S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 134(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dowding, I., Debener, S., Müller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Annual International Conference IEEE English Medicine Biology Society*.

Dreher, J.-C., Kohn, P., & Berman, K. F. (2006). Neural coding of distinct statistical properties of reward information in humans. *Cerebral Cortex*, 16(4), 561–73.

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458.

Failing, M., & Theeuwes, J. (2018). Selection history: how reward modulates selectivity of visual attention. *Psychonomic Bulletin & Review*, 25(2), 514–538. https://doi.org/10.3758/s13423-017-1380-y

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614), 1898–1902.

Geib, B. R., Cabeza, R., & Woldorff, M. G. (2021). Linking the rapid cascade of visuo-attentional processes to successful memory encoding. *Cerebral Cortex*, 31(4), 1861–1872. https://doi.org/10.1093/cercor/bhaa295

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, 49(3-4), 277–90.

Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Science*. 3, 65–73. https://doi.org/10.1016/S1364-6613(98)01271-6

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates low-level saccade target selection: evidence from rapidly generated saccades in the global effect paradigm. *Journal of Vision*, 13(13), 4–4. https://doi.org/10.1167/13.13.4

Hogarth, L., Dickinson, A., & Duka, T. (2010). Selective attention to conditioned stimuli in human discrimination learning: untangling the effects of outcome prediction, valence, arousal, and uncertainty, in *Attention and Associative Learning. From Brain to Behavior*, eds C. J. Mitchell and M. E. Le Pelley (Oxford: Oxford University Press), 71–97.

Huang, L., & Pashler, H. (2007). Working memory and the guidance of visual attention: consonance-driven orienting. *Psychonomic Bulletin & Review*, 14(1), 148–153. https://doi.org/10.3758/BF03194042

ImageMagick Development Team. (2021). ImageMagick. Retrieved from https://imagemagick.org

JASP Team (2022). JASP (Version 0.16.3) [Computer software].

James, W. (1890). *The Principles of Psychology*. Henry Hold and Company.

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of erp latency differences: a comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, 45(2), 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.x

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, 57(3), 193–243.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, 8, 213.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Luck, S. J., & Kappenman, E. S. (2012). *Oxford handbook of event-related potential components*. Oxford University Press.

Luck, S. J., Fan, S., & Hillyard, S. A. (1993). Attention-related modulation of sensory-evoked brain activity in a visual search task. *Journal of Cognitive Neuroscience*, 5, 188–195.

Lutz, K., Pedroni, A., Nadig, K., Luechinger, R., & Lutz, J. (2012). The rewarding value of good motor performance in the context of monetary incentives. *Neuropsychologia*, 50(8), 1739–1747. https://doi.org/10.1016/j.neuropsychologia.2012.03.030

Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*. 82, 276–298. https://doi.org/10.1037/h0076778

Marturano, F., Brigadoi, S., Doro, M., Dell'Acqua, R., & Sparacino, G. (2020). A Time-Frequency Analysis for the Online Detection of the N2pc Event-Related Potential (ERP) Component in Individual EEG Datasets. *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),* pp. 1019-1022, https://doi.org/10.1109/EMBC44109.2020.9175462

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

Mazza, V., Dallabona, M., Chelazzi, L., & Turatto, M. (2011) Cooperative and opposing effects of strategic and involuntary attention. *Journal of Cognitive Neuroscience*, 23, 2838–2851. https://doi.org/10.1162/jocn.2011.21634

NOMO Design. (2021). *Runway Series*. https://shop.nomodesign.com/

Oemisch, M., Watson, M. R., Womelsdorf, T., & Schubö, A. (2017). Changes of attention during value-based reversal learning are tracked by n2pc and feedback-related negativity. *Frontiers in Human Neuroscience*, 11, 540–540. https://doi.org/10.3389/fnhum.2017.00540

Pearce, J., & Hall, G. (1980). A model for Pavlovian learning: variation in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*. 87, 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Pearce, J., and Mackintosh, N. J. (2010). Two theories of attention: a review and possible integration, in *Attention and Associative Learning. From Brain to Behavior*, eds C. J. Mitchell & M. E. Le Pelley (Oxford: Oxford University Press), 11–39.

Pütz, C., van den Berg, B., & Lorist, M. M. (2022). Dynamic modulation of neural feedback processing and attention during spatial probabilistic learning. *Iscience*, 25(5). https://doi.org/10.1016/j.isci.2022.104302

Qu, Z., Hillyard, S. A., & Ding, Y. (2017). Perceptual learning induces persistent attentional capture by nonsalient shapes. *Cerebral Cortex*, 27(2), 1512–1523. https://doi.org/10.1093/cercor/bhv342

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: evidence from pupillometry and event-related fmri. *Neuroimage*, 37(3), 1017–1031. https://doi.org/10.1016/j.neuroimage.2007.04.066

Satterthwaite, T. D., Ruparel, K., Loughead, J., Elliott, M. A., Gerraty, R. T., Calkins, M. E., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2012). Being right is its own reward: load and performance related ventral striatum activation to correct responses during a working memory task in youth. *Neuroimage*, 61(3), 723–729. https://doi.org/10.1016

Sternberg, R.J., Sternberg, K. (2015). *Cognitive Psychology*. Cengage Learning. 7th Ed.

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring lrp onset effects in factorial designs. *Psychophysiology*, 38(5), 816–827. https://doi.org/10.1111/1469-8986.3850816 van den Berg, B., Appelbaum, L. G., Clark, K., Lorist, M. M., & Woldorff, M. G. (2016). Visual search performance is predicted by both prestimulus and poststimulus electrical brain activity. *Scientific Reports*, 6(1). https://doi.org/10.1038/srep37718

van den Berg, B., Geib, B. R., San Martin, R., & Woldorff, M. G. (2019). A key role for stimulus-specific updating of the sensory cortices in the learning of stimulus-reward associations. *Social Cognitive and Affective Neuroscience*, 14(2).

Van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention? *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1450.

van Rijn, D., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving Learning Gains by Balancing Spacing and Testing Effects. *Proceedings of the 9th International Conference on Cognitive Modeling,* 187, 108-114.

Wang, M., Yang, P., Jin, Z., Zhang, J., & Li, L. (2019). The success of the representation maintenance affects the memory-guided search processing: an erp study. *Experimental Brain Research*, 237(2), 503–509. https://doi.org/10.1007/s00221-018-5430-0

Wei, P., & Ji, L. (2021). Reward expectation modulates n2pc for target selection: electrophysiological evidence. *Psychophysiology*, 58(8). https://doi.org/10.1111/psyp.13837

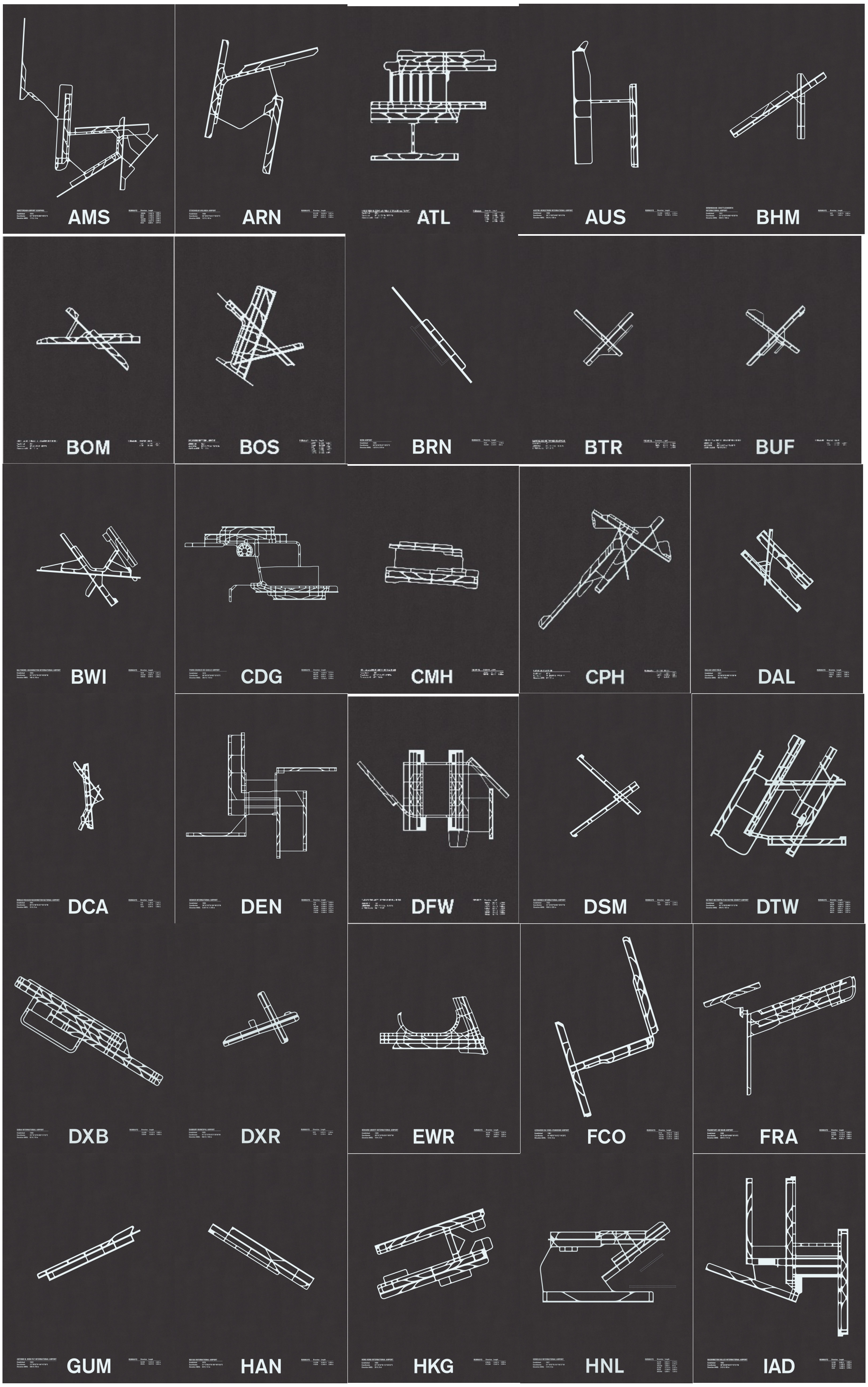
Wolfe, J. M., Cave, K. R., Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–433.

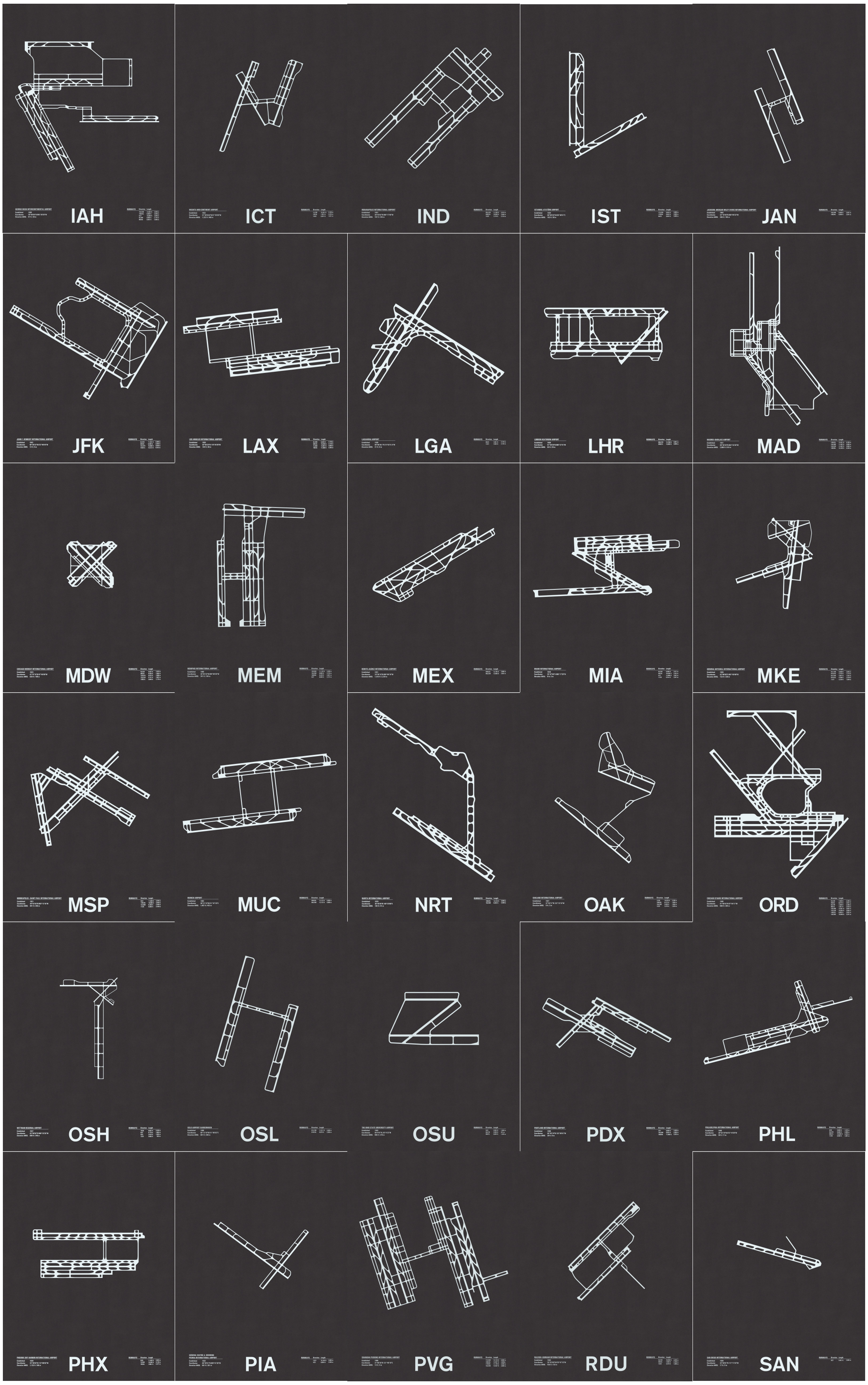
Woodman, G. F. & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology Human Perceptual Performance*. 29(1), 121-38.

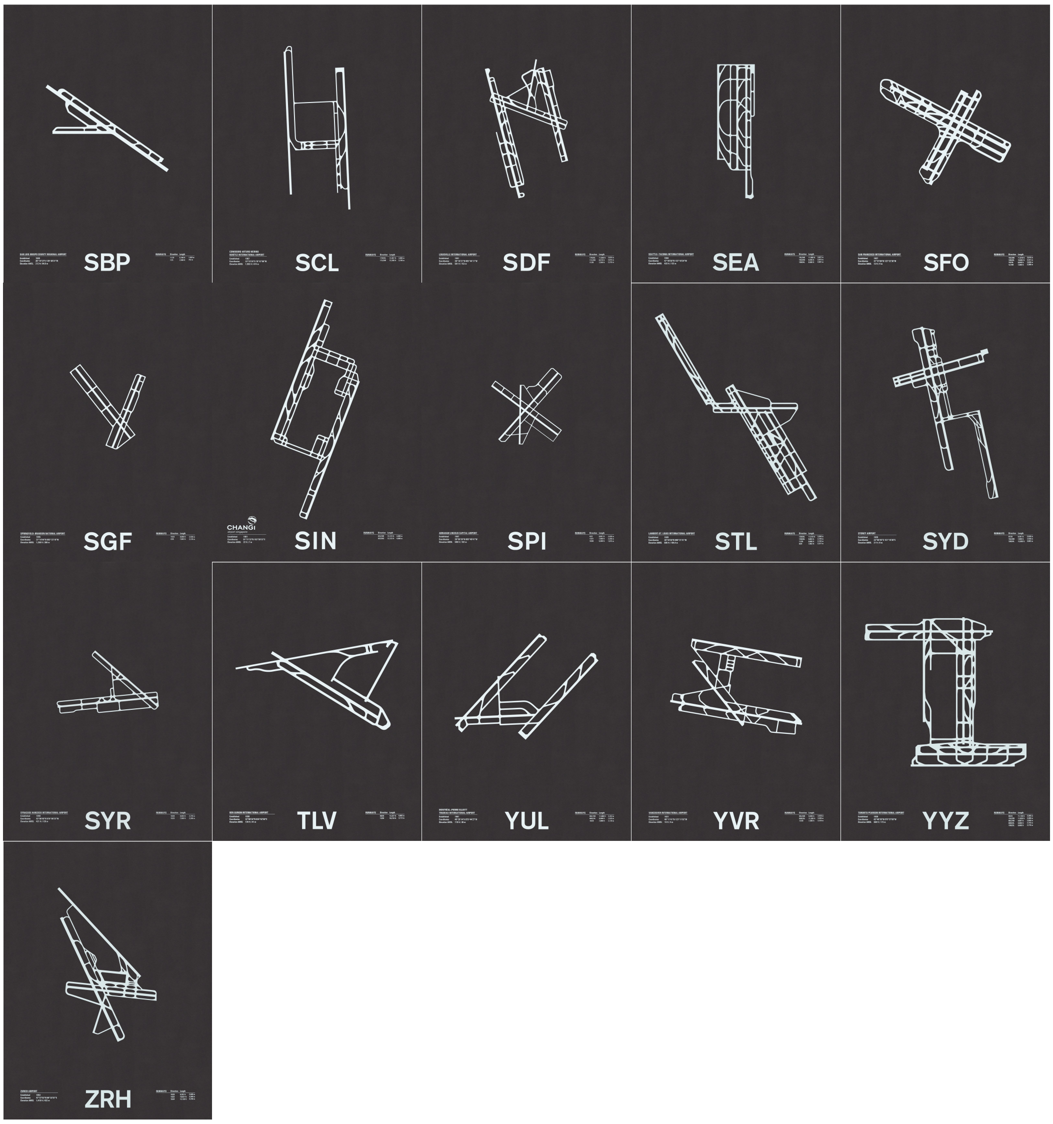
Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral cortex,* 17(1), 124. https://doi.org/10.1093/cercor/bhm065

Appendix A

Airport Runway Layouts




IAH
ICT
IND
IST
JAN
JFK
LAX
LGA
LHR
MAD
MDW
MEM
MEX
MIA
MKE
MSP
MUC
NRT
OAK
ORD
OSH
OSL
OSU
PDX
PHL
PHX
PIA
PVG
RDU
SAN


SBP
SCL
SDF
SEA
SFO
SGF
CHANGI
SIN
SPI
STL
SYD
SYR
TLV
YUL
YVR
YYZ
ZRH

Appendix B

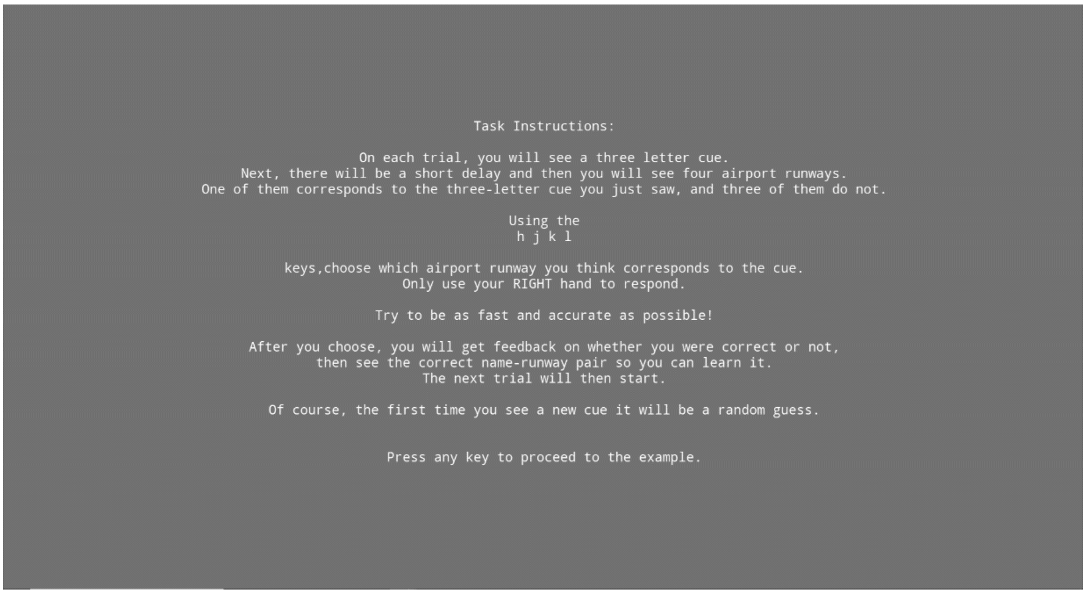
Runway Sizes

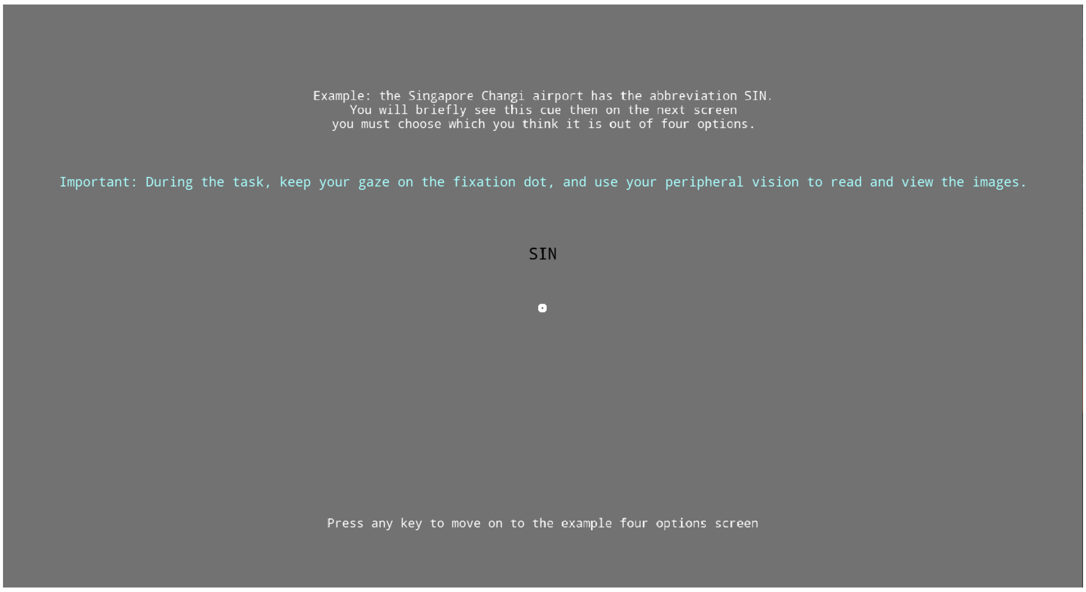
Code	Runways	Size	Code	Runways	Size	Code	Runways	Size
AMS	6	3	GUM	2	1	OSH	4	1
ARN	3	3	HAN	2	1	OSL	2	2
ATL	5	3	HKG	2	2	OSU	3	1
AUS	2	2	HNL	6	2	PDX	3	2
BHM	2	1	IAD	4	3	PHL	4	2
BOM	2	1	IAH	5	3	PHX	3	2
BOS	6	2	ICT	3	2	PIA	2	1
BRN	2	1	IND	3	2	PVG	4	3
BTR	3	1	IST	3	1	RDU	3	2
BUF	2	1	JAN	2	1	SAN	1	1
BWI	3	2	JFK	4	3	SBP	2	1
CDG	4	3	LAX	4	2	SCL	2	2
CMH	2	2	LGA	2	2	SDF	3	2
CPH	3	3	LHR	2	2	SEA	3	2
DAL	3	2	MAD	4	3	SFO	4	2
DCA	3	1	MDW	5	1	SGF	2	1
DEN	6	3	MEM	4	3	SPI	3	1
DFW	7	3	MEX	2	1	STL	4	2
DSM	2	1	MIA	4	2	SYD	3	2
DTW	6	3	MKE	5	1	SYR	2	1
DXB	2	3	MSP	4	2	TLV	2	2
DXR	2	1	MUC	2	2	YUL	3	2
EWR	3	2	NRT	2	3	YVR	3	2
FCO	4	2	OAK	4	2	YYZ	5	3
FRA	4	3	ORD	8	3	ZRH	3	2

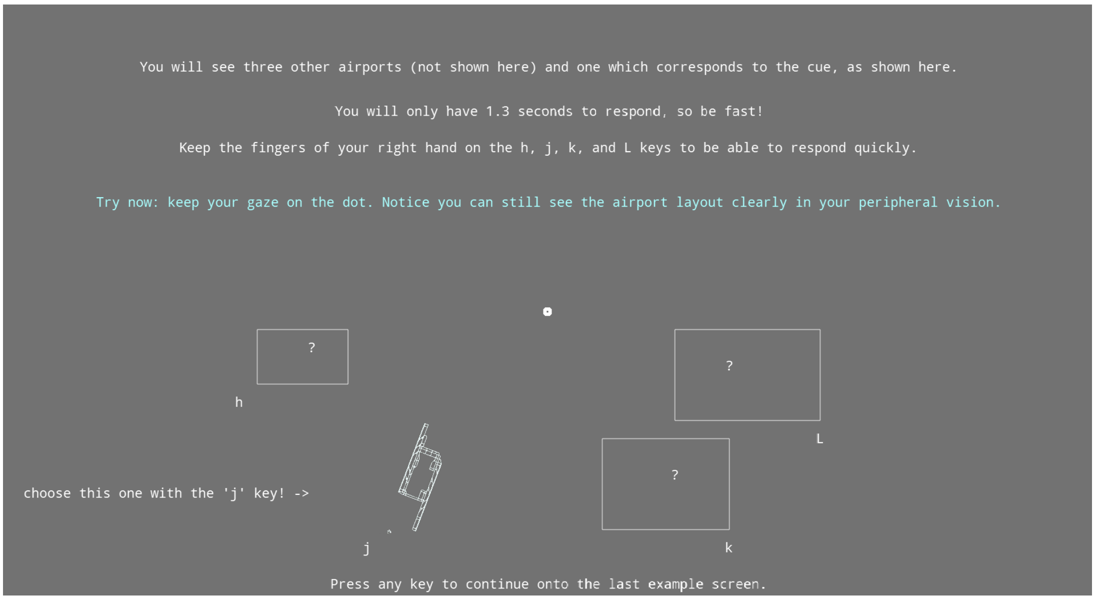
Note: Underlined airports indicate final 40 used for main experiment.

Appendix C

Instructions






You will see three other airports (not shown here) and one which corresponds to the cue, as shown here.
You will only have 1.3 seconds to respond, so be fast!
Keep the fingers of your right hand on the h, j, k, and L keys to be able to respond quickly.
Try now: keep your gaze on the dot. Notice you can still see the airport layout clearly in your peripheral vision.
?
?
h
L
choose this one with the 'j' key! ->
?
j
k
Press any key to continue onto the last example screen.


Lastly, after each trial you will see what the correct answer is so you can commit it to memory.
Keep your gaze on the center dot when you view this!
You will surprise yourself: you will find you are still able to learn the shapes without moving your eyes directly to them.
SIN
Press any key to continue onto the last instruction.


Try your hardest to learn the airport runway layouts and remember them for future trials.
In order to make these easier to learn, you will learn them in sets.
After each set is done, you will have the opportunity to have a break.
Take breaks when you need them!
After the break you will have a new set of airports to learn.
Have fun!
Press any key to begin the task.

Appendix D

Versions of the Task

Table D1.

Three versions of the experiment tested in the pre-study.

Version	*N*	Response timeout (ms)	Response keys	Airports learned	Total pool of airports
1	4	2000	h, n, m, k	30	75
2	11	1500	h, n, m, k	40	75
3	11	1300	h, j, k, l	40	51

Version 1

As a preliminary test of the task, this version contained only six sets of five airports (total of 30), drawn randomly out of the whole pool of 75. To check if participants were learning the material, the accuracy was assessed and found to indicate learning. Mean accuracy reached 92% (*SD* = 6%) in the last repetition.

Version 2

Once it was clear the task was not too difficult, two goals were set for the second version of the task: decrease the use of the elimination strategy and instruct participants to act as if they were participating in an EEG experiment. To decrease time to identify more than one airport in the array, time to respond was decreased to 1500 ms. An instruction was added to keep one's gaze on the fixation cross for the entire duration of the experiment, so that the results could be extrapolated to an EEG setting. The researcher was not keeping track of participant eye movements, so this could not be enforced. Participants were still learning the material [accuracy last repetition *M* = 82% (*SD* = 11%)] despite the decrease in time to respond.

Version 3

To further decrease use of the elimination strategy, time to respond was decreased to 1300 ms. The pool of airports was reduced from 75 to 51 by eliminating airports that looked very similar, as well as airports with less than two runways or more than four. Because three participants reported uncomfortable finger placement, the response keys were changed from h,n,m,k to h,j,k,l. This allowed a more natural hand placement because the h,j,k,l keys are in a row, rather than a U-shape.

Appendix E

Visualization of Behavioral Results per Version

A. B.

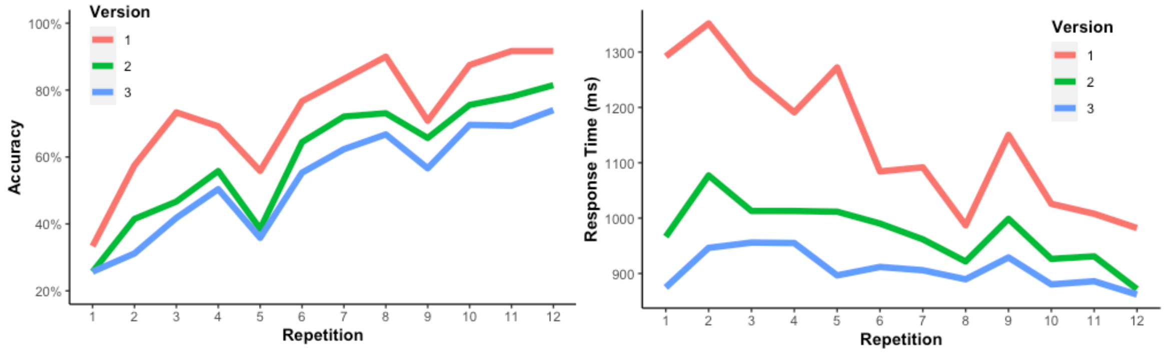


Figure E1. Accuracy and Response time plotted per repetition for each version of the task. A) Accuracy is highest in the first version, with the most likely explanation being the time to respond at 2000 ms. Decreasing the time to respond to 1500 ms (green) and 1300 ms (blue) decreases overall accuracy, but the learning rate is maintained. B) Response time improvement is much more pronounced in versions 1 & 2, indicating that decreasing the time allowed to respond improves overall response time but leaves less room for improvement.

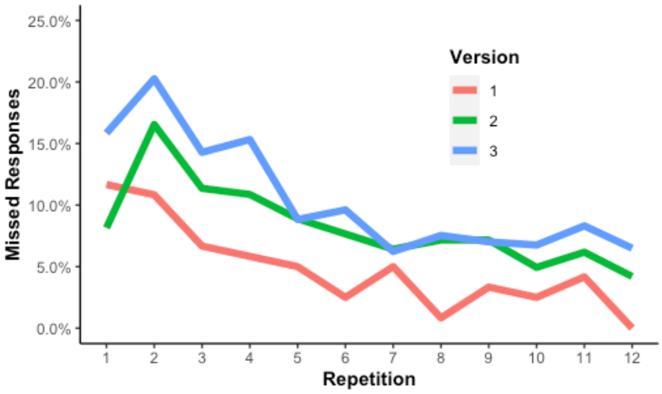


Figure E2. Missed responses per repetition for each version.
More time to respond (red) appears to decrease the likelihood of timeout.

While Goal 1 of the pre-study was met by only looking at accuracy, we visually assessed additional metrics that the participants were improving at the task. These results validated that all versions showed successful learning of the stimuli (Figure E1A), and a decrease in response time from the second to the last repetition (Figure E1B). Lastly, because we decreased time to respond for each version, we assessed whether participants were still able to respond to a similar degree as previous versions (Figure E2).

Appendix F

Extended Elimination Strategy Results

Goal 2 of the pre-study was to decrease use of other strategies during the task. The strategy of elimination is defined as identifying the target not by direct knowledge of the target, but by the knowledge of the distractors. It was described to the participants as, 'I know it's not this one, and I know it's not that one, so it must be one of these'. The strategy of elimination was assessed for two reasons: primarily, if participants are able to identify the target in early repetitions not because they know the item, but because they can identify the distractors, this may lead to an attentional bias in the absence of learning the association. Secondly, this could lead to attention being better allocated to the distractors rather than the target as a method to succeed at the task, which also would not allow us to assess attention to the target as a metric for learning. For these reasons the strategy of elimination was assessed in the pre-study both quantitatively, by looking at accuracy in relation to knowledge of distractors, and qualitatively, by discussing it with the participants.

To assess this quantitatively, we compared the accuracy within the first five trials of all sets. In each of these trials, the target was being seen for the first time, and participants should not have been able to score above chance level (25%). However, the number of recognizable distractors increases from trial 1-5, as the distractors are always pulled from the same set of five airports as the target. If knowledge of the distractors were being used, then we would expect to see an increase in accuracy from trial 1-5, as the number of recognizable distractors goes up.

As it appears from Figure F1, a 2000 ms window to respond (red line) does allow for some use of the elimination strategy. A 60% mean accuracy on the first time seeing an item indicates that recognition of distractors allowed for identification of the unknown target.

Indeed, a RM-ANOVA on the effect of trial number on accuracy for participants in version 1 was significant, F (4, 12) = 3.462, p = .042.

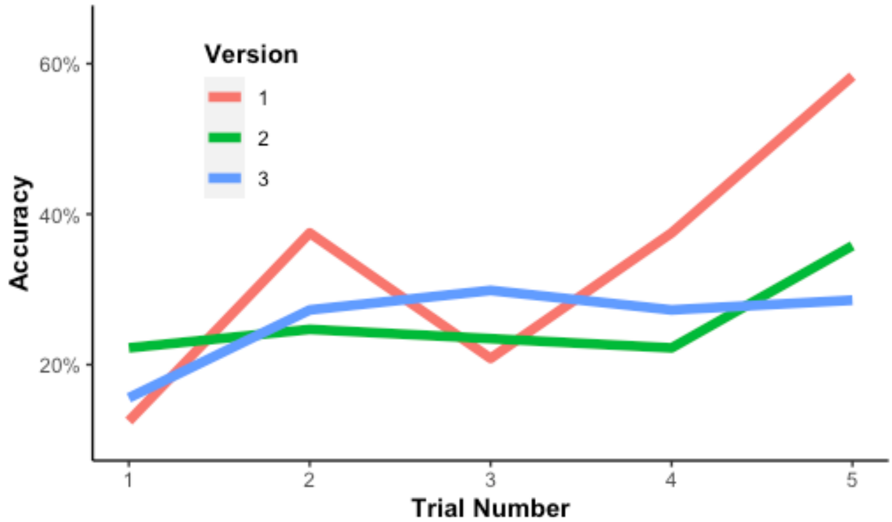


Figure F1. Accuracy per version on the first five trials per set (all are repetition 1). Trial 1 has no recognizable distractors. Trial 2 has 0-1 recognizable distractors. Trial 3 has 1-2 recognizable distractors. Trial 4 has 2-3 recognizable distractors. Trial 5 has 3 recognizable distractors.

Another RM-ANOVA on the effect of trial number on accuracy in version 2 showed no effect of trial number on accuracy F (4, 40) = 1.424, p = .265. This indicates that decreasing the time to respond to 1500 ms for version 2 of the task appears to decrease the ability to use this strategy (F1, green line).

For version 3, as can be seen in Figure F1 in blue, accuracy appears to increase from trial 1-2, but the difference between trials 2-5 seems negligible. The fact that accuracy on the first trial was below chance may be due to the high percentage of missed responses on the first trial (M = 25%, SD = 43%) compared to the second trial (M = 8%, SD = 27%). Despite this, a RM-ANOVA on the effect of trial number (1-5) on accuracy showed no effect, F (4, 40) = 1.845, p = .139, indicating that the increasing knowledge of the distractors could not be utilized in the shorter (1300 ms) time frame.

Qualitative results support this result. While participants in version 1 were not asked about this, all participants in versions 2 & 3 were explicitly asked about using the strategy of elimination. While the quantitative results show no indication of this strategy in version 2, still 45% (5/11) participants reported use of this strategy. This supported the further decrease in the timeout from 1500 ms in version 2 to 1300 ms in version 3. The decision was supported by the result of only one participant (1/11) reporting using this strategy in version 3. We considered the combined quantitative and qualitative results sufficient to go ahead with this task design for the EEG experiment.

Appendix G

Extended Stimuli Elimination Process

The third goal of the pre-study was to narrow down the pool of airports to be used in the main study. Seventy-six airports were found on the website shop.nomodesign.com as potential learning material. One airport was used as the example in the instructions (SIN) leaving 75 airports to be used in the task. As the experiment was designed to take one hour, only 40 airports could be learned during this time while still allowing for 12 repetitions of each airport. In versions 1 & 2 of the task, each participant saw 40 airports randomly chosen from the pool of 75. This meant that some participants learned airports that others did not. For the final version of the task used for the EEG experiment, we reduced the pool of 75 down to 40 in order to minimize these task differences between participants.

Primary exclusion process

Between versions 2 & 3 of the task, the pool was narrowed from 75 to 51 with the following method:

1. Since 53% (8/15) of the participants (combined version 1 & 2) mentioned using size of the airports to identify them, a frequency table was made displaying the number of airports containing each different number of runways (Table G1). To create a pool of airports more similar in size, all airports with more than four runways or less than two were removed from the pool, leaving 62.

Table G1

Frequency of number of runways

Runways	One	Two	Three	Four	Five	Six	Seven	Eight
Airports	1	25	21	16	5	5	1	1

2. Six airports were found to look like small Xs (BOM, BTR, BUF, DSM, DXR, and PIA) so five of them were removed, leaving 57.

3. Three airports were found to look like one straight line (HAN, BRN, and GUM). Because this shape is small and distinct, all three were removed, leaving 54.

4. Three airports were removed given our visual observation that they were unusual. OSH had experienced deterioration of the lines during editing. CDG had a very distinct circular shape that stood out from the others, and IAD was unusually large despite being within the allowed number of runways. This exclusion resulted in the 51 airports used in version 3 of the pre-study.

Secondary exclusion process

In order to reach the final 40 needed for the EEG experiment, a RM-ANOVA was performed to compare the effect of number of runways on accuracy. Number of runways had a significant effect on accuracy, F (2, 50) = 5.798, p = .005, with post hoc tests showing that airports with four runways differed in accuracy than airports with three runways t (25) = 3.23, p = .01.

Additionally, we visually estimated a new metric for size, by ranking each object on a scale of 1-3. This was done in case number of runways (while more objective) was not as representative of the image size experienced by a participant. Each airport's number of runways and size can be found in Appendix B. The same RM-ANOVA using the subjective metric of size provided a similar result. There was an effect of size on accuracy F (2, 50) = 3.984, p = .025. Post hoc tests showed that larger (size 3) airports were significantly different than medium (size 2) airports, t (25) = 2.77, p = .031.

Table G2

Frequencies of remaining number of runways

Runways	Two	Three	Four
Airports	18	20	13

With a clear effect of size (in both subjective and objective metrics of size), and an indication that the larger airports drove this effect, we removed from the pool all four-runway airports that were also visually estimated to be a large size (size 3). As can be seen from Table G2, only 13 airports contained four runways, so only two airports with four runways needed to remain in the set. We removed the remaining four-runway airports except the two smallest ones (visually estimated), leading to the final set of stimuli containing 40 airports.