

# Stable individual differences in strategies within, but not between, visual search tasks

A. D. F. Clarke<sup>1,\*</sup>, J. L. Irons<sup>2</sup>, W. James<sup>3</sup>, A. B. Leber<sup>2</sup>, and A. R. Hunt<sup>3</sup>

<sup>1</sup>*Department of Psychology, University of Essex, Colchester, UK*

<sup>2</sup>*Department of Psychology, The Ohio State University, Columbus, USA*

<sup>3</sup>*School of Psychology, University of Aberdeen, Aberdeen, UK*

*\*Corresponding author: a.clarke@essex.ac.uk*

November 30, 2018

## Abstract

A striking range of individual differences has recently been reported in three different visual search tasks. These differences in performance can be attributed to strategy, that is, the efficiency with which participants control their search to complete the task quickly and accurately. Here we ask if an individual's strategy and performance in one search task is correlated with how they perform in the other two. We tested 64 observers in the three tasks mentioned above over two sessions. Even though the test-retest reliability of the tasks is high, an observer's performance and strategy in one task did not reliably predict their behaviour in the other two. These results suggest search strategies are stable over time, but context-specific. To understand visual search we therefore need to account not only for differences between individuals, but also how individuals interact with the search task and context. These context-specific but stable individual differences in strategy can account for a substantial proportion of variability in search performance.

## 1 Introduction

As is common in cognitive psychology, most visual search literature has focused on how the average par-

ticipant performs in the task, despite it being well known that there is a great deal of variability between one subject and the next. From Treisman's work on Feature Integration Theory[22] to the latest incarnation of the Guided Search Model[24], we now have a good understanding of what makes some search targets effortless to find, while others require careful inspection. However, these theories and models have neglected the question of why some observers find visual search so much harder than others. These differences can emerge from several different sources of variation: tiredness[15], information-processing ability, speed-accuracy trade-off, motivation, visual impairments[18], and search strategies[1]. Although their existence has previously been noted[15], a rigorous examination of individual differences in visual search is a challenge that has not been taken up by many researchers, and questions about their importance and stability remain relatively under explored.

In the current study, we focus on one source of individual differences in visual search: strategy. By search strategies, we refer to a collection of behaviours that all observers can freely choose from when they search a display. Examples include choosing to adopt a systematic left-to-right and top-to-bottom strategy[7], or the type of guided search behaviour in which locations that are more likely to contain the target are prioritised[24].

A striking example of the effect of strategy is given by Boot and colleagues[1]. They asked participants to monitor a cluttered display for an object changing colour or suddenly appearing. Large individual differences were found with respect to the number of saccades participants made while monitoring the stimulus, and this was negatively correlated with detection performance. Proulx[20] demonstrated that participants differed in how they prioritised visual features (colour or orientation) when searching through a set of objects for one with a particular colour-orientation combination. These differences in search strategy explained some of the variation in the extent to which participants were distracted by an irrelevant object of a unique size. However, participants were largely unaware of the search strategy they employed, and the majority were unable, or unwilling, to comply with instructions to employ a particular strategy.

Eye movement strategies have also been shown to be an important source of individual differences in visual search efficiency. Nowakowska, Clarke & Hunt[17] designed a simple search paradigm to discriminate between optimal[16] and stochastic [3] search strategies. The task involved searching through arrays of line segments arranged such that those on one side of the display all had a very similar orientation ("homogeneous"), while those on the other side had higher variance ("heterogeneous", see Figure 1 for an example). This meant that targets appearing on the homogeneous side would be highly salient, and targets on the heterogeneous side would be much harder to find. With these stimuli, the optimal eye movement strategy is to only search the heterogeneous half, as targets on the homogeneous side can be detected with peripheral vision. We will refer to this paradigm here as the Split-Half Line Segment task (SHLS). While some participants initially searched the displays near optimally, others carried out strategies counter to this, failing to even match the performance of the stochastic searcher. Furthermore, the degree to which they made saccades in line with the optimal search strategy was strongly correlated with their reaction times. A related version of this paradigm has been used in research investigating eye movement strategies in response to (simulated) hemianopia[18, 19], with similar conclusions:

the full spectrum of individual differences in strategy was observed, from near-optimal to the opposite of optimal. It is therefore not possible to conclude whether optimal or stochastic models better describe search without first explaining individual variability.

A similar range of strategies, from random to near-optimal, has been found by Irons & Leber with the Adaptive Choice Visual Search (ACVS) paradigm[9], designed to explore how feature-based attention is used. This paradigm involves stimuli made up of small coloured boxes (red, blue, green and a fourth colour that varies from red, through purple, to blue and back again) with numerals written inside them (See Figure 1). The target is defined as a red or blue box containing one of four numerals (e.g., 2-5), and on each trial one target of each colour is present. The participant's task is to find one of either target as quickly as possible and report the numeral. On trials in which the fourth colour is red (or close to red), participants should search through the blue boxes and report the blue target, as there will be fewer distracters. As the fourth colour changes to through purple to blue, participants should update their strategy and search for the red target. The results showed that participants varied substantially along two key dimensions: how frequently they used the more effective target colour to search (varying from chance performance to near optimal), and how often they changed between colours. Further work[10] has shown that these differences are stable over time (between one and ten days) with test-retest correlations of around  $r = 0.83$  (95% confidence interval = [0.72, 0.90]).

Another example of differences in search strategy comes from the foraging literature [14, 12]. In this context, foraging is a visual search task in which there are multiple targets on each trial. Participants were asked to search through a set of items from four categories, with two categories classed as targets. In the conjunction condition (searching for red-horizontal and green-vertical line segments among red-vertical and green-horizontal distracters), most observers searched in runs, finding all the targets of one target category, and then switching and finding the targets in the other category. This strategy has previously been observed in animal foraging

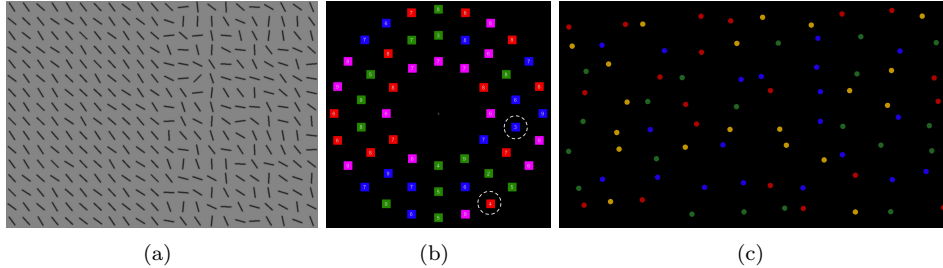


Figure 1: Example stimulus from the (a) *split-half line segment* (SHLS), (b) *adaptive choice visual search* (ACVS) and (c) *mouse click foraging task* (MCFT) paradigms.

[5], and suggests holding one complex target template in mind at a time is a better strategy than switching templates. However, a sub-set of observers, termed ‘super-foragers’, were able to change between search target categories with no cost to performance. This study has been run on touch displays[14] (observers are asked to tap on as many targets as they can as quickly as possible) and a 3D computer game (in which participants navigate a squirrel through a virtual environment picking up targets with different feature combinations)[IN PREP], with super-foragers present in both versions. While test-retest reliability has not been measured explicitly for the foraging paradigm, the task was used as a measure to assess the effect of a six day long mindfulness retreat on cognitive performance[8]. From a re-analysis of these data, we can estimate that the test-retest reliability for the mean run length is  $r \approx 0.7$  for the feature condition and  $r \approx 0.88$  for the conjunction search.

Previous research has investigated the relationship between these behaviours to psychometric measures, but to date, these differences have not shown strong correlations with other attributes. In the ACVS paradigm, Irons & Leber found no evidence of a correlation between the proportion of optimal choices made by observers and measures of visual working memory; trait impulsivity[9]; novelty seeking; need for cognition; and intolerance of uncertainty[10]. Similarly, the differences foraging behaviour are not accounted for by working memory or inhibitory control[11].

A common theme emerging from these studies is

the observation that individual strategies vary in their degree of effectiveness or optimality. However, “visual search” encompasses a wide range of specific tasks, each tapping into a different aspect of behaviour (e.g. feature-based attention, information sampling). The aim of the present study is to investigate the extent to which individual differences are stable across different visual search paradigms. Are observers who use the optimal strategy in the SHLS paradigm also more optimal in the ACVS task? Does it make sense to talk about ‘super-searchers’ who show above average performance in a range of search tasks (analogous to the ‘super-recognizers’ of the face-recognition literature [21])? Are the super-foragers consistently better or worse than more typical searchers in the other two paradigms? As a secondary question, we will measure the test-retest reliability of the differences found in the SHLS paradigm, and compare it with existing estimates of reliability for ACVS and MCFT.

## 2 Methods

The methods and planned analysis for this study were registered on the Open Science Framework<sup>1</sup> before data collection started.

<sup>1</sup><https://osf.io/scv4u/>

## 2.1 Participants

64 students from the University of Aberdeen took part in this study<sup>2</sup>. Participants were compensated for their time with either course credit or £15. All participants gave informed consent. The study was approved by the University of Aberdeen Psychology Ethics Committee.

Sample size was determined in part by a power analysis, and in part due to constraints with counter-balancing.  $n = 64$  participants means we should be able to detect correlations with  $r > 0.34$  with  $\alpha = 0.05$ ,  $\beta = 0.80$  between the different visual search paradigms. The sample is therefore of sufficient size to detect relatively small correlations.

## 2.2 Materials and Procedures

The study consists of three paradigms from the visual search literature in which large individual differences have been found [17, 9, 14]. Example stimuli can be seen in Figure 1. A brief overview of each paradigm is given below, with full details in *supplementary materials*. The three tasks were completed over two sessions, approximately one week apart. The SHLS was run in both sessions allowing us to measure test-retest reliability. The order in which participants completed the tasks was counter-balanced. There are 16 different possible orders of tasks/conditions; four participants completed each order for a total of 64.

The display was presented on a 17-inch CRT monitor with a resolution of  $1400 \times 1050$  ( $n = 40$ ) or  $1600 \times 1200$  ( $n = 24$ ). Stimulus generation, presentation and data collection were controlled by MATLAB and the psychophysics and eyelink toolboxes [2, 4] run on a Powermac. Participants sat  $\approx 47$ cm from the screen.

### 2.2.1 Split-half Line Segments

Stimuli consisted of arrays of black oriented line segments against a grey background. The target was oriented  $45^\circ$  clockwise, while the distractor items

had a random orientation with a mean of  $45^\circ$  anti-clockwise. The variance was low ( $18^\circ$ ) on one half of the display to create a homogeneous texture, and high ( $95^\circ$ ) on the other side to create a heterogeneous texture. This means that when the target is present on the homogeneous half, it can be easily be detected with peripheral vision, but when it is in the heterogeneous half, it is much harder to detect. There were 160 trials in total and homo- and heterogeneous sides of the display were randomly varied from trial to trial. The position of the dominant eye was recorded using a desktop-mounted EyeLink 1000 eye tracker (SR Research, Canada).

This paradigm was carried out twice, once in each testing session, to give us an estimate of how consistent participants are in their search strategy over time.

### 2.2.2 Adaptive Choice Visual Search

Each search display was composed of red, blue, green and variable-coloured small squares arranged in three concentric rings around fixation. Variable distractors changed colours from trial-to-trial according to a 24 trial cyclical pattern: the distractors would be red for five trials, then across a period of seven trials, they would gradually change colour from red to blue. The variable distractor would then be blue for five trials, and then gradually transition back to red.

A white digit appeared inside each square. Participants were informed that two targets - a red square and a blue square each with a digit between 2 and 5 - were embedded in every search display. The two target digits were always different, to enable us to distinguish the chosen target. The remaining red, blue and variable squares all contained digits between 6-9. Green squares could contain any digit between 2-9. The location of the targets and distractor within the search display were randomized on each trial. Participants were only required to find one target on each trial, and they were free to search for either one.

### 2.2.3 Mouse Click Foraging Task

In the feature foraging task, search displays contained small red, green, yellow and blue circles. For half of

<sup>2</sup>data from an additional 11 participants was discarded due to being recorded with an inappropriate screen resolution.

the participants, targets were red and green circles, and for the other half of participants, targets were blue and yellow circles. Participants were asked to collect all of the targets within a trial by using the mouse to click on each target. Clicking on a target caused it to disappear from the display. If the participant clicked erroneously on a non-target, the trial immediately ended and a replacement trial begun. In the conjunction foraging task, search displays were composed of both circles and squares. For half of the participants, the shapes were red and green, and for the remaining participants the shapes were blue and yellow. Targets were defined by conjunctions of colour and shape (e.g., red squares and green circles, with red circles and green squares as distractors). The assignment of targets and distractors was assigned at random for each participant. The procedure was otherwise identical to the feature foraging task.

### 3 Results

#### 3.1 Replication of each task

A brief summary of participants' behaviour is given below. More time is spent on SHLS as the test-retest validity of it has not previously been assessed. Further analysis and details can be found in the *supplementary materials*.

##### 3.1.1 Split-half Line Segments

Our results are consistent with the original SHLS study[17]: we find a large range of individual differences in search reaction time and accuracy (see Figure 2). Furthermore, these differences are stable across the two sessions, with Pearson's  $r \in [0.71, 0.89]$  (95% confidence interval) for accuracy in finding hard targets. We get similar scores for the correlation in reaction times between sessions *a* and *b* for hard targets, ( $r \in [0.54-0.81]$ ), easy targets ( $r \in [0.52-0.80]$ ) and target absent trials ( $r \in [0.66-0.86]$ ).

We can also look at the initial search strategies adopted by our participants 2(c, d). Again, we see large and stable individual differences across the two

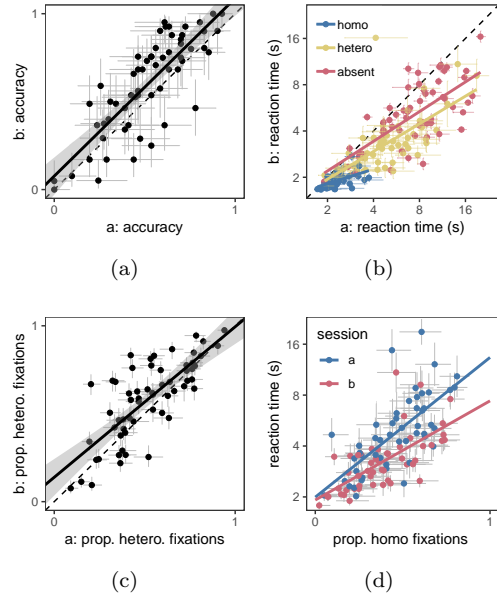


Figure 2: Correlation between the two sessions of the SHLS paradigm for (a) accuracy (TP-heterogeneous trials only); (b) reaction times and (c) search strategy (TA trials only). (d) Initial search strategy correlates with reaction times in both sessions. Each point represents a participant and the error-bars indicate 95% confidence intervals.

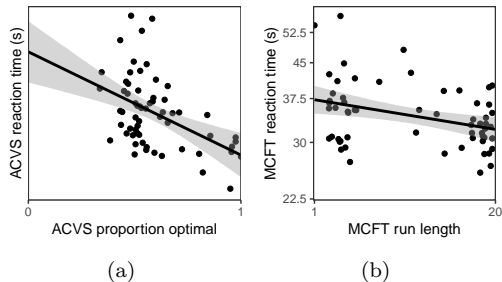


Figure 3: Correlation between strategy and reaction times for (a) ACVS and (b) MCFT. Each point represents a participant.

sessions (test-retest  $r \in [0.63, 0.86]$  for the proportion of the first five saccades to the heterogeneous half of the display for target absent trials). More importantly, as with [17], we see that the search strategies give a good correlation with reaction times in both session a,  $r \in [0.52, 0.82]$  and session b,  $r \in [0.50, 0.80]$ .

### 3.1.2 Adaptive Choice Visual Search

We measured an individual’s strategy as the percent of plateau trials in which the individual chose the optimal target (i.e., the target with the fewest distractors: When the variable distractor was red, the optimal choice was blue, and vice versa). The results for the ACVS were consistent with previous findings [9, 10]. We can clearly see from figure 3(a) that there are individual differences in the proportion of optimal targets reported (range 33.62% - 100.00%,  $\bar{x} = 59.15$ ,  $s = 16.54$ ) and the mean ( $\log_2$ ) reaction times (range 1.90 - 4.80 seconds). As with the SHLS task, the degree to which participants follow the optimal strategy is correlated with reaction times ( $r \in [-0.65, -0.25]$ ).

### 3.1.3 Mouse Click Foraging Task

The main measure of interest was average run length per trial in the conjunction condition, with a run defined as a succession of one or more of the same target type, which was followed and preceded by the other target or no target. The average run length

was the mean number of target selections in a run. The multiple-target foraging results were in line with previous findings [14, 12], with shorter run lengths for feature foraging ( $\bar{x} = 3.16$ ,  $s = 3.14$ ) than conjunction foraging ( $\bar{x} = 11.73$ ,  $s = 7.09$ ). This suggests more frequent foraging for multiple targets concurrently when those targets were defined by features than by conjunctions. Figure 3(b) depicts the individual differences in terms of run length and the correlation with reaction time ( $r \in [-0.55, -0.10]$ ).

## 3.2 Correlations Between Tasks

We have successfully replicated the previous findings around individual differences in visual search strategy in each of the three tasks. Furthermore, the SHLS task has been shown to have good test-retest reliability, similar to that of the ACVS and MCFT tasks. Given this, we now investigate the extent to which an individual’s performance in one of the tasks tells us about how well they will do in the other two.

The results show that the correlations between the strategy metrics in the three tasks (Figure 4) are weak. Perhaps even more surprisingly, there is also little evidence for meaningful correlations between reaction times in the different tasks. Even if we optimistically take all data together as suggesting a robust correlation in reaction times from one task to another, the mean correlation over the three tasks is only  $r = 0.2$ , implying that this correlation accounts for  $R^2 = 0.04 = 4\%$  an individual’s performance.

## 4 Discussion

We successfully replicated the wide range of individual differences in strategy and performance that had previously been observed in each of these three visual search paradigms, with a larger sample size than the original experiments. Surprisingly, however, the between-paradigm correlations give  $R^2 = 0.04$ , and this is a generous interpretation that would fail to pass the usual criteria for null hypothesis significance testing. Knowing how one person will behave in one of these paradigms apparently tells us very little about how they will perform in the others. This can

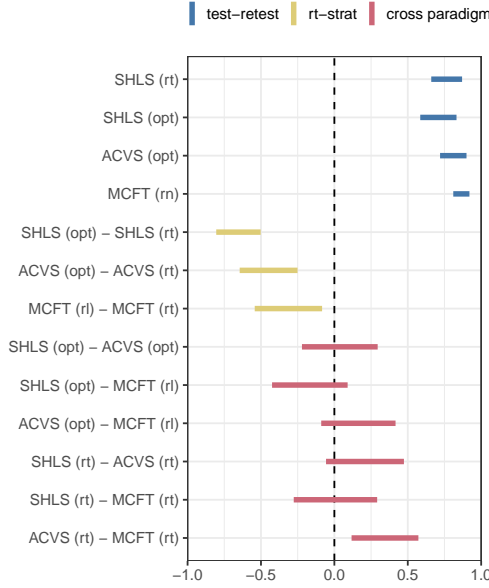


Figure 4: The between- and within-task correlations for the three different search tasks. The bars indicate the 95% confidence intervals for Pearson’s correlation coefficient. Blue bars represent test-retest scores for each task for reaction times (*rt*), optimality (*opt*) or run length (*rl*). Yellow bars indicate how well the strategy measures predict reaction times, while the red bars show that performance in one task is not a good indication of performance in another, either for reaction times of strategy.

be contrasted with the relatively high test-retest correlations of all three of the tasks individually, which range from a lower bound of 0.58 to an upper bound of 0.92 .

There are many reasons why two measurements might be uncorrelated, such as range restriction or measurement noise. The test-retest correlations on each of the individual visual search tasks rule out many of these alternatives, leaving a true absence of shared variance between these tasks as a likely explanation. One might have expected reaction time to be at least modestly correlated from one search task to the next, even in the absence of correlations in strategy, as a general factor like an individual’s speed-accuracy trade-off, motivation or conscientiousness might lead to better or worse overall performance across all the tasks. However, even reaction times were not highly correlated between tasks. What we did observe were correlations between our measures of strategy and reaction time within each task, suggesting our strategy metrics are predictive of search performance, and reinforcing the interpretation that an individual who is strategic in one task is not necessarily strategic in the others.

Although the tasks in this experiment all have visual search in common, they also have unique aspects that appear to have resonated with particular individuals’ strengths, and not others. Our definition of a successful strategy in the SHLS task was fixating the locations that provide new information. In the ACVS task, a successful strategy meant appropriately altering search goals to match changes in the environment. In the MCFT task, success involved minimizing cognitive load by minimizing target switching. Each of these tasks clearly taps into unique aspects of visual search strategies, and performance on one of these aspects has no bearing on any of the others.

Individual differences pose a challenge for efforts to devise a comprehensive model of visual search. Our understanding of the underlying mechanisms of visual search is based predominantly on experiments that systematically vary details of the search task and measure their effects on performance. This approach has led to many important insights, for example, about the kinds of visual features that can be used to guide attention (e.g. [22, 25]; how attentional

control settings filter out distractors (e.g. [6, 26]); and biases in attention, such as a bias towards unexplored locations (e.g. [13]). These insights are based on search performance averaged over some number of individual participants. For all three of the experiments included in the current study, however, the average performance would be highly misleading, as it would describe very few of the individuals who completed the tasks. In the original SHLS study [17], for example, the original aim of the experiment was to assess whether search behaviour could be better described by an optimal [16], versus a stochastic [3], model. Considering only the average search performance, the stochastic model was a good explanation. Underlying that average performance, however, was a very large spectrum of search behaviour, replicated here, some of which would be clearly categorized as optimal, and some as stochastic, and some as neither. The original question needed to be refined into new question: for whom is search optimal and for whom is it stochastic? The current findings add even further challenges for researchers interested in understanding visual search, by suggesting we need to not only account for individual differences, but also the interaction of a given individual with a particular search context.

We view these findings, not as a disappointing null result, but as thought-provoking and exciting. Vogel and Awh[23] argued that studying individual differences in cognitive psychology (in their case, working memory) provides valuable insight to containing potential theories of the underlying cognitive mechanisms. Our results suggest that context and structure of the task also needs to be taking into account and we hope that studying how an individual’s behaviour varies across different search tasks will lead to the development of a more comprehensive theory of search.

## Acknowledgements

We would like to thanks Anna Nowakowska, Árn Kristjánsson and Ian Thornton for sharing data and their helpful comments and suggestions. We would like to thank Charles Rigitano and Jacqueline Von Seth who helped with data collection. We acknowl-

edge the funding support of the James S. McDonnell Foundation (Scholar Award to ARH) and NSF BCS-1632296 to ABL.

## References

- [1] Walter R Boot, Arthur F Kramer, Ensar Becic, Douglas A Wiegmann, and Tate Kubose. Detecting transient changes in dynamic displays: The more you look, the less you see. *Human Factors*, 48(4):759–773, 2006.
- [2] David H Brainard. The psychophysics toolbox. *Spatial vision*, 10:433–436, 1997.
- [3] Alasdair DF Clarke, Patrick Green, Mike J Chantler, and Amelia R Hunt. Human search for a target on a textured background is consistent with a stochastic model. *Journal of vision*, 16(7):4–4, 2016.
- [4] Frans W Cornelissen, Enno M Peters, and John Palmer. The eyelink toolbox: eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4):613–617, 2002.
- [5] Marian Dawkins. Shifts of ‘attention’ in chicks during feeding. *Animal Behaviour*, 19(3):575 – 582, 1971.
- [6] Charles L Folk, Roger W Remington, and James C Johnston. Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human perception and performance*, 18(4):1030, 1992.
- [7] Iain D Gilchrist and Monika Harvey. Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, 14(4-8):704–715, 2006.
- [8] Mirjam Hartkamp and Ian M Thornton. Meditation, cognitive flexibility and well-being. *Journal of Cognitive Enhancement*, 1(2):182–196, 2017.
- [9] Jessica L Irons and Andrew B Leber. Choosing attentional control settings in a dynamically



- changing environment. *Attention, Perception, & Psychophysics*, 2016.
- [10] Jessica L Irons and Andrew B Leber. Characterizing individual variation in the strategic use of attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, pages 1–18, 2018.
  - [11] Ómar I. Jóhannesson, Árni Kristjánsson, and Ian M. Thornton. Are foraging patterns in humans related to working memory and inhibitory control? *Japanese Psychological Research*, 59(2):152–166.
  - [12] Ómar I Jóhannesson, Ian M Thornton, Irene J Smith, Andrey Chetverikov, and Árni Kristjánsson. Visual foraging with fingers and eye gaze. *i-Perception*, 7(2):2041669516637279, 2016.
  - [13] Raymond M Klein. Inhibition of return. *Trends in cognitive sciences*, 4(4):138–147, 2000.
  - [14] Árni Kristjánsson, Ómar I Jóhannesson, and Ian M Thornton. Common attentional constraints in visual foraging. *PloS one*, 9(6):e100752, 2014.
  - [15] Norman H Mackworth. The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1(1):6–21, 1948.
  - [16] Jiri Najemnik and Wilson S Geisler. Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3):4–4, 2008.
  - [17] Anna Nowakowska, Alasdair D F Clarke, and Amelia R Hunt. Human visual search behaviour is far from ideal. *Proceedings of the Royal Society B: Biological Sciences*, 2017.
  - [18] Anna Nowakowska, Alasdair DF Clarke, Arash Sahraie, and Amelia R Hunt. Inefficient search strategies in simulated hemianopia. *Journal of Experimental Psychology: Human Perception and Performance*, 42(11):1858, 2016.
  - [19] Anna Nowakowska, Alasdair D.F. Clarke, Arash Sahraie, and Amelia R. Hunt. Practice-related changes in eye movement strategy in healthy adults with simulated hemianopia. *Neuropsychologia*, 2018.
  - [20] Michael J Proulx. Individual differences and metacognitive knowledge of visual search strategy. *PLoS One*, 6(10):e27043, 2011.
  - [21] Richard Russell, Brad Duchaine, and Ken Nakayama. Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, 16(2):252–257, 2009.
  - [22] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.
  - [23] Edward K Vogel and Edward Awh. How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Current Directions in Psychological Science*, 17(2):171–176, 2008.
  - [24] Jeremy Wolfe, Matthew Cain, Krista Ehinger, and Trafton Drew. Guided search 5.0: Meeting the challenge of hybrid search and multiple-target foraging. *Journal of vision*, 15(12):1106–1106, 2015.
  - [25] Jeremy M Wolfe. Guided search 4.0. *Integrated models of cognitive systems*, pages 99–119, 2007.
  - [26] Steven Yantis and Howard E Egeth. On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3):661, 1999.