The Role of Attention in Perceiving Multiple Visual Dimensions: Testing Boolean Map Theory Using a Summary Statistics Paradigm

Matthew Martin

A thesis in
the Department of Psychology

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Arts (Psychology) at
Concordia University
Montreal, Quebec, Canada

December, 2017

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Matthew Martin

Entitled: The Role of Attention in Perceiving Multiple Visual Dimensions: Testing Boolean Map Theory Using a Summary Statistics Paradigm

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Psychology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________ Chair
Andrew Chapman

______________________ Examiner
Rick Gurnsey

______________________ Examiner
Benjamin Eppinger

______________________ Thesis Supervisor(s)
Aaron Johnson

____

Approved by ______________________
Virginia Penhue Chair of Department or Graduate Program Director

____
André Roy Dean, Faculty of Arts and Social Sciences

Date February 6, 2018

ABSTRACT

The Role of Attention in Perceiving Multiple Visual Dimensions:

Testing Boolean Map Theory Using a Summary Statistics Paradigm

Matthew Martin

It has been demonstrated that observers are able to extract the statistical properties of a set of objects, such as the mean size of a set of circles, with surprising efficiency. However, debate surrounds the role of attention in this process, with some theorists claiming that it is ‘automatic’ and done ‘in parallel’, arising from a ‘preattentive’ processing stage. Conversely, other theorists claim that there is a limited attentional capacity for the perception of summary statistics. Further, there is more specific debate surrounding the processes by which summary statistics are computed across different dimensions of visual features simultaneously, such as mean orientation and size. Some findings suggest a cost of encoding two dimensions simultaneously (Emmanouil & Treisman, 2008), whereas others have found evidence of unlimited capacity encoding (Attarha & Moore, 2015b). The Boolean map theory of visual attention would claim that the limiting attentional factor in perceiving multiple visual dimensions simultaneously is stimulus exposure duration. In the current study, participants were shown a set of stimuli which varied in colour, orientation, and size for a fixed exposure duration of 200ms. They were tasked with attending to either one, two, or all three dimensions present in the stimuli. Results indicated that psychophysical performance decreased in terms of both decreased accuracy and increased response times as the number of dimensions attended to increased. The results suggest a bottleneck of selective attention caused by a fixed and difficult exposure duration, consistent with the predictions of Boolean map theory.

Acknowledgements

I would like to thank all my friends and family who have helped me throughout my Master's degree. First and foremost, I would like to thank my mom and dad who have given me endless support and love throughout my life and continue to put up with me. Love you guys.

I would also like to thank my supervisor, Aaron Johnson for all of his support and guidance over the past two years. Thanks for all of the opportunities you have given me and for all of the geeky psychophysics I have learned from you. I would also like to thank all the members of the Concordia Vision Lab, especially my research assistant, Brandon Huard-Nicholls who dutifully spent many hours collecting data for me. Also a big thank-you to all of the lab members, friends, colleagues and friendly acquaintances who were willing to sit through hours of psychophysics over the course of my degree. Thank you all.

Table of Contents

List of Figures

List of Tables

The Role of Attention in Perceiving Multiple Visual Dimensions:
Testing Boolean Map Theory Using a Summary Statistics Paradigm

Visual attention can be defined as the processes by which the cognitive system selects some visual information from our environment to be processed to a greater extent than other ‘unattended’ information. The current thesis is concerned with attention for visual features within a scene or display such as colour, orientation, and size. The primary question is how observers attend to and process multiple types of visual features simultaneously. The field of visual search has shown that there are important differences between attending to a single type of visual feature (also referred to as a visual *dimension*, e.g., colour) versus two or more types of features simultaneously (e.g., size or orientation; Healey & Enns, 2012; Wolfe & Horowitz, 2004). A target which differs from distractors along one feature dimension (e.g., red amongst blue distractors) tends to ‘pop-out’ and yield very efficient detection when searched for. This phenomenon is referred to as *feature search*. Conversely, a target that shares a dimension with the distractors and which requires differentiation across two (or more) dimensions (e.g. colour and shape) seems to lose the pop-out effect and yields inefficient detection when searched for. This phenomenon is referred to as *conjunction search*. A longstanding and influential theory in the field of visual search is feature-integration theory, which proposes that visual features are “registered early, pre-attentively, and in parallel” (p. 98), whereas effortful selective (serial) attention is required to combine two or more features to perform a conjunction search (Treisman & Gelade, 1980). A marker of ‘pre-attentive’ processing in feature-integration theory is whether the presence of more or less distracting items (i.e., the set size) hinders search performance. When a task is ‘pre-attentive’, including more items in the display has relatively little impact on the response time to detect the target. Presumably, every item in the set is searched through all at once (i.e., in parallel), so the slope of the line representing the relationship between the number of items in the set and response time is relatively flat. When selective attention is involved (e.g. in conjunction search), the slope of the line between response time and set size is much steeper. Since items must be searched through one-by-one (i.e. serially), the task takes longer to complete. According to feature integration theory, parallel processing is a marker of pre-attentive processing.

Set summary statistics

Visual search examines the speed with which observers can detect a target amongst a set of distracting stimuli. The field is related to other research examining how sets of objects are themselves represented. More recent work has focused on how observers perceive the average (or *summary statistic*) features of a set (or *ensemble*) of objects. Arieli (2001) conducted a series of foundational experiments which demonstrated that observers have a tendency to represent the average feature of a set of objects, rather than the features of individual objects in a set. For example, when shown a set of circles that varied in size, observers were much more likely to indicate that a probe circle was a part of the set when its size was the average of the set of circles, rather than the size of a circle that was actually present in the set. Arieli also found that the thresholds for a mean-discrimination task, where participants must determine if a probe circle is larger or smaller than the mean of a set, were largely the same across different set sizes. This provided evidence in favour of set-size independent processing, and Arieli concluded that the visual system represents the average of a set of features, and not the individual elements. He also went on to speculate that this could explain a number of findings in the visual search literature, including the pop-out effect: Targets pop out when their feature is highly different from the average of the set.

Work since Arieli (2001) has elaborated on the role of attention in perceiving the average properties of a set. There has been a debate about whether or not representing summary properties is a pre-attentive process. Early work by Chong and Treisman (2003) showed that mean discrimination thresholds are relatively unaffected by stimulus exposure duration or simultaneous versus successive presentation. They also demonstrated that estimates of the mean size of a set of circles are almost as accurate as those of a single circle, and are relatively unaffected by the distribution of sizes, be it normal, uniform, two-peaked, or homogeneous. In follow-up research, Chong and Treisman (2005) showed participants a set of green and red circles of heterogeneous sizes. Participants had to report the mean size of the circles belonging to one or the other colour. They were either given a pre-cue to selectively attend to one colour, or a post-cue, where advance information about which colour of circle would be probed was not provided. The researchers found no pre-cue advantage in computing the average size of the relevant coloured circles, indicating to the researchers that representing average properties of sets is “automatic”, and “precedes the limited capacity bottleneck that forces selective attention” (p.

899). These claims have been a matter of debate in the subsequent literature. For example, some researchers have found inattentional blindness to set summary statistics as well as dual task interference when forming set summary representations (Jackson-Nielsen, Cohen, & Pitts, 2017). Other researchers have shown that pre-cueing offers the same accuracy benefit for set summaries as for single-object features, indicating a similar effect of attention for both set summaries and single objects (discussed further below; Huang, 2015).

Aside from extracting mean size, researchers have also shown that observers are able to efficiently extract the mean of other visual features including orientation, colour, and texture (shape); direction of motion, speed, and position of objects; and emotion and identity of faces (Attarha & Moore, 2015a, 2015b; Brady & Alvarez, 2011; Brand, Oriet, & Sykes Tottenham, 2012; Maule & Franklin, 2015; Whitney, Haberman, & Sweeny, 2014). As visual displays rarely have a single feature dimension, some researchers have begun to examine how observers are able to form average representations across sets that differ along two visual dimensions. For example, Emmanouil and Treisman (2008) used a pre/post-cue design and showed participants two sets of circles on the right and left of a screen that differed in their relative size and speed of motion. The task was to select which side of the screen had the larger average size or which side had the faster average speed. In the pre-cue condition, participants were told which dimension would be queried (size or speed of motion) prior to viewing the objects. In the post-cue condition, participants were given no advance indication of which dimension would be queried. Pre-cueing the relevant dimension was shown to improve accuracy when making judgements of mean size, but not mean speed of motion. Follow-up experiments showed that the pre-cueing effects were larger for both visual dimensions when different objects were used to vary the size and speed of motion (such as stationary circles which varied in size, and moving Xs which varied in speed of motion). They also demonstrated that there was a significant pre-cue advantage for determining the average size of a set of circles and the average orientation of a set of lines. The authors concluded on the basis of these findings and previous research that “there is a cost in the statistical processing of two means only when these are computed on separate dimensions, as opposed to a single one” (p. 952).

One possible alternative explanation for the findings of Emmanouil and Treisman (2008) is that participants had to form average representations across two sets of stimuli, as well as across the two dimensions. Specifically, in the post-cue condition, participants had to form two representations of mean size, and two representations of mean speed (or orientation), given that there were two sets of stimuli on the screen. Performance decrements may reveal processing limitations in forming multiple representations *within* a dimension, as opposed to merely *between* dimensions, as the authors argued. This criticism lead Attarha and Moore (2015b) to conduct a series of experiments aimed at examining the relative impact of forming multiple average representations within a dimension versus forming average representations between two dimensions within a single set of stimuli.

In their experiments, Attarha and Moore (2015b) used circular sinusoidal gratings that varied in orientation and size. In their first experiment, four sets of gratings were presented in the four quadrants of the screen. Three of the sets of gratings had the same average size and orientation, whereas one set was the 'oddball' set, the average size and orientation of which differed from the other three by +/- 15° of orientation or +/- 0.46° of visual angle. Participants were tasked with reporting the average size and orientation of the oddball set in a four-alternative forced choice task (4AFC; 'left and small', 'left and large', 'right and small', 'right and large'). Presumably, the task required the formation of a representation of average size and orientation for each of the four sets. The researchers used an extended version of the simultaneous-sequential task to determine if there are capacity limitations in forming multiple within-dimensions summaries (Scharff, Palmer, & Moore, 2011). In the simultaneous condition, all four of the sets of were presented for a certain amount of time (e.g. 200ms), whereas in the sequential condition, half of the sets were presented first for 200ms, followed by the other half for 200ms. This difference in stimulus onset asynchrony allowed the researchers to probe whether the number of sets to be processed at a given time had an effect on accuracy. If performance were equal between processing four sets at the same time versus two, it would be evidence in favour of the hypothesis that the number of sets to average over has no impact on performance (i.e. it is done in parallel, or of 'unlimited capacity'). Conversely, if performance were better in the sequential condition where fewer sets need to be processed simultaneously, it would be evidence in favour of the hypothesis that forming multiple set summary representations is of 'limited capacity' (i.e., the visual system can only process a limited number of sets at once; see Attarha & Moore, 2015b, and Scharff et al., 2011, for a more complete description of the method).

Over the course of the experiment, accuracy performance was kept between floor and ceiling by a coarse tracking procedure which altered the stimulus exposure duration according to performance in the preceding block. If performance in the simultaneous condition of the previous block was within 10% of chance or perfect performance, the stimulus exposure durations of both the simultaneous and sequential conditions were increased or decreased by 10ms on the next block. The authors do not describe a procedure used to affect the perceptual discriminability of the oddball versus majority sets, other than manipulating the stimulus exposure duration. The result of the experiment demonstrated that performance was worse in the simultaneous condition compared to the sequential condition, indicating that there are capacity limitations in forming multiple representations of two dimensions simultaneously (i.e., representing multiple 'within-dimension summaries', is of 'limited capacity', following Attarha and Moore's nomenclature). A previous study had already shown that this was the case for the dimension of orientation alone (Attarha & Moore, 2015a).

After demonstrating capacity limitations in forming multiple within-dimension summary representations, and that these limitations were not due to memory decay, Attarha and Moore (2015b) conducted an additional experiment which aimed to test whether there are capacity limitations in forming a single summary representation across two dimensions simultaneously (i.e., forming multiple 'between-dimension summaries'). In this experiment, participants were presented with a single set of 16 circular sinusoidal gratings which varied in orientation and size (see Figure 1). Participants had to perform three tasks over the course of three testing sessions: report the average orientation ('left', 'right'), report average size ('small', 'large'), or report average size and orientation ('left and small', 'left and large', 'right and small', 'right and large') of all 16 gratings. As before, in the simultaneous condition, the entire set was presented all at once, whereas in the sequential condition half of the set was presented first, followed by the second half. Equal performance between the simultaneous and sequential conditions would be taken as evidence of unlimited capacity processing.

In all three task conditions (orientation, size, and orientation and size) they found that performance was the same for the simultaneous and sequential conditions, indicating that forming a single set representation across one or two dimensions is an unlimited-capacity process. The authors concluded that "multiple within-dimension summary representations are mediated through at least some limited-capacity processes… whereas *between-dimension summaries are mediated entirely through unlimited- capacity processes*" (Attarha & Moore, 2015b, p. 13, emphasis added).

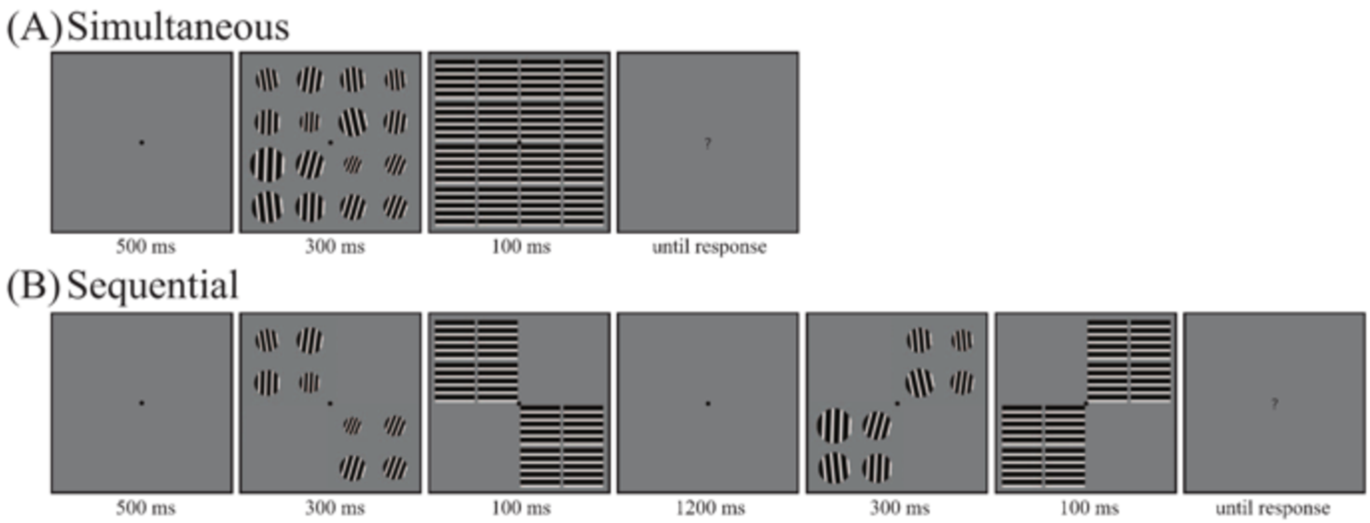

Figure 1. Stimuli from Attarha & Moore's (2015b) third experiment.

One possible critique of the findings of Attarha and Moore (2015b) is that in their experiment, the durations of stimulus exposure times were different between the orientation, size, and dual task conditions. As in their first experiment, the researchers used a coarse tracking design that altered the stimulus exposure durations on a block-by-block basis so that performance was between floor and ceiling in the simultaneous condition. If performance was within 10% of perfect or chance performance, stimulus exposure durations for both the simultaneous and sequential conditions were decreased or increased by 10ms in the next block. The reported average stimulus presentation durations for the orientation, size, and orientation and size tasks were 60, 90, and 230ms, respectively. It may be that stimulus exposure duration is a critical variable in forming summary representations across multiple conditions (as some theories of visual attention would predict, see below). By not keeping this variable constant, Attarha and Moore (2015b) may have inadvertently biased their results in favour of finding no capacity limitations. Additionally, while it may be true that forming summaries across two dimensions can be done without evidence of processing limitations, there may be an upper limit to the number of dimensions that can be processed simultaneously. To the present author's knowledge, no experiment has yet examined observers' capacity to form summary representations across more than two dimensions simultaneously.

Boolean map theory

The dichotomy between pre-attentive processes versus selective attention in representing set summaries is reminiscent of the discussion that has surrounded visual search in the wake of feature-integration theory. The form of the debate is perhaps being driven in part through an attempt to explain the phenomena of set summary statistics in terms of feature integration theory itself (e.g., Chong & Treisman, 2003, 2005). It should be noted that subsequent work in the field of visual search has shown that the strict dichotomy between pre-attentive (parallel), feature search and selective attention (serial) conjunction search does not adequately capture the full range of phenomena. For example, it has been shown that single-feature search can lose the pop-out effect and exhibit inefficient search behaviour when the perceptual discriminability between the target and distractors is low. More recent theories favour a continuum of search efficiencies and seek to define visual attributes which can ‘guide’ selective attention (Wolfe & Horowitz, 2004).

A recent theory of visual attention, Boolean map theory, may offer an account of the data structures and processes that underlie visual attention in more detail than does the feature-integration theory (Huang & Pashler, 2007; Huang, Treisman, & Pashler, 2007). Briefly, Boolean map theory claims that observers are able to represent only one visual feature within a dimension at a time (e.g., green or red, but not both). Representing multiple different dimensions simultaneously can be done, but requires extra computational steps to compute the conjunction or disjunction of multiple features. The Boolean map theory of visual attention would claim that it is possible for observers to form representations across more than one visual dimension simultaneously, but each extra dimension to be attended requires additional processing time to compute. Boolean map theory also claims that the ability to form a representation (or representations) across a set of stimuli is set-size independent. Thus, multiple-dimension summaries are of unlimited capacity, but stimulus exposure duration is a critical variable in observers’ ability to form representations across multiple dimensions of visual features. Performance decrements associated with selective attention might therefore be observed if stimulus exposure duration is kept at a constant-but-difficult level across tasks that involve forming representations across one or more dimensions of visual features.

A few more details about The Boolean map theory of visual attention are worth noting before describing the structure of the present study. Boolean map theory seeks to answer the question of what visual properties are consciously accessible to an observer at a given time. In answering the question, the theory makes a critical distinction between *access* and *selection*. Access refers to the type of information content and structure that can be consciously accessed or reported by an observer. It can be thought of as the data structure that visual attention uses to represent the world. Selection refers to the mechanisms by which the data structure is created through the process of top-down control.

Access

Boolean map theory claims that observers can only access a single feature value within a dimension (e.g. all red objects, or all horizontal) at a given time, along with the spatial locations of its occurrence. For example, if the colour red is attended to (i.e., selected), then each location where red appears in a scene is represented (or mapped) along with an explicit representation of the colour value at each location (a ‘feature label’). Critically, no two values within a dimension (e.g. red and green) can have an explicit feature label at the same time. Thus, multiple feature values within a dimension are not simultaneously accessible according to the theory, whereas if a feature has been selected, multiple locations are simultaneously accessible. The *Boolean* aspect of the theory describes the fact that the visual world is represented in two binary (i.e., Boolean) levels: regions which are selected and regions which are not. The basic data structure of visual attention according to the theory is called a *labeled Boolean map* (Huang & Pashler, 2007).

Observers can also select across an entire visual dimension such as colour, but doing so prevents access to explicit representations of individual features within that dimension if the features are heterogeneous (e.g. the labeled location of specifically red or green objects would not be accessible in a scene with different colours). The resulting dimensional label still provides access to the spatial layout of the set of objects defined by the visual dimension. For example, if an observer selects by the colour dimension, no explicit representation of individual colour values (e.g., red, green) is possible, but the spatial layout and the statistical properties of each coloured object of the scene are still accessible.

Finally, observers can select across a set of locations. If the selected locations all contain the same feature value (e.g. all are red), a feature label can be attached to each location, allowing for explicit access to the value at each location. If the selected locations are heterogeneous on one dimension, an overall description of the statistical features of the set of locations is still accessible (e.g. mean colour; Figure 2 displays various types of Boolean maps that can be created from a set of stimuli).

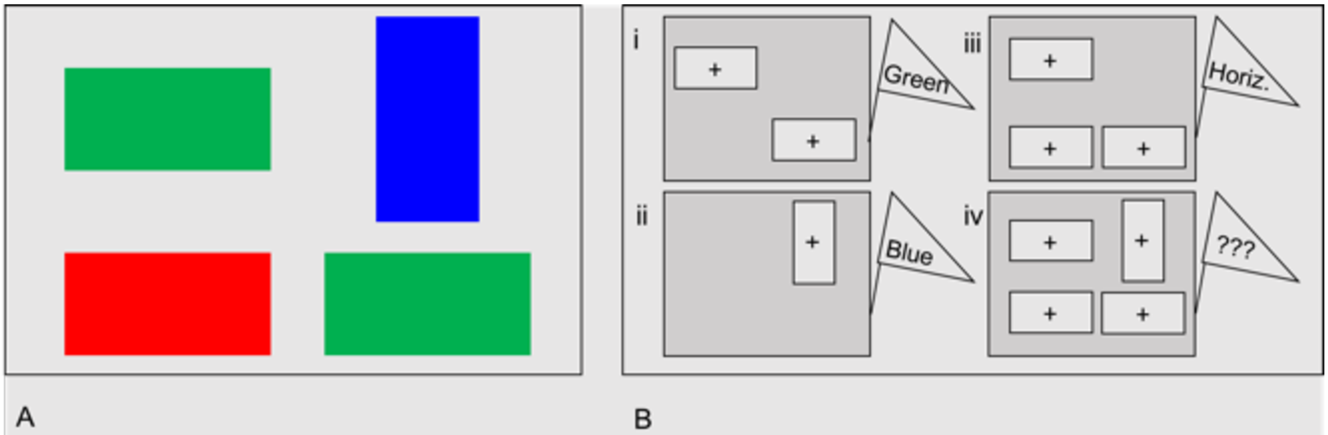

Figure 2. (A) Sample stimuli. (B) Four possible labeled Boolean maps of (A). (Bi) The colour green is selected and the locations of each green stimulus are mapped and given a label. (Bii) The colour blue is selected, mapped, and given a label. (Biii) Horizontally oriented stimuli are selected, mapped and given labels. (Biv) All four location values are selected, with the result that individual feature values are inaccessible, but an overall statistical description of a selected visual dimension is still accessible.

Selection

Observers can use the top-down process of selection to add additional visual dimensions to the labeled Boolean map through the Boolean operations of union or intersection. For example, observers can attend to all objects which are green *and* horizontal (through the process of intersection/conjunction) or all objects which are green *or* horizontal (through the process of union/disjunction). Importantly, the Boolean operations of intersection and union require computational steps that must be performed serially in order for multiple dimensions to be represented simultaneously. Thus, according to Boolean map theory, the difficulty in performing tasks which require representing two different dimensions (such as conjunction search tasks) comes from limitations of performing the Boolean operations associated with selection (i.e., performing the intersection operation), but are not limitations of access (i.e., representing multiple dimensions at once).

Selection across multiple dimensions is constrained by a bottleneck of the time taken to compute the relevant Boolean operations of intersection or union. Observers require a certain minimum amount of stimulus exposure time to be able to compute the relevant Boolean maps. While the theory does not explicitly state how long the computational processes of creating

Boolean maps requires, this duration may be possible to derive empirically. Huang and Pashler (2007) reported an experiment comparing performance in a feature search task (i.e, requiring selection across one dimension) versus a conjunction search task (i.e., requiring selection across two dimensions - colour and orientation; Experiment 5). Stimulus exposure durations were either 100, 200, or 400ms and results were measured in terms of accuracy in determining the location of a target (either the left or right of the screen). They observed that performance was superior for feature search under short exposure times (100 or 200ms), but was superior for conjunction search under long exposure times (400ms). The authors argued that this suggests that conjunction search is performed using a slower dimension-by-dimension subset search strategy.

The authors also point to other studies with stimulus exposure durations ranging from 28-378ms as being too fast to combine signals from two different dimensions (Moore & Egeth, 1998; Shih & Sperling, 1996). The precise time constraints on the ability to form representations across multiple dimensions still requires further study. However, the findings of Attarha and Moore (2015b) might suggest that forming a summary representation across two dimensions – specifically, orientation and size - requires approximately 230ms. However, forming of a summary representation across a single dimension (either orientation or size) only requires a shorter exposure duration of approximately 60 and 90ms, respectively.

Summary statistics and Boolean map theory

Interestingly, Boolean map theory makes no distinction between the features of single objects and those of sets of objects (Huang, 2015). The same attentional processes apply equally to both sets and singular objects. It is therefore possible to test the claims of Boolean map theory using a summary statistics paradigm, with the predictions being analogous to those regarding single objects. According to the theory, the locations of multiple objects can be selected for and encoded simultaneously. When multiple locations are selected, a process known as the *location-feature routine* takes the locations as inputs and returns a feature value associated with all of those locations. If the feature values of each location are homogeneous (e.g. they are all red), the value returned is a single feature value (e.g. red). If the feature values are heterogeneous (e.g. red and green), the value returned is a summary representation of the features (e.g. the mean colour; Huang & Pashler, 2007).

Huang (2015) examined the predictions of Boolean map theory with regard to set summary statistics. He used a pre/post-cueing design where participants were shown a combination of two of the following: a single coloured circle (red or green), a single white rectangle (horizontal or vertically oriented), a set of coloured circles (majority red or green), or a set of white rectangles (majority horizontal or vertical). The stimuli were shown in two different quadrants of the screen after either an uninformative spatial cue, or an informative spatial cue which designated which quadrant would contain the relevant information to be reported. After the stimuli were presented, the post-cue would indicate which quadrant the participant would have to report from. The task was to indicate the colour or orientation of the cued stimulus, be it the feature of the single object or the mean of the set. The dependent variable was the relative advantage in accuracy for the cued versus non-cued stimulus. Huang found that the pre-cueing advantage was the same regardless of whether the task involved reporting the properties of a single object or a set, consistent with Boolean map theory.

It is important to point out that Huang's (2015) task could be accomplished by forming a representation across only one dimension, either colour or orientation. The stimuli were presented for 50ms, so this short exposure duration would likely have interfered with the ability to form a labeled Boolean map with information from two dimensions. The current thesis is concerned with the processes of selection across multiple visual dimensions in sets of objects.

The current thesis

The predictions of Boolean map theory regarding the selection processes related to forming representations across multiple dimensions have not been thoroughly explored in the context of set summaries. It is possible to test the claims of Boolean map theory in this domain while also shedding light on the more general claims about the attentional processes involved in set summary statistics. The aim of the current thesis is to test the predictions of Boolean map theory regarding set summary statistics, particularly with regard to the ability to represent multiple visual dimensions simultaneously. It also aims to explain some of the previous (arguably discrepant) findings of the set summary statistics literature under the framework of Boolean map theory.

It is the contention of this thesis that the representation of set summary statistics is an attention-driven process, as stipulated under the framework of Boolean map theory. Limitations of selective attention are predicted to arise under the right sort of attentional constraints. Regarding representing multiple dimensions simultaneously, the relevant constraint would be stimulus exposure duration, and not set size. The findings of Attarha and Moore (2015b) which were interpreted as showing evidence of unlimited capacity processing of multiple dimensions were derived by using the variable of stimulus exposure duration as a means of keeping accuracy performance between floor and ceiling. Doing so may have obscured the potential finding of limitations of selective attention imposed by stimulus exposure duration. Limitations in forming set summary representations across multiple dimensions should be observed when stimulus exposure duration is kept at a fixed and difficult level. Decreases in accuracy and increases in response time (RT) are expected to occur as participants are required to attend to more visual dimensions simultaneously under fixed exposure durations.

To test this hypothesis, Experiment 1 (Chapter 2) sought to explore how observers form summary representations across one, two, or three dimensions of visual features simultaneously. Dimension values for size (large vs small) and orientation (vertical vs horizontal) were determined based on pilot testing, and full-colour RGB values for red and blue were used as the dimension values for colour. Experiment 2 (Chapter 3) repeats this experiment, but used an adaptive psychometric threshold procedure to equate performance (targeted at 85% accuracy) in the single feature condition across all three visual dimensions.

Experiment 1

The present experiment sought to explore how observers form summary representations across one, two, or three dimensions of visual features simultaneously. Boolean map theory would predict that observers can form summary representations across multiple dimensions, so long as there is sufficient stimulus presentation time to do so, and so long as observers do not have to form multiple representations of feature values within a single dimension. If stimulus presentation duration is held constant at a difficult level, performance should decrease as the number of dimensions to be attended to increases.

Attarha and Moore (2015b) showed participants a single set of circular sinusoidal gratings which varied in orientation and size. They asked participants to either report the mean orientation (‘left’ or ‘right’), mean size (‘large’ or ‘small’), or the mean of both dimensions (‘left and large’, ‘left and small’, ‘right and large’, or ‘right and small’). As a means of keeping performance between floor and ceiling, the experimenters used a coarse tracking procedure which varied the stimulus presentation duration across blocks. If participants were within 10% of floor or ceiling, the stimulus presentation duration was increased or decreased by 10ms. This resulted in unequal stimulus presentation durations across the orientation, size, and dual tasks, with average stimulus durations of 60, 90, and 230ms, respectively. The main independent variable in their study was whether the stimuli were presented simultaneously or sequentially. They found the same pattern of no advantage for sequential presentation across the three task types, which they interpreted as showing that forming multiple between-dimension summaries is an “unlimited capacity process” (p. 13), meaning that forming mean representations across a set is not impacted by the number of items to be processed simultaneously. The findings are also consistent with Boolean map theory, which would claim that all selectively attended objects are attended to simultaneously, regardless of set size, although the authors did not interpret their findings in light of this theory.

Attarha and Moore’s (2015b) experiment is notable in that the stimuli were the same across the three tasks, so any difference in performance could not be attributed to differences in the design of the stimuli. The tasks also required participants to form a single summary representation of the relevant dimension(s), as opposed to multiple within-dimension summaries, which may have been a limiting factor in Emmanouil and Treisman’s (2008) study. However, there are also a number of shortcomings of Attarha and Moore’s design which bear mentioning.

Their stimuli consisted of circular sinusoidal gratings that varied in orientation and size. Each of the gratings had four cycles, regardless of their size. This presents a potential confound in that the spatial frequency of the circles differed systematically with their size, with smaller circles having higher spatial frequency. It is therefore possible that observers were detecting differences in size, spatial frequency, or both when performing the ‘size’ task.

Additionally, the paradigm made use of two two-alternative forced-choice (2AFC) tasks and one four-alternative forced-choice (4AFC) task, with arcsine-corrected percentage correct as the sole measure of performance. Participants were instructed to answer as accurately as possible, with speed not being emphasized. The emphasis on accurate responding may have elicited a speed-accuracy trade-off, which may have shown a different pattern of results in RT data, if they had been reported. Further, while Attarha and Moore’s (2015b) extended simultaneous-sequential paradigm permits comparison of the *pattern* of results, the accuracies in each task are difficult to compare directly, given that 2AFC and 4AFC tasks have different guess rates (Macmillan & Creelman, 2005). However, a sensitivity analysis of their results reveals d' = 0.95, 1.05, and 0.99 for the simultaneous conditions of the orientation, size, and dual tasks, respectively. This would appear to validate the claim that the three tasks elicited roughly equal performance in terms of sensitivity.

Third, while Attarha and Moore (2015b) found evidence for unlimited-capacity (set-size independent) processing of the dimensions of size and orientation, there may be an upper limit to the attentional system’s ability to represent multiple dimensions simultaneously. Although not precisely specified in the theory itself, the authors of Boolean map theory suggest that creating labeled Boolean maps requires some kind of working memory, which is of limited capacity (Huang & Pashler, 2007). Therefore, evidence of limited-capacity processing may also be observed if the number of dimensions required to be attended to is increased beyond two.

Finally, while Attarha and Moore (2015b) found evidence of unlimited capacity processing of the orientation and size dimensions, an extension of this research into new dimensions may reveal important differences in ensemble encoding. There is some evidence of a hierarchy of summary statistics processing in that accuracy in computing the average properties of faces (such as emotion or identity) bear little correlation to computing average colour or orientation of triangles or Gabors (Haberman, Brady, & Alvarez, 2015). There is additional evidence that the form and colour of objects are processed in anatomically distinct regions of the ventral pathway of visual cortex (Rentzeperis, Nikolaev, Kiper, & van Leeuwen, 2014; Viviani & Aymoz, 2001; Zeki et al., 1991; Zeki, 1978). Therefore, it may be the case that the perception of colour could behave differently from the perception of mean orientation and size.

To address these issues, the stimuli of the current experiment consisted of rectangles which varied across colour hue, orientation, and size. The stimuli were presented to participants at a fixed duration of 200ms in all conditions. Participants were tasked with reporting the majority feature(s) across each of the possible combinations, depending on the block condition. The purpose of the experiment was to engage the processes of attention that facilitate summaries across one, two, or three dimensions and observe differences in performance across the relevant tasks. Observing dual or triple-task interference could help clarify visual attention's ability to form summary representations across multiple visual dimensions. On the basis of Boolean map theory, it is predicted that performance should decrease as the number of dimensions attended to increases, due to constraints associated with the attentional processes of selection.

Method

All aspects of the method were approved by the Concordia University Human Research Ethics Committee (HREC# 10000119), in compliance with Canadian Tri-council policy on ethical treatment of human participants. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2012).

Participants

An a-priori power analysis was conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) based on a pilot test of the experiment with four subjects. The analysis indicated that at least three observers were needed to achieve at least 90% power for our anticipated repeated-measures ANOVA, using an eta square of .9 and an epsilon of .52. The target number was increased to 10 in case the effect was less than expected and 11 participants were collected in total. Participants were recruited online through Concordia University's Psychology participant pool system, with participants receiving bonus class credit in exchange for participation. Demographic information including age (M = 21.64, SD = 3.11) and sex (1 male) was collected after participants gave their informed consent. Participants were then screened for visual acuity using the Freiburg Visual Acuity Test (Bach, 2007) using the apparatus described below. All participants had normal vision (logMar range: -0.18 to 0.14). Participants were also screened for normal colour vision using the HRR Pseudoisochromatic Plates (Cole, Lian, & Lakkis, 2006). Participants were then instructed about the task they were to complete with the aid of printed visual graphics with stimuli from the experiment. No participants were excluded from analysis.

Apparatus

Stimuli were presented on a 50.50cm cathode ray tube monitor (ViewSonic G220fb, 1024x768 pixel resolution, refresh rate of 100Hz) controlled by an iMac computer (OS X 10.9.5). with an NVIDIA GeForce GT 750M 1024 MG graphics card. Stimuli were created using the Psychophysics Toolbox Version 3.0.11 (Brainard, 1997a; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) for MATLAB (version 2014a, The Mathworks, Natick, MA). Participants sat in a height-adjustable chair at a table with a chin rest placed at a viewing distance of 180cm. Participants used the chin rest for visual acuity screening, but were allowed to sit as they were most comfortable during Experiment 1; measurements of visual angle for stimuli are therefore approximations based on an assumed viewing distance of 180cm. The room was well-lit during the experiment so that participants could view the response keys.

Stimuli. Stimuli consisted of a set of 25 equidistant rectangles arranged in a 5x5 grid centered on the screen against a black background. The stimuli covered a total of 6.36°. The rectangles varied in colour hue (blue or red), orientation (horizontal or vertical), and size (big or small). In each trial, 19 out of the 25 (76%) rectangles were randomly assigned to have the same feature within a dimension (e.g. 19 blue, 19 small, and 19 horizontal rectangles). Small rectangles were 0.16 x 0.47°, large rectangles were 0.31 x 0.62°. Horizontal/vertical rectangles were oriented at either 0 or 90°. Full colour hue values were used for red and blue (i.e. red = 100%, RGB = [1.0, 0.0, 0.0]; or blue = 100%, RGB = [0.0, 0.0, 1.0]), with no modifications to saturation.

A Mondrian mask, generated by custom code in MATLAB, consisting of a set of overlapping rectangles of all colours that filled the entire screen (15.97°) was presented after each stimulus (Hesselmann, Hebart, & Malach, 2011).

Procedure.

There were seven blocked conditions in which every participant took part. Each experimental block was preceded by a practice block in which participants received feedback in the form of high/low tones after correct/incorrect responses. Experimental blocks did not include feedback. Each block was designed to test participants' ability to attend to and encode one or more featural dimensions present in the set of rectangles. Participants were informed about which dimensions they would be queried at the start of each block through an instruction screen.

On each trial, participants would be presented with a fixation dot for 1000ms, followed immediately by the stimulus for 200ms, then the mask for 600ms. After presentation of the mask, the participant was queried about the majority colour, orientation, or size of the preceding stimulus using a 2-alternative forced choice. The two alternatives were presented 0.80° left- or right-of-centre of the screen, in the form of rectangles that contained only the relevant feature being queried. For example, colour was queried with a blue square on the left and a red square on the right. Orientation and size used white rectangles (see Figure 3).

There were three single-task conditions in which participants would be queried about the majority feature of one dimension (either colour, orientation, or size) at the end of each trial for the entire duration of the block. They would thus have to attend to and encode only that dimension. For example, in the colour condition, participants would be asked to report the colour of the majority of the rectangles for the entire duration of the block.

In dual-task conditions, participants would be queried about the majority feature of one of two possible dimensions (either colour/orientation, colour/size, or orientation/size) at the end of each trial for the duration of the block. For example, in the colour and orientation condition, participants could be asked to report either the majority colour or the majority orientation of the set of rectangles at the end of a trial, but would not be asked to report both. The order in which features were queried was shuffled randomly so that participants could not anticipate which feature to attend to during a trial.

In the triple-task condition, participants would be queried about the majority feature of one of three possible dimensions (colour/orientation/size) at the end of each trial, in a manner similar to the dual-task conditions. The experiment therefore took a form akin to a pre/post-cueing paradigm, where the cue was provided at the start of each block. In single-task blocks participants could direct their attention to the relevant stimulus dimension without any uncertainty, whereas in multi-task blocks, uncertainty about task requirements during a trial would force participants to attend to more than one dimension.

Each experimental block consisted of 12 trials per stimulus-response condition (i.e., majority blue/red, horizontal/vertical, big/small). In single-dimension blocks this equated to 96 trials. In two-dimension blocks, there were double the amount of trials (192), and triple for the three-dimension block (288 trials). Practice blocks had a third of the trials of experimental blocks (i.e. 32, 64, 96 trials, respectively). Participants could take breaks between blocks. Each participant completed all seven block conditions in random order. The entire experiment lasted approximately 75 minutes.

Data collected from the participant included, exhaustively: age, sex, response (left/right), trial result (correct/incorrect), and response time (milliseconds between onset of question and key response). Information about the block type, stimulus properties such as the majority colour, size, and orientation presented, and test condition (whether the test was for colour, orientation, or size) were also recorded for each trial.

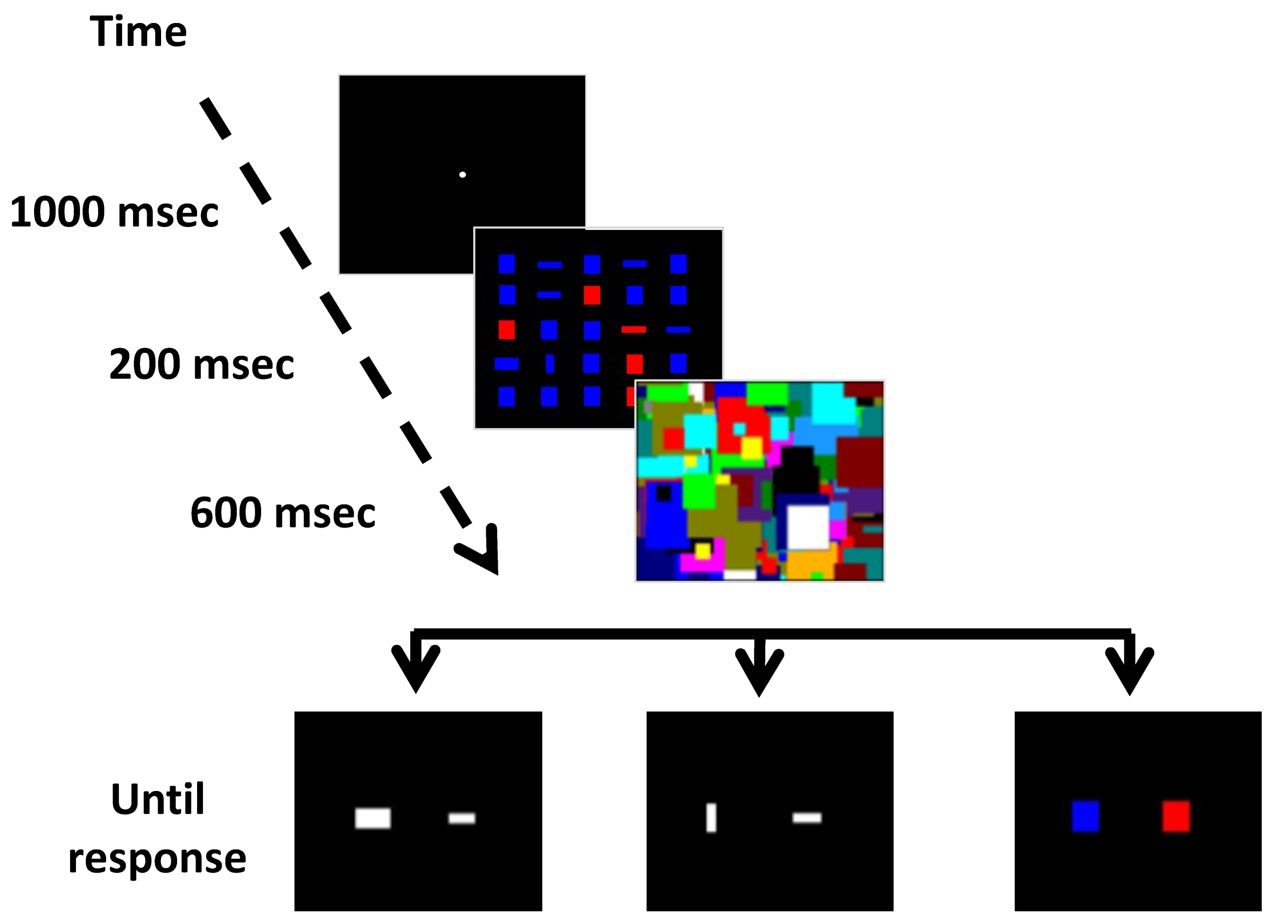

Figure 3. Trial events for Experiment 1. Stimuli sizes are increased for illustration purposes.

Data Analysis

A few definitions will be helpful in understanding the results. There were three main *tasks* of the experiment, namely, report the majority colour, majority orientation, or majority size of the stimuli. Each participant engaged in seven blocked *conditions* consisting of one or more tasks. There were three single-task conditions, three dual-task conditions, and one triple-task condition. The *number of tasks* in each condition corresponds to the *number of dimensions* required to attend to in order to complete the condition successfully.

The analysis was aimed at testing the question of whether attending to more featural dimensions has a negative impact on performance. Performance was measured in terms of accuracy and RT. The approach of analysis went from coarse to fine-grained. First, the average performance of the three one-dimension tasks was compared to the average performance of the three two-dimensional tasks and the one three-dimensional task. Second, performance at the level of the seven blocked conditions of the experiment was descriptively explored. Third, performance on each task type within the seven block conditions (e.g., the average performance on colour task trials within each condition that contained the colour task) was explored. The purpose of the third analysis was partly descriptive, but also allowed for comparing differences in performance within dual and triple task conditions to the baseline single-task condition. Each analysis was conducted using R version 3.3.0 (R Core Team, 2013; RStudio Team, 2015) and JASP version 0.8.1. (JASP Team, 2017; Morey & Rouder, 2015; Rouder, Morey, Speckman, & Province, 2012; Rouder, Speckman, Sun, Morey, & Iverson, 2009).

The first set of analyses were conducted using standard one-way repeated measures ANOVA, with post-hoc dependent sample *t*-tests applied to each pairwise comparison, using Bonferroni correction where appropriate. The relevant effect sizes in terms of η^2 for ANOVAs, and bias-corrected dependent-samples Hedges' *g* (*g**) for each pairwise comparison are also reported (Kline, 2004). Additionally, 95% confidence intervals (CI) around the mean of each condition are reported for the descriptive analyses. For all analyses, between-subjects variance was removed using the method proposed by Loftus and Masson (1994); namely, subtracting the difference between the subjects' mean and the grand mean from each of the participants' condition scores. This was done so that the variance reflects the effect of a condition on subjects' scores, and not individual differences between subjects. It should be noted that this correction does not affect the results of significance testing or Bayes Factors, but it does narrow the CIs due to reducing the variance which therefore also increases the effect size.

Bayesian analyses were conducted so as to avoid many of the pitfalls associated with traditional null hypothesis significance test (NHST) analyses (Dienes, 2011; Kline, 2004; Wetzels et al., 2011). Furthermore, Bayesian analyses are capable of determining whether there is evidence in favour of the null hypothesis, as opposed to there being insufficient evidence to support either hypothesis (Wetzels et al., 2011). Following the recommendations of Kruschke (2016) and Wetzels et al. (2011) the results of these experiments are reported using both NHST and Bayesian techniques, since no established standards yet exist for reporting the results of Bayesian analyses. Readers are invited to interpret the statistics of their choice, be it NHST or Bayesian.

For two-factor comparisons, the analysis includes estimates of the means with the corresponding 95% Bayesian credibility interval (BCI) around the means. The 95% BCI represents the range of values which have a 95% probability of containing the true mean (Dienes, 2008). Bayes Factors (*BF*s) for each multi-factor comparison and two-factor comparison are also reported along with an estimate of the proportional error on the *BF*. Finally a Bayesian estimate of effect size (median δ) with its 95% BCI is reported for each two-factor comparison.

Bayes Factors represent the relative posterior probability of one hypothesis over another. The term BF_{10} represents the relative probability of the research hypothesis over the null hypothesis, where the null hypothesis is non-directional. Conversely, the term BF_{01} represents the relative probability of the null hypothesis over the research hypothesis when the null hypothesis is non-directional. BF_{+0} represents the relative probability of the research hypothesis when the research hypothesis predicts differences greater than zero; a BF_{-0} represents the probability of the opposite alternative prediction (Wetzels et al., 2011).

A BF_{10} equal to 20 would indicate that the alternative hypothesis is 20 times more likely to explain the data than the null hypothesis. A BF_{10} greater than 10 would be considered strong evidence in favour of the alternative hypothesis. Those between 10 and 3 are considered substantial (or moderate) evidence, and those between 0 and 3 are considered anecdotal evidence in favour of the alternative hypothesis (Wetzels et al., 2011).

The *BF*s for two-condition contrasts were calculated with JASP using the default Cauchy prior with an r scale effect size of 0.707. The robustness of each test was checked with prior scale of 1.0 and 1.5 to ensure that the selection of prior distributions did not bias the results (Rouder et al., 2009). Bayesian ANOVAs were conducted using an r scale fixed effects of 0.5, r scale random effects of 1.0 and r scale covariates of 0.354. Models used the default of 10,000 samples for Monte Carlo simulation (Kruschke, 2016; Rouder et al., 2012).

Results

Accuracy - Primary analysis

The experiment was aimed at testing the hypothesis that attending to more visual dimensions decreases performance in terms of lower accuracy and higher RT. The experiment required participants to engage in one or more tasks within a condition, requiring attention towards one, two, or three of the visual dimensions present in the stimuli. The primary analysis examined mean differences in accuracy across the mean of the three single-task conditions, the mean of the three dual-task conditions and the single triple-task condition. For this purpose, a 1 x 3 within-subjects ANOVA was performed. The assumption of normality was tested using the Shapiro-Wilkes test (Shapiro & Wilk, 1965) and revealed no violation, $W = .98$, $p = .791$. Likewise, the assumption of sphericity was tested with Mauchly's method (Mauchly, 1940), and revealed no violation, W = .62, $p = .118$. No scores were excluded from the analysis and no transformations to the data were made. The ANOVA revealed that the conditions were statistically different, $F(2,20) = 26.29$, $p < .001$, $\eta^2 = .72$, with a tendency towards lower accuracy as more tasks were present in the condition (Table 1, Figure 4). Follow-up Bonferroni-corrected post-hoc *t*-tests (Dunn, 1961) assuming that accuracy would be greater in conditions with fewer tasks (Table 2) revealed that mean accuracy in the single-task conditions was higher than mean accuracy in either the dual-task conditions, or the triple-task condition. However, mean accuracy in the dual-task conditions was not found to be statistically higher than accuracy in the triple-task condition. This analysis would suggest that overall accuracy declined as participants were required to attend to more visual dimensions, but the effect of attending to three versus two dimensions may not have been as strong as attending to one versus two dimensions.

Table 1.

Descriptive Statistics of Accuracy Across Number of Tasks Within a Condition

Number of Tasks	*M*	*SD*	95% CI	95% BCI
1 Task	93.36	2.79	[91.49, 95.24]	[91.69, 95.04]
2 Tasks	86.64	1.53	[85.61, 87.66]	[85.72, 87.56]
3 Tasks	84.36	2.84	[82.46, 86.27]	[82.66, 86.07]

Note. $N = 11$ in all cases. CI = Confidence interval (NHST), BCI = Bayesian credibility interval. Confidence intervals are constructed using within-subjects corrected variance and are not corrected for multiple comparisons (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

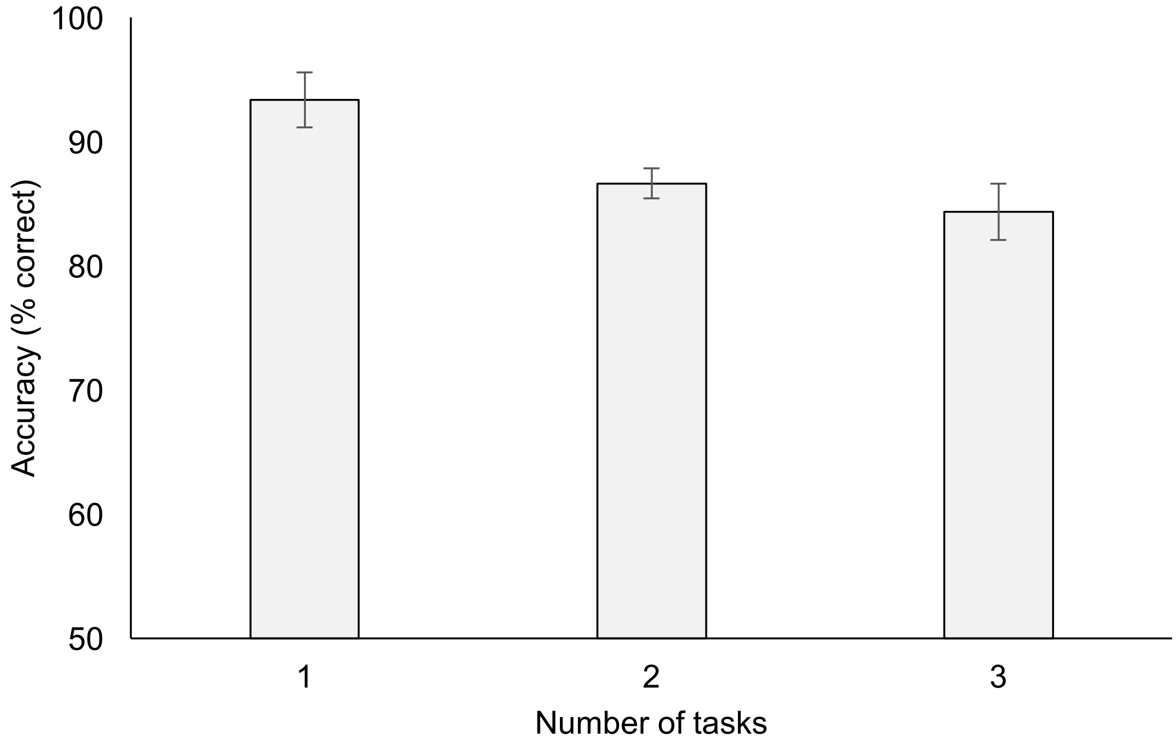

Figure 4. Mean accuracy across the number of tasks within a condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

Table 2.

Pairwise Comparisons of Accuracy Across Number of Tasks Within a Condition

Contrast	t^{a}	*df*	p_{bonf} [b]	W^{c}	*p*	M_{diff}	SE_{diff}	r_{ab}	*95%* CI[d]	*g**	*95%* CI[e]
1 Task – 2 Tasks	6.38	10	< .001	.92	.325	6.73	1.05	-.24	[4.37, 9.08]	2.74	[2.41, 4.40][f]
1 Task – 3 Tasks	5.51	10	< .001	.94	.514	9.00	1.64	-.85	[5.36, 12.64]	2.93	[1.21, 4.65]
2 Tasks – 3 Tasks	2.09	10	.095	.89	.135	2.27	1.09	-.30	[-0.15, 4.70]	0.91	[1.49, 0.80][f]

[a] These are one-tailed Student's *t*-tests and assume that more dimensions lead to lower accuracy.

[b] *p*-values are corrected for three comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 24.42% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of *g** takes into account the correlation between the dependent samples, r_{ab}.

[f] Difference between *SD*s was significant ($p < .05$); Lower and upper boundaries of 95% CI represent effect size (Glass' Δ) based on treatment group (more tasks) *SD*, and control group (fewer tasks) *SD*, respectively.

A Bayesian ANOVA revealed a BF_{10} of 2,327,000, Error % = 0.62, indicating that the alternative hypothesis is approximately 2.33 million times more likely to explain these data than the null hypothesis. Follow-up Bayesian paired-samples *t*-tests assuming the hypothesis that accuracy would be greater in conditions with fewer tasks found strong evidence that single-task conditions resulted in higher accuracy scores in comparison to dual- and triple-task conditions (BF_{+0} = 696.79, 253.54, respectively). However, the analysis only found anecdotal evidence (BF_{+0} = 2.74) in favour of the hypothesis that dual-task conditions resulted in higher accuracy than the triple-task condition (see Table 3). The results of the Bayesian analysis are consistent with the NHST analysis, in that there is strong evidence that participants were less accurate as they were required to attend to more dimensions, but the effect appears to be uneven between attending to one dimension versus two and two dimensions versus three.

Table 3.

Bayesian Pairwise Comparisons of Accuracy Across Number of Tasks Within a Condition

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{+0}[a]	Error %	Median	95% BCI
1 Task – 2 Tasks	696.79	1.36×10^{-8}	1.70	[0.74, 2.77]
1 Task – 3 Tasks	253.54	1.97×10^{-7}	1.45	[0.56, 2.44]
2 Tasks – 3 Tasks	2.74	3.29×10^{-5}	0.53	[0.07, 1.16]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be greater than zero.

Accuracy - Seven condition analysis

An inspection of the results across the seven blocked condition types suggests that performance in the colour condition was near ceiling (Table 4, Figure 5). Indeed, 3/11 (27%) participants had 100% accuracy in the colour condition and 5/11 (45%) had 99% accuracy. This may explain the results of the dimension-level analysis that suggested that there was a relatively small decrease from two dimensions to three. It appears that the colour task was likely too easy to be attentionally demanding, such that the colour task had a relatively small impact on performance when it was combined with the other tasks in the dual- or triple-task conditions. It can also be observed that performance in the dual tasks that involved colour were higher, in comparison to the orientation and size dual task condition. Performance in the orientation and size dual task was near that of the triple task, again potentially indicating that inclusion of colour in the triple task did not impact performance in that condition.

However, additional analysis of the accuracy may reveal that performance in each task was still reduced by the influence of the other tasks. Therefore, the final analysis conducted here examines performance on each task *within* each blocked condition. For example, the analysis examines the average performance on colour trials for the colour only condition; the colour and size condition; the colour and orientation condition; and the colour, orientation, and size condition. Differences from the baseline condition of colour only could elucidate whether each additional task reduced performance on the primary task of interest.

Table 4.

Descriptive Statistics of Accuracy Across Seven Condition Types

Condition	*M*	*SD*	95% CI[a]	95% BCI
C	98.64	4.90	[95.34, 101.93]	[95.69, 101.58]
O	90.73	2.82	[88.83, 92.62]	[89.03, 92.42]
S	91.27	5.23	[87.76, 94.79]	[88.13, 94.41]
CO	88.00	6.36	[83.73, 92.27]	[84.18, 91.82]
CS	87.45	4.11	[84.69, 90.22]	[84.98, 89.92]
OS	84.91	4.36	[81.98, 87.84]	[82.29, 87.53]
COS	84.36	3.70	[81.88, 86.85]	[82.14, 86.58]

Note. $N = 11$ for all conditions. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

[a] Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) and are not adjusted for multiple comparisons.

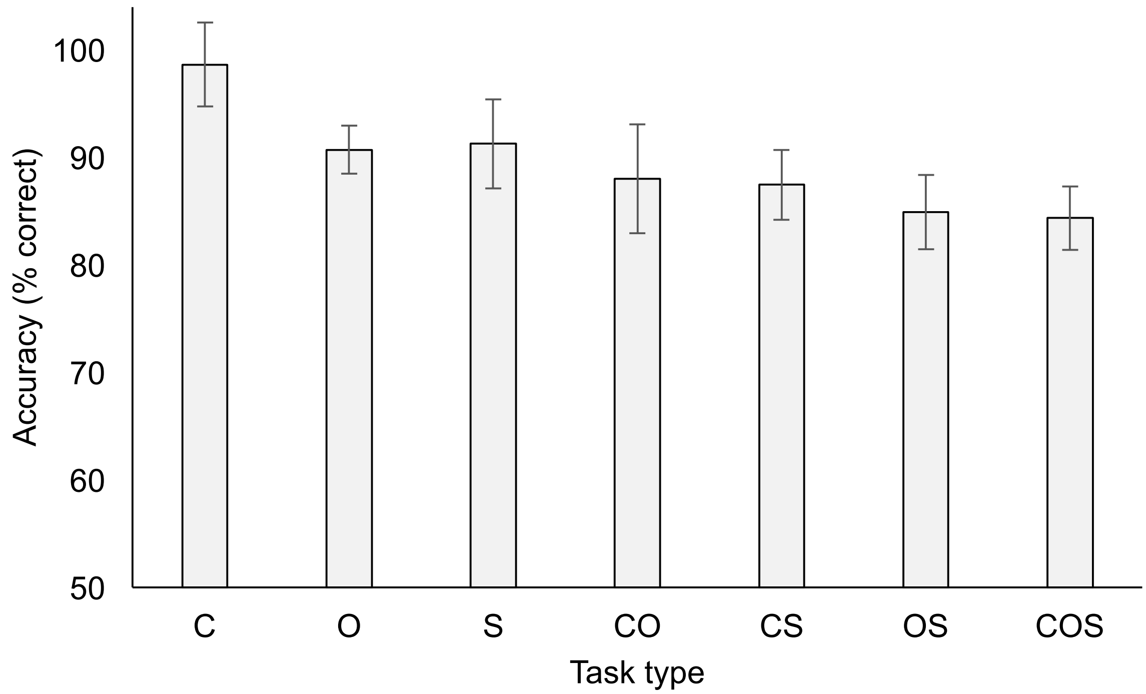

Figure 5. Accuracy across seven condition types.

Note. Error bars represent 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) and are not corrected for multiple comparisons. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Accuracy - Task within condition analysis

The final tertiary analysis was conducted to examine the effect of other tasks in a condition compared to each task's respective baseline. It is composed of two parts. First, the descriptive statistics of performance on each task within each condition are examined. Second, the difference scores of task performance within a multi-task condition minus their single-task baseline scores are provided, along with the relevant dependent-samples *t*-tests, effect sizes, and Bayes Factors. For these analyses, between-subjects variance was removed using the task type mean as the grand mean (e.g., the grand mean for the colour task was calculated as the mean performance on the colour task for each condition that contained it).

It can be observed from the descriptive statistics of each task type within each condition (Table 5; Figures 6, 7, and 8) that performance in a task tends to decrease if more tasks are present within a condition. As an additional means of analysis, the difference between task performance in multi-task conditions and performance in the baseline single-task condition was examined in order to determine if the presence of other tasks within a condition affected performance on the task of interest. For example, colour task performance within the colour and orientation condition was subtracted from the baseline single-task performance for colour, as was colour task performance in the colour and size, and colour, orientation, and size conditions. Therefore, three difference scores were computed for each task type (i.e. colour, orientation, size), yielding nine in total. If performance in each multi-task condition was worse than its respective baseline, this indicates that the presence of other tasks was interfering with the ability to attend to / perceive the visual dimension in question. Furthermore, the relative magnitude of difference from baseline could be used as an indication of performance decrement (in terms of raw difference scores, effect sizes, or Bayes factors). It was hypothesized that performance decrements would be greatest in the triple-task conditions relative to the dual-task conditions.

Table 5.

Descriptive Statistics of Task Accuracy Within Each Condition

Task Within Condition	*M*	*SD*	95% CI	95% BCI
C	98.58	2.41	[96.96, 100.20]	[97.13, 100.03]
C W/IN CO	95.27	1.70	[94.13, 96.40]	[94.25, 96.28]
C W/IN CS	94.32	2.37	[92.73, 95.91]	[92.89, 95.74]
C W/IN COS	93.66	1.53	[92.63, 94.68]	[92.74, 94.57]
O	90.53	5.04	[87.15, 93.91]	[87.50, 93.56]
O W/IN CO	80.49	9.82	[73.90, 87.09]	[74.59, 86.39]
O W/IN OS	85.23	4.71	[82.07, 88.39]	[82.40, 88.06]
O W/IN COS	80.30	5.45	[76.64, 83.96]	[77.03, 83.58]
S	91.10	5.58	[87.35, 94.85]	[87.74, 94.45]
S W/IN CS	80.59	5.86	[76.65, 84.53]	[77.06, 84.11]
S W/IN OS	84.38	4.27	[81.51, 87.24]	[81.81, 86.94]
S W/IN COS	79.36	5.22	[75.85, 82.87]	[76.22, 82.50]

Note. $N = 11$ in all conditions. Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of each condition type (colour, orientation, or size), and are not corrected for multiple comparisons. If correcting for nine comparisons, confidence intervals would be 49.57% larger. W/IN = within, C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

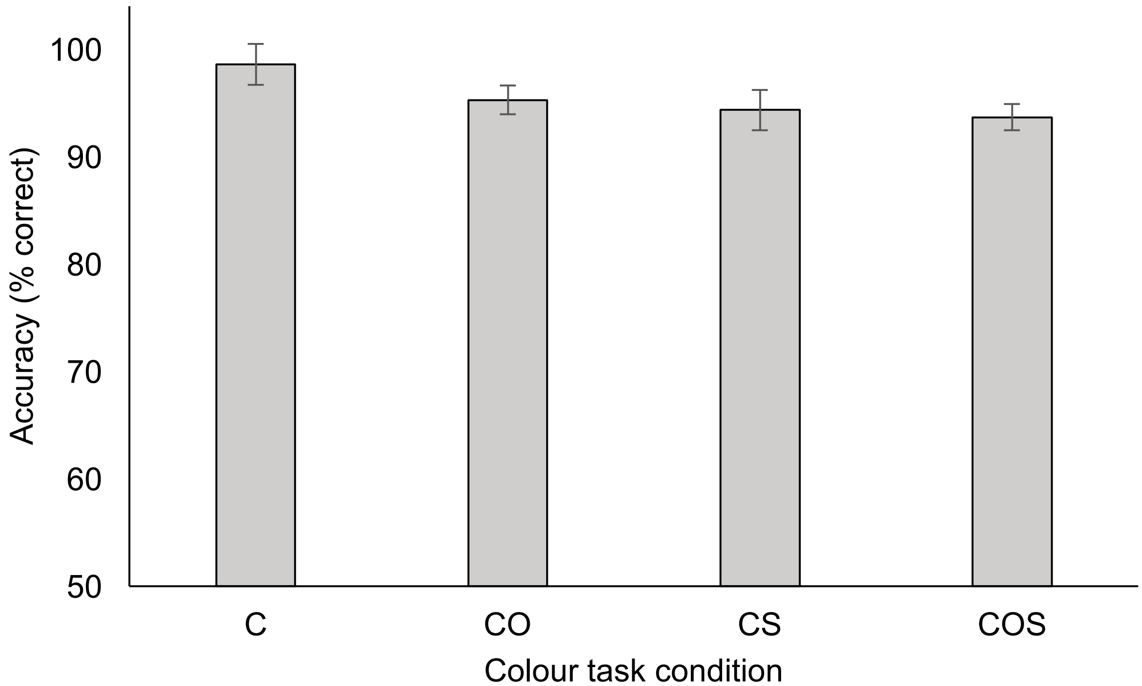

Figure 6. Accuracy on the colour task within each colour-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of colour task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. C = colour, CO = colour and orientation, CS = colour and size, COS = colour, orientation, and size.

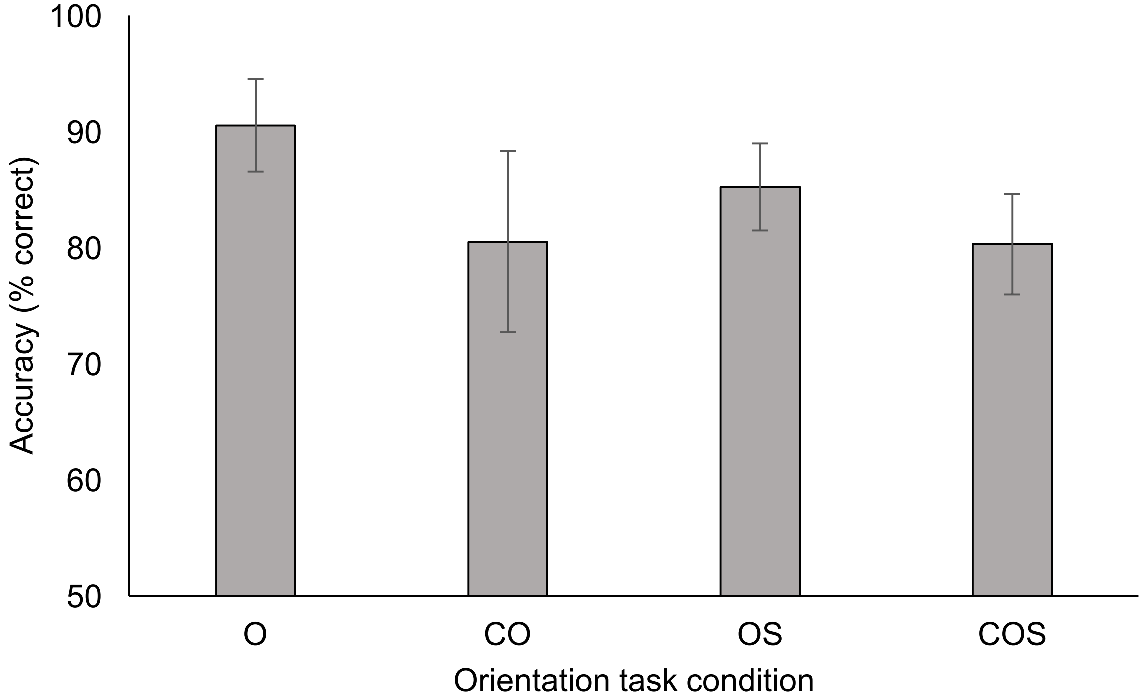

Figure 7. Accuracy on the orientation task within each orientation-containing condition.
Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of orientation task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. O = orientation, CO = colour and orientation, OS = orientation and size, COS = colour, orientation, and size.

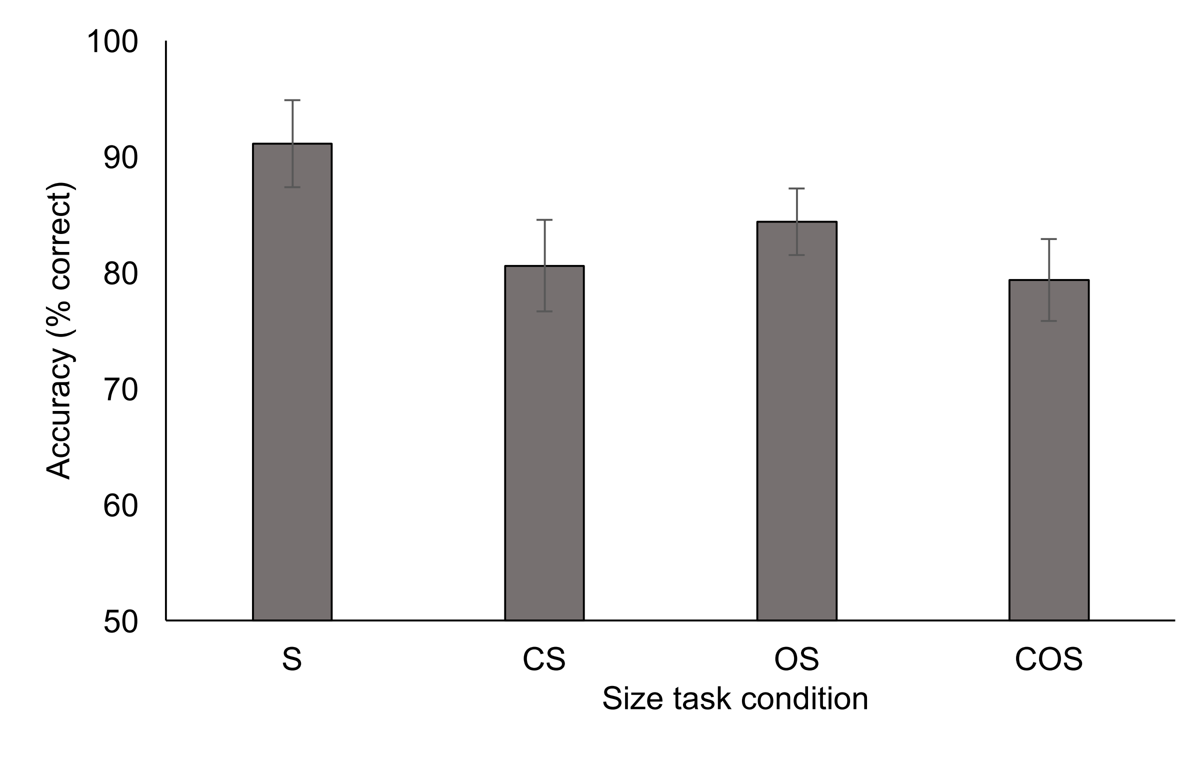

Figure 8. Accuracy on the size task within each size-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of size task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. S = size, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Examination of the differences from baseline (Tables 6, 7) suggest that task accuracy in most of the multi-task conditions was significantly lower compared to baseline, even after Bonferroni correction. The notable exception appears to be the orientation task, where only the triple-task condition was statistically significantly lower than baseline, or had a $BF_{-0} > 10$, indicating strong evidence in favour of the alternative hypothesis. It can also be observed that there is a general tendency towards larger mean differences when there are three tasks compared to two (both in terms of raw scores and effects sizes), although the imprecision around these observations represented by the confidence intervals makes it unclear if the effects would be reliable. A higher-powered study would likely be able to achieve more precise estimates of these effects. However, these findings should also be taken with caution, given that the colour task was at ceiling and likely behaved differently from the other two task types. In sum, it appears as though the main hypothesis of the experiment was confirmed by this analysis, in that attending to more than one dimension lead to lower accuracy scores, although perhaps not in a straightforward linear fashion.

Table 6.

Accuracy Difference from Baseline for Multi-Task Conditions

Contrast	t^a	df	p_{bonf}^b	W^c	p	M_{diff}	SE_{diff}	r_{ab}	95% CI[d]	$g*$	95% CI[e]
C W/IN CO - C	-3.34	10	.034	.92	.286	-3.31	0.99	-.26	[-5.52, -1.11]	-1.46	[-2.60, -0.32]
C W/IN CS - C	-3.52	10	.025	.93	.365	-4.26	1.21	-.42	[-6.96, -1.56]	-1.64	[-2.86, -0.41]
C W/IN COS - C	-4.56	10	.005	.94	.459	-4.92	1.08	-.64	[-7.33, -2.52]	-2.24	[-3.69, -0.79]
O W/IN CO - O	-2.46	10	.151	.63	< .001	-10.04	4.08	-.62	[-19.13, -0.95]	-1.18	[-1.99, -1.02][f]
O W/IN OS - O	-2.84	10	.079	.96	.740	-5.30	1.87	.20	[-9.46, -1.15]	-1.00	[-1.87, -0.13]
O W/IN COS - O	-4.63	10	.004	.94	.529	-10.23	2.21	.02	[-15.15, -5.30]	-1.79	[-2.93, -0.65]
S W/IN CS - S	-3.76	10	.017	.95	.585	-10.51	2.80	-.31	[-16.74, -4.28]	-1.68	[-2.89, -0.47]
S W/IN OS - S	-3.43	10	.029	.95	.688	-6.72	1.96	.15	[-11.09, -2.36]	-1.24	[-2.18, -0.30]
S W/IN COS - S	-3.76	10	.017	.89	.155	-11.74	3.13	-.84	[-18.71, -4.78]	-1.99	[-3.42, -0.56]

[a] These are one-tailed Student's *t*-tests and assume that accuracy would be lower in multi-task conditions.

[b] *p*-values are corrected for nine comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 49.57% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of $g*$ takes into account the correlation between the dependent samples, r_{ab}.

[f] Difference between SDs was significant ($p < .05$); lower and upper limits of 95% CI represent effect size (Glass' Δ) based on treatment group SD, and control group SD, respectively.

Table 7.

Bayesian Analysis of Accuracy Difference from Baseline for Multi-Task Conditions

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{-0}[a]	Error %	Median	95% BCI
C W/IN CO - C	15.26	1.26×10^{-6}	-0.85	[-1.60, -0.20]
C W/IN CS - C	19.36	8.50×10^{-7}	-0.89	[-1.67, -0.22]
C W/IN COS - C	78.19	3.25×10^{-7}	-1.19	[-2.07, -0.39]
O W/IN CO - O	4.52	1.45×10^{-5}	-0.62	[-1.29, -0.10]
O W/IN OS - O	7.63	1.00×10^{-7}	-0.72	[-1.44, -0.13]
O W/IN COS - O	85.00	4.49×10^{-7}	-1.22	[-2.09, -0.41]
S W/IN CS - S	26.96	5.98×10^{-7}	-0.96	[-1.76, -0.25]
S W/IN OS - S	17.28	1.03×10^{-6}	-0.87	[-1.65, -0.22]
S W/IN COS - S	26.91	5.99×10^{-7}	-0.97	[-1.75, -0.27]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors are directional and assume the difference would be less than zero.

Response time - Primary analysis

Response time was analyzed in a similar manner to accuracy data. A 1 x 3 repeated measures ANOVA was conducted across the mean RT of the three single-task conditions, the mean of the three dual-task conditions, and the triple-task condition. A Shapiro-Wilkes test revealed no violation of the assumption of normality ($W = .97$, $p = .552$), nor did Mauchly's test of sphericity reveal any violation, $W = .622$, $p = .118$. The analysis revealed that the three conditions differed from each other, $F(2,20) = 26.46$, $p < .001$, $\eta^2 = 0.73$, revealing a tendency of greater RT as more tasks were present in the conditions (Table 8, Figure 9). Follow-up Bonferroni-corrected paired-samples *t*-tests assuming greater RT as more tasks were present in the conditions (Table 9) revealed that RT in the single-task conditions was lower than RT in either the dual-task conditions, or the triple-task condition. However, RT in the dual-task conditions was not found to be statistically greater than RT in the three-dimension task. This analysis would suggest that participants were generally longer to respond when they were required to attend to more dimensions, but the effect of attending to three versus two dimensions may not have been as strong as attending to one versus two dimensions.

A Bayesian repeated-measures ANOVA revealed a $BF_{10} = 879{,}200{,}000$, Error % = 0.63, indicating that the alternative hypothesis is approximately 879 million times more likely to explain these data than the null hypothesis. Follow-up Bayesian paired-samples *t*-tests for each two-level contrast (Table 10) revealed strong evidence in favour of the hypothesis that single-task conditions had lower RT than either the dual-task conditions, or the triple-task condition, but only found anecdotal evidence in favour of the hypothesis that RT was greater in the triple-task condition compared to the dual-task conditions. The Bayesian analysis thus found a similar pattern to the NHST results.

The analysis revealed that single-task conditions produced reliably lower RT than dual- or triple-task conditions. The analysis did not find credible evidence that dual-task conditions had greater RT than the triple-task condition. This is a similar pattern to those found in the accuracy data, and may again be due to the relatively low difficulty of the colour task not impacting attention enough to elicit differences in RT when combined with additional tasks.

Table 8.

Descriptive Statistics of Response Time Across Number of Tasks Within a Condition

Number of Tasks	*M*	*SD*	95% CI	95% BCI
1 Task	395.50	72.55	[346.70, 444.20]	[351.90, 439.10]
2 Tasks	683.40	73.04	[634.30, 732.40]	[639.50, 727.30]
3 Tasks	767.30	88.01	[708.10, 826.40]	[714.40, 820.10]

Note. $N = 11$ for all conditions. CI = Confidence interval (NHST), BCI = Bayesian credibility interval. Confidence intervals are constructed using within-subjects corrected variance and are not corrected for multiple comparisons (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

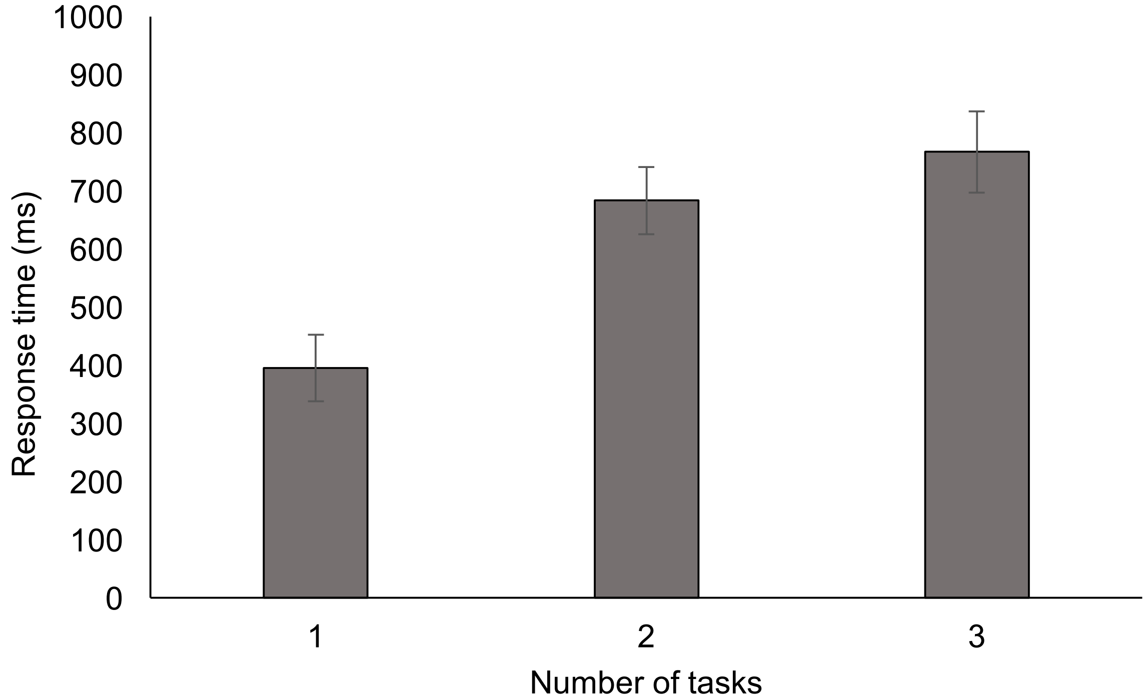

Figure 9. Mean response time across the number of tasks within a condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

Table 9.

Pairwise Comparisons of Response Time Across Number of Tasks Within a Condition

Contrast	t^a	df	$p_{bonf}{}^b$	W^c	p	M_{diff}	SE_{diff}	r_{ab}	95% CI^d	g^*	95% CI^e
1 Task – 2 Tasks	-8.23	10	< .001	.93	.405	-287.88	34.97	-.27	[-365.80, -209.97]	-3.63	[-4.26, -1.22]
1 Task – 3 Tasks	-8.57	10	< .001	.89	.123	-371.78	43.36	-.60	[-468.40, -275.16]	-4.23	[-5.47, -1.78]
2 Tasks – 3 Tasks	-1.93	10	.125	.93	.388	-83.90	43.59	-.61	[-181.00, 13.22]	-0.95	[-2.09, 0.19]

[a] These are one-tailed Student's *t*-tests and assume that more dimensions lead to higher response time.

[b] *p*-values are corrected for three comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 24.42% larger.

[e]. The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of g^* takes into account the correlation between the dependent samples, r_{ab}.

Table 10.

Bayesian Analysis of Response Time Across Number of Tasks Within a Condition

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{-0}	Error %	Median	95% BCI
1 Task – 2 Tasks	4587.33	8.61×10^{-10}	-2.23	[-3.56, -1.06]
1 Task – 3 Tasks	6285.83	7.37×10^{-10}	-2.30	[-3.62, -1.11]
2 Tasks – 3 Tasks	2.20	3.19×10^{-5}	-0.49	[-1.10, -0.06]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be less than zero.

Response time - Seven condition analysis

A descriptive analysis of RT in each of the seven conditions (Table 11, Figure 10) suggested a similar finding to the above, namely that RT tended to increase as the number of tasks increased. It is also worth highlighting that performance in the orientation and size condition was close to that of the colour, orientation and size condition, which were both higher than the dual-task conditions which involved colour. This again, may be owing to the colour task being too easy to impact performance negatively when it is combined with other tasks within a condition.

Table 11.

Descriptive Statistics of Response Time Across Seven Condition Types

Condition	*M*	*SD*	95% CI[a]	95% BCI
C	352.30	124.15	[268.90, 435.70]	[277.70, 426.90]
O	427.00	88.13	[367.80, 486.20]	[374.10, 480.00]
S	407.10	129.24	[320.30, 493.90]	[329.50, 484.80]
CO	651.00	112.30	[575.60, 726.50]	[583.50, 718.50]
CS	624.00	67.83	[578.40, 669.50]	[583.20, 664.70]
OS	775.10	203.8	[638.20, 912.00]	[652.60, 897.50]
COS	767.30	113.16	[691.20, 843.30]	[699.30, 835.30]

Note. $N = 11$ for all conditions. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

[a] Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) and are not adjusted for multiple comparisons.

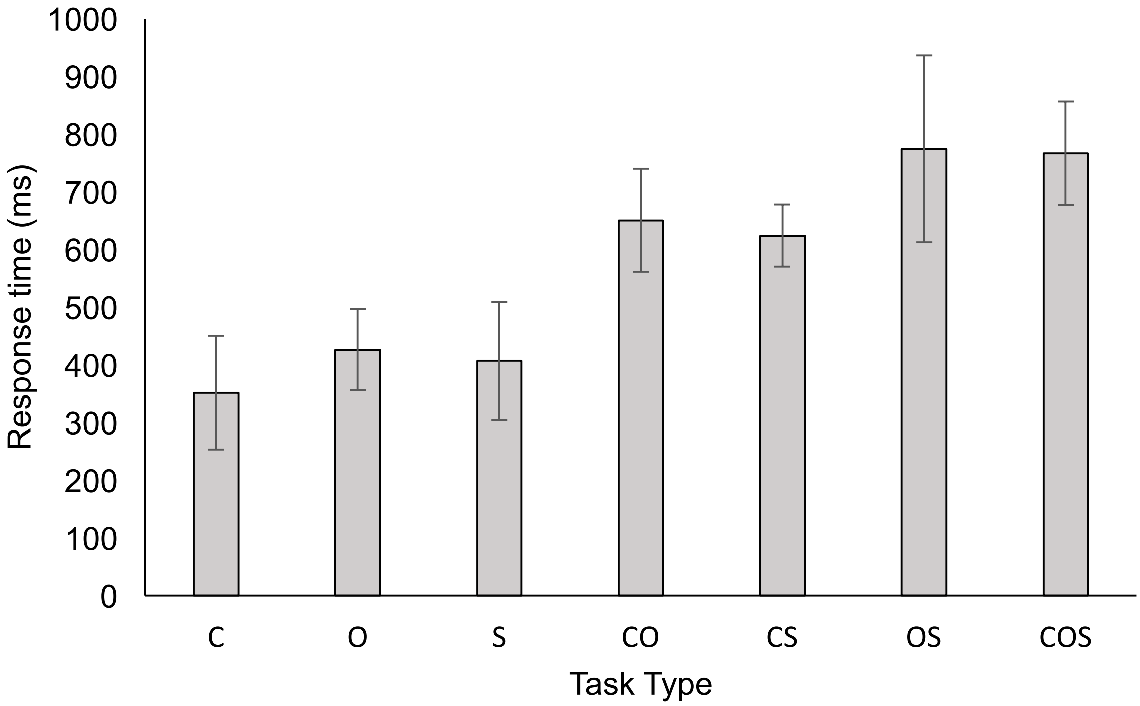

Figure 10. Response time across seven condition types.

Note. Error bars represent 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) and are not corrected for multiple comparisons. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Response time - Task within condition analysis

A final tertiary analysis on RT for each task within a condition was performed to examine the effect of other tasks present within a multi-task condition in comparison to their respective baseline single-task RT. First, each task's average response time within each condition is descriptively examined. Second, each multi-task condition's difference from baseline is examined, with the relevant paired samples *t*-tests, effect sizes, and Bayes Factors. As with accuracy, between-subjects variance for RT was removed using the grand mean of the task type. Direct comparisons across task types are therefore not advised.

An examination of the descriptive statistics (Table 12; Figures 11, 12, 13) reveals a general trend towards longer response times in multi-task conditions, with the longest response times in triple-task conditions. The dependent-samples *t*-tests comparing the difference from baseline confirmed that RT for each task within a multi-task condition was statistically significantly higher than their respective baseline, even after Bonferroni correction. It can also be observed that there was a general tendency towards larger mean differences when there were three tasks compared to two (in terms of raw scores and effects sizes), although the variability in the data (as represented by the confidence intervals) makes it unclear if these are reliable estimates. These analyses would indicate that participants were taking longer to respond in conditions with more tasks, indicating more cognitive load, likely related to increased attentional demand. However, these findings should also be taken with caution, given that the colour task was at ceiling and likely behaved differently from the other two task types.

.

Table 12.

Descriptive Statistics of Task Response Time Within Each Condition

Task Within Condition	*M*	*SD*	95% CI	95% BCI
C	352.30	110.41	[278.10, 426.50]	[286.00, 418.60]
C W/IN CO	615.60	109.03	[542.40, 688.80]	[550.10, 681.10]
C W/IN CS	570.90	55.75	[533.50, 608.40]	[537.40, 604.40]
C W/IN COS	659.80	68.99	[613.40, 706.10]	[618.30, 701.20]
O	427.00	114.63	[350.00, 504.00]	[358.10, 495.90]
O W/IN CO	686.40	119.74	[606.00, 766.90]	[614.50, 758.40]
O W/IN OS	805.30	191.81	[676.40, 934.20]	[690.00, 920.60]
O W/IN COS	851.10	152.02	[749.00, 953.20]	[759.70, 942.40]
S	407.10	121.62	[325.40, 488.80]	[334.00, 480.20]
S W/IN CS	677.00	105.70	[606.00, 748.00]	[613.50, 740.50]
S W/IN OS	744.90	175.31	[627.10, 862.60]	[639.50, 850.20]
S W/IN COS	790.90	118.10	[711.60, 870.20]	[719.90, 861.90]

Note. $N = 11$ for all conditions. Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of each condition type (colour, orientation, or size), and are not corrected for multiple comparisons. If correcting for nine comparisons, confidence intervals would be 49.57% larger. W/IN = within, C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

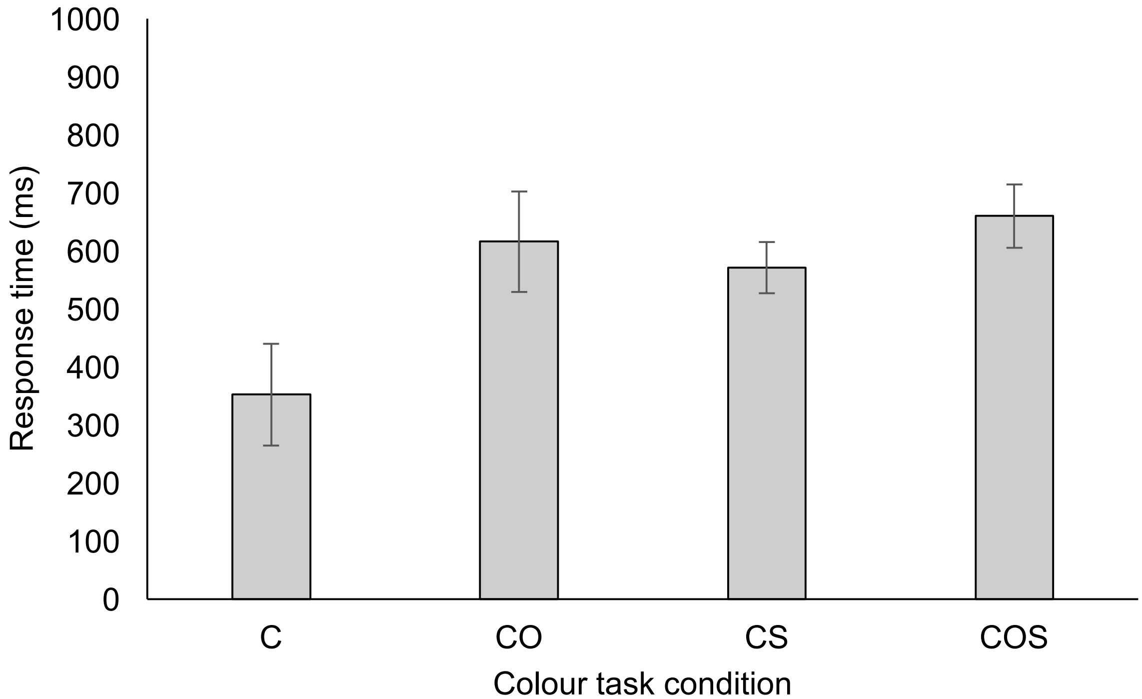

Figure 11. Response time on the colour task within each colour-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of colour task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. C = colour, CO = colour and orientation, CS = colour and size, COS = colour, orientation, and size.

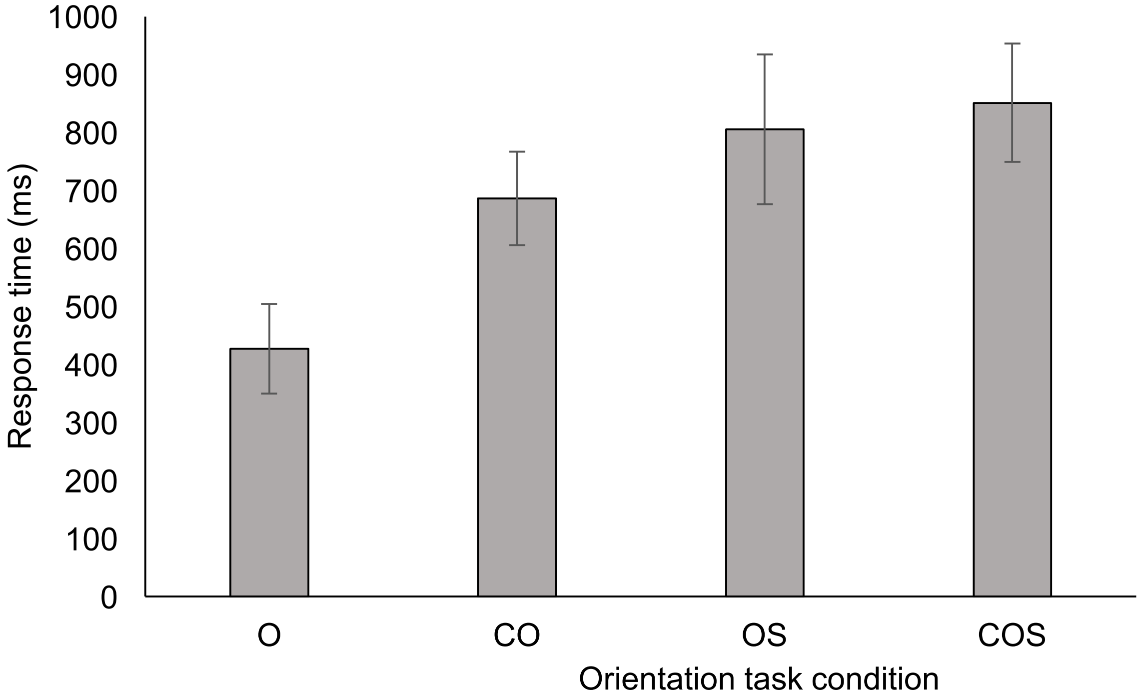

Figure 12. Response time on the orientation task within each orientation-containing condition. *Note*. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of orientation task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. O = orientation, CO = colour and orientation, OS = orientation and size, COS = colour, orientation, and size.

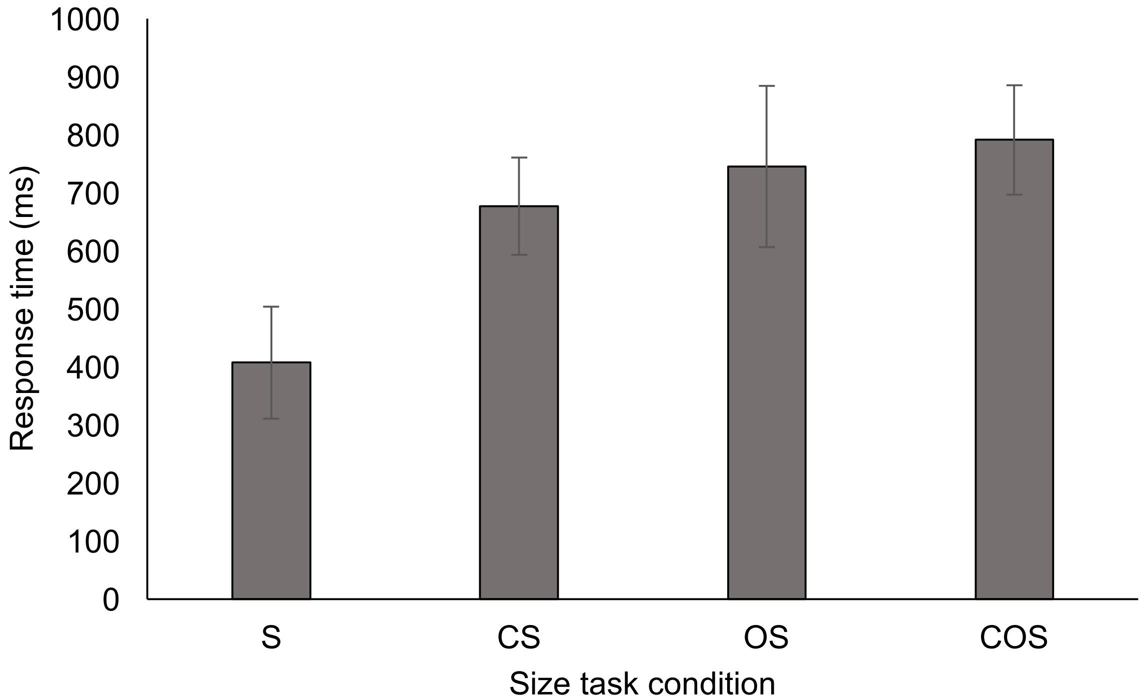

Figure 13. Response time on the size task within each size-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of size task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. S = size, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Table 13.

Response Time Difference from Baseline for Multi-Task Conditions

Contrast	t^a	*df*	$p_{bonf}{}^b$	W^c	*p*	*Mdiff*	*SEdiff*	r_{ab}	95% CI[d]	*g**	95% CI[e]
C W/IN CO - C	4.30	10	.007	.90	.194	263.30	61.24	-.71	[126.80, 399.80]	2.20	[0.74, 3.66]
C W/IN CS - C	5.32	10	.002	.88	.115	218.60	41.07	-.27	[127.10, 310.10]	2.29	[1.98, 3.92][f]
C W/IN COS - C	7.06	10	< .001	.99	.993	307.50	43.58	-.26	[210.40, 404.60]	3.06	[1.43, 4.70]
O W/IN CO - O	5.31	10	.002	.93	.440	259.40	48.83	.05	[150.60, 368.20]	2.03	[0.82, 3.24]
O W/IN OS - O	4.69	10	.004	.75	.002	378.30	80.68	-.49	[198.50, 558.10]	2.20	[0.79, 3.60]
O W/IN COS - O	6.85	10	< .001	.84	.032	424.10	61.87	-.17	[286.20, 561.90]	2.89	[1.33, 4.44]
S W/IN CS - S	5.37	10	.001	.96	.812	269.90	50.29	-.07	[157.90, 382.00]	2.17	[0.89, 3.46]
S W/IN OS - S	4.25	10	.008	.95	.622	337.70	79.52	-.56	[160.60, 514.90]	2.05	[0.67, 3.43]
S W/IN COS - S	7.07	10	< .001	.98	.983	383.80	54.3	-.13	[262.80, 504.80]	2.94	[1.37, 4.50]

[a] These are one-tailed Student's *t*-tests and assume that accuracy would be lower in multi-task conditions.

[b] *p*-values are corrected for nine comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 49.57% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of *g** takes into account the correlation between the dependent samples, r_{ab}.

[f] Difference between *SD*s was significant ($p < .05$); lower and upper bounds of the 95% *CI* represents effect size (Glass' Δ) based on treatment group *SD*, and control group *SD*, respectively.

Table 14.

Bayesian Analysis of Response Time Difference from Baseline for Multi-Task Conditions

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{+0}[a]	Error %	Median	95% BCI
C W/IN CO - C	55.60	5.23×10^{-8}	1.10	[0.35, 1.95]
C W/IN CS - C	203.47	2.81×10^{-9}	1.41	[0.54, 2.38]
C W/IN COS - C	1435.09	4.10×10^{-8}	1.89	[0.85, 3.03]
O W/IN CO - O	200.99	8.03×10^{-9}	1.40	[0.52, 2.35]
O W/IN OS - O	92.20	5.54×10^{-7}	1.22	[0.40, 2.10]
O W/IN COS - O	1160.47	8.36×10^{-8}	1.82	[0.78, 2.96]
S W/IN CS - S	214.68	2.47×10^{-8}	1.41	[0.54, 2.40]
S W/IN OS - S	51.92	4.62×10^{-7}	1.10	[0.35, 1.94]
S W/IN COS - S	1452.80	3.84×10^{-8}	1.88	[0.81, 3.04]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be greater than zero.

Discussion – Experiment 1

The aim of the current experiment was to demonstrate attentional limitations related to perceiving set summaries across multiple visual dimensions simultaneously. Boolean map theory would predict that the ability to form representations across multiple dimensions requires more stimulus exposure duration for each additional dimension to be attended. The experiment sought to show that the constraint of a constant and difficult stimulus exposure duration would elicit performance decrements in conditions requiring simultaneous attention to multiple visual dimensions. Attention to multiple visual dimensions was operationally defined as the number of tasks within a condition, with multiple tasks requiring attention to multiple visual dimensions. The overall trend of the results was consistent with the hypothesis that perceiving multiple visual dimensions simultaneously is subject to the constraints of selective attention (i.e., the constraints of selection under Boolean map theory). The constraint of a fixed stimulus exposure duration appears to elicit evidence of limitations in attentional ability related to the processes of selection. When attention is taxed by attending to more feature dimensions within the set, performance decreases, albeit not in a linear relation to the number of tasks present. This would seem to be inconsistent with some previous claims that the perceptual averaging process occurs "automatically", "in parallel", and "precedes the limited capacity bottleneck that forces selective attention" (Chong & Treisman, 2005, p. 899).

Experiment 1 of the current thesis appears to have shown a bottleneck that forces selective attention, although not caused by the number of items to process simultaneously, as was investigated by Attarha and Moore (2015b). The bottleneck according to Boolean map theory comes from the requirement to form more complicated labeled Boolean maps across multiple dimensions within a limited amount of time. The distinction between Boolean map theory and feature integration theory is important to bear in mind, since they make different claims about selective attention. In feature integration theory, selective attention is involved in serial object-by-object search, in which the number of items to operate over has an impact on performance. In contrast performing a search (or averaging) operation over all objects in parallel is a feature of 'pre-attentive' processing. In Boolean map theory, all attended objects are processed simultaneously, but each level of processing (such as performing the intersection or union operation to include information from more dimensions) must take place serially. Thus, according to Boolean map theory, the present experiment's findings suggest that selective attention was not able to unite information from multiple dimensions within the fixed stimulus exposure time of 200ms, leading to decreases in performance as the number of computations required to represent multiple dimensions increased.

The results of the present experiment can shed light on those of Attarha and Moore (2015b). Recall that they required participants to form summary representations across a set of 16 sinusoidal gratings that varied in orientation and size. Their participants engaged in three tasks, requiring them to form a representation of the average size, orientation, or average orientation and size. Their main experimental manipulation was whether their 16 sinusoidal gratings were presented either all simultaneously or sequentially (half at a time). Their stimulus exposure duration was not fixed, but was used as a means of keeping accuracy between floor and ceiling. This resulted in the stimuli being presented for an average of 60, 90, and 230ms, for the orientation, size, and orientation and size tasks, respectively. This may have biased their results towards showing no processing cost between the three conditions, which they interpreted as demonstrating that "the establishment of multiple between-feature summary representations depends entirely on parallel, unlimited- capacity processes" (p. 12).

Their claim about unlimited capacity processing can be interpreted either narrowly or broadly. A narrow interpretation would simply state that forming summary statistics across multiple visual dimensions is set-size independent. Boolean map theory would seem to agree with that statement, and this claim is not tested in the current thesis, since no set-size manipulations are undertaken. A broad interpretation of Attarha and Moore's claim would be that selective attention is not required to form representations of summary statistics across multiple visual dimensions. The findings of Experiment 1 of the current thesis would argue against this broad interpretation, since performance costs were observed as a result of forming representations across multiple dimensions, indicative of selective attention.

The results presented in this chapter are perhaps more in line with Emmanouil and Treisman's (2008) findings, which found a pre-cue advantage when participants were required to form summary representations across both speed of motion and size. However, Emmanouil and Treisman's study also required participants to form multiple summaries within a dimension in the post-cue condition, (but not the pre-cue condition) which makes it difficult to isolate the source of their observed effect. The current experiment required the formation of only one within-dimension summary, and still found an effect akin to a pre-cue advantage (with the pre-cue being provided at the start of each block, instead of at the start of each trial).

Some peculiarities exist within these data which are not readily explained by the hypothesis that attending to more dimensions requires more attention. Notably, the decrease in task performance from one dimension to two dimensions is much greater than the decrease from two dimensions to three dimensions. Boolean map theory would predict that attending to three dimensions would require more computational steps than attending to two, which would lead to even greater performance decrements in the triple-task condition compared to the dual-task conditions.

This pattern of results could be explained by observing that the colour condition accuracy was at or near ceiling, whereas performance in the orientation and size conditions were at ~90%. Second, the performance in each dual-task condition that had colour as one of the tasks had higher performance than the dual-task condition that had both orientation and size tasks. Third, the performance in the triple-task condition was close in performance to the orientation and size condition. It is possible that the colour task was too easy, and did not require much attention to complete. Therefore, when it was added to the two-dimension or three-dimension conditions, it did not decrease the performance of the block, compared to performance from the orientation or size task alone. The low difficulty of the colour task is likely owing to the fact that it was not subject to a pilot study examining the stimulus intensities required to achieve below-ceiling results in Experiment 1, unlike what was done to determine the proper size of the stimuli. If the colour task were made harder in order to require more attention, then this would likely decrease performance on each task that includes it. Consequently, one would expect that this would make the trend of decreased task performance with increased number of features more linear, and the difference between dual- and triple-task performance would be more apparent.

There is also the possibility that the non-linear decrease in performance can be partially attributed to a global task-switching cost. Researchers have shown that accuracy decreases and RT increases due to the presence of multiple tasks within a condition (Monsell, 2003). Not only is there a cost associated with switching tasks from trial to trial (termed the *switch cost*), but there is also a more general cost of performing multiple tasks within a block, as opposed to performing only one task within a block (termed the *mixing cost*). The effect of both costs is termed the *global switch cost* in the present report. Given the design of Experiment 1, there is likely a global switch cost affecting the results. However, the effect of the global switch cost is examined only for the results of Experiment 2, after the other above-mentioned issues with the experimental design had been controlled for.

Experiment 2

Experiment 2 was designed to equate the difficulty of the three tasks, and thereby avoid ceiling effects observed in the colour task of Experiment 1. If this could be achieved, it would give a clearer picture of the effects of attending to multiple dimensions simultaneously. Experiment 2 also improved aspects of the design of Experiment 1, which contained a few methodological and theoretical shortcomings. First, the colour task in Experiment 1 can be conceptualized as requiring forming summaries across two different dimensions, given that red and blue are on different dimensions of LAB colour space. Experiment 2 changed the design such that the colour task was defined in terms of saturation of the colour red. Second, a potential exploit could have affected the results of Experiment 1. Since each individual rectangle had a 76% chance of belonging to the majority dimensional feature (and thus being the correct response at the end of the trial), a participant could have adopted a fixation strategy of attending to a single rectangle (such as the rectangle near the fixation dot) in order to achieve above-chance results. Although most participants performed well above 75% correct in most conditions, it is impossible to screen out this possibility from the data of Experiment 1, so Experiment 2 added a random jitter to the location of the stimulus to throw off any potential fixation strategy.

It is worth noting that even if it were the case that participants were adopting a fixation strategy, it would not invalidate the results of Experiment 1. It would imply that participants were not computing summary statistics, but rather attending to features of a singular object. Boolean map theory makes the same predictions about groups of objects and singular objects that share features, so the decreases in performance seen in Experiment 1 could still be interpreted in light of Boolean map theory.

As per Experiment 1, was expected that attentional limitations would be observed in the formation of set summary representations across multiple dimensions. Boolean map theory would predict that if stimulus exposure duration is kept at a fixed and difficult level, the process of selection will not be able to compute representations across multiple visual dimensions. It is hypothesized that as the number of tasks within a condition increases and thus the number of feature dimensions to be attended also increases, accuracy on the task/condition should decrease and RT should increase.

Method

Experiment 2 was designed to improve aspects of Experiment 1, most notably by attempting to equate the difficulty of the three main single-tasks both within and across subjects by using a multi-step thresholding procedure. Experiment 2 was largely a replication of Experiment 1, but with some additional elements to improve on the design. Unless otherwise specified, methods of Experiment 2 were the same as Experiment 1.

Experiment 2 had three phases. Phase 1 consisted of three training blocks at high stimulus intensities to familiarize participants with the three main tasks of the experiment. Phase 2 consisted of a series of thresholding blocks aimed at estimating the 85% threshold of size, orientation, and colour saturation discrimination for each participant. Phase 3 was nearly identical to the design of Experiment 1, but used the 85% threshold estimates of the preceding thresholding blocks as the stimulus intensities for each block.

Participants

Assuming a power equal to or greater than in Experiment 1, a target of 10 valid participants was set for Experiment 2. Participants either received bonus course credits for participation or were paid a $60 honorarium for participating.

Two out of a total 30 participants were excluded from the analysis due to their inability to achieve minimum threshold performance in the colour task (see below for thresholding exclusion criteria). Fifteen participants were excluded from the analysis because they were at chance performance (<60%) in one of the experimental blocks. Participants were excluded on this basis because performance at chance makes it unclear whether participants were truly engaged in the task. One participant was excluded due to technical malfunctions with MATLAB which caused the experiment to terminate early, and one was excluded due to colour blindness. The remaining 11 participants (1 male) had an average age of 21.91 (*SD* = 3.21).

Apparatus

Stimuli were presented on the same iMac computer (OS X 10.12.3) and screen as before. Stimuli were created using the Psychophysics Toolbox Version 3.0.14 (Brainard, 1997b; Kleiner et al., 2007; Pelli, 1997) for MATLAB (version 2016b, The Mathworks, Natick, MA). Participants were required to use the chin rest for all tasks in Experiment 2; measurements of visual angle for stimuli are based on a fixed viewing distance of 180cm.

Stimuli

The 5 x 5 box of rectangles was no longer centered in the screen, but its central location was randomly jittered in the x and y coordinates uniformly between -1.17 and 1.17°. This was to prevent a fixation strategy of participants only attending to one rectangle as opposed to the global average.

Experiment 2 was observer-paced and designed in the manner of an ERP study (Luck, 2014). After pressing the spacebar, the fixation dot would turn white immediately and the stimulus would appear after a gap that varied randomly between 300 and 600ms.

In Experiment 2, stimuli size and colour were variable. The minimum width and length of a rectangle was 0.16° (10 pixels). Large rectangles increased their width and length within a range from 0.016 to 0.31° (1 to 20 pixels). Orientation discriminability was varied by changing the length of the rectangles. The increase in length could range from 0.016 to 0.47° (1 to 30 pixels). Colour discriminability was varied by changing the saturation of red. Full-colour 'dark' red used used RGB values of [1.0, 0.0, 0.0], whereas 'light' red used RGB values ranging from .01 to .20 in the blue and green channels. Green and blue values closer to 1.0 added more white light in the stimuli and thus were easier to distinguish from dark red. For this report, 'percent white' is used as shorthand for the amount of white signal in the stimuli. Five percent white corresponds to [1.0, .05, .05], i.e., the percentage value added to the blue and green channels. Twenty percent white corresponds to [1.0, 0.20, 0.20]; a lighter red signal easier to distinguish from dark red. For the presentation of the response query, the two choices were presented at the same intensities that the stimuli were during the trial (see Figure 14).

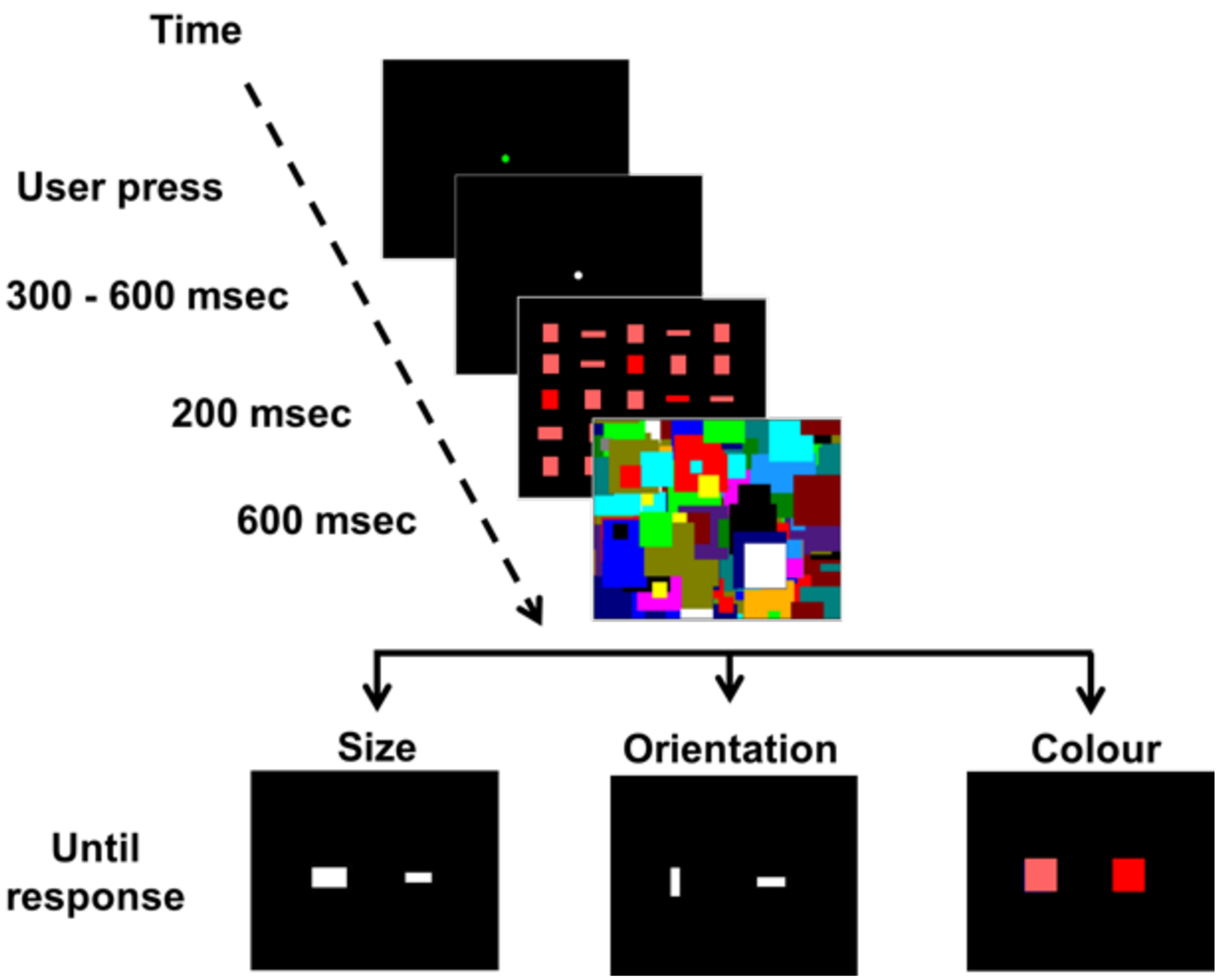

Figure 14. Trial events for Experiment 2.

Note. Stimuli sizes and saturation differences are increased for illustration purposes.

Procedure

Phase 1. Participants were first given three training blocks with feedback at high stimulus intensities. Size and orientation were kept at their Experiment 1 levels; light red consisted of 20 percent white rectangles. Participants completed one practice block of 48 trials for each single-task condition, in random order. This portion of the experiment was intended to be relatively easy for participants to complete and was merely used to familiarize the participants with each of the three main tasks.

Phase 2. Participants then went on to complete three thresholding tasks aimed at finding each participant's 85% accuracy threshold for size, orientation, and colour. This level of accuracy was chosen to avoid both floor and ceiling effects. Each thresholding block presented the same type of stimuli as in Phase 1, but varied the intensity of the visual dimension of interest within a range of linearly spaced stimulus intensities. In the size condition, the stimulus intensities ranged from 0.016 to 0.31° (1 to 20 pixel) increases in 10 linearly spaced increments. The orientation condition varied the change in length from .016 to 0.47° (1 to 30 pixels) in 15 linearly spaced increments. The colour condition varied the percentage white from 1% to 20% in 40 linearly spaced increments. These values were chosen as a result of pilot experiments which demonstrated that most participants performed near chance at the lowest stimulus intensities and near ceiling at the highest stimulus intensities.

The thresholding procedure used the psi method (Kontsevich & Tyler, 1999) implemented using the Palamedes toolbox in MATLAB (Prins & Kingdom, 2009). The psi method is a Bayesian adaptive method that seeks to minimize entropy on each trial while estimating the slope and position (threshold) of a psychometric function (Kingdom & Prins, 2010). The Weibull function was assumed for each task because the stimulus intensities were linearly spaced, with a stimulus intensity of $x = 0$ corresponding to an absence of signal. Given these characteristics, the Cumulative Normal and Logistic functions were ruled out as possibilities (Kingdom & Prins, 2010).

The psi method would search for alpha (threshold) and beta (slope) values, and assumed a fixed lapse rate (lambda) of .02, and a guess rate (gamma) of .50. The alpha prior range was linearly spaced from the minimum stimulus intensity of the task type to the maximum, with a grain of 201. The beta prior range was determined on the basis of pilot experiments which used a method of constant stimuli design and maximum likelihood fitting to estimate the slope. The minimum observed slope was multiplied by 0.1 and the maximum was multiplied by 10 in order to create a wide prior range. For all tasks, the beta prior range was uniform and linearly spaced from the base 10 logarithm of the adjusted minimum to the base 10 logarithm of the adjusted maximum, with a grain of 201. For colour, the minimum and maximum observed slopes were 1.22 and 133.11, for orientation, they were 0.48 and 47.97, and for size they were 0.66 and 20.80.

Participants completed a training block of 64 trials prior to each experimental block of 128 trials. Participants first completed a size threshold task, followed by an orientation threshold task, followed by a colour threshold task. While the stimulus intensity of the dimension under question was varied according to the psi method, the other two dimensions' stimulus intensities were held constant. The initial values used for size were 0.13° (8 pixel) size increases for length and width, for orientation were 0.16° (10 pixel) length increases, and for colour were 4.3% white. These were derived from the average 85% threshold values observed in a previous pilot experiment.

The data generated by the psi method in each respective task were fit using a maximum-likelihood search function using the Palamedes toolbox (Prins & Kingdom, 2009) in MATLAB. The Weibull function was assumed, with gamma (guess rate) and lambda (lapse rate) values set at .5 and .02, respectively. Alpha (85% threshold) and beta (slope) were free parameters for which the function searched. The alpha search space was defined as being between the minimum and maximum stimulus intensities of the threshold block, in linear increments of 0.001. The beta search space was between 0 and 3 in 101 logarithmically spaced points.

After each successive experimental block of the thresholding task, the resulting 85% threshold estimate was used as a fixed value for each following threshold block. Training data were not used to alter stimulus intensities. This thresholding design was used because it was believed on the basis of pilot experiments that the colour saturation threshold was most sensitive to changes in the size of the stimuli, whereas size and orientation were relatively unaffected by colour saturation. Size increases were also believed to affect the orientation threshold more than the reverse. To anticipate the results, the performance in each single-task condition in Phase 3 were not exactly at 85%, with colour performance being ~70%. However, the thresholding task achieved its primary aim of bringing performance to below-ceiling levels.

Participants were excluded from Phase 3 of the study if the function had to extrapolate beyond the maximum tested values to reach the 85% accuracy threshold (i.e., if the 85% threshold was above the maximum stimulus intensity, the participant was excluded).

Phase 3. The experimental procedure for Phase 3 of Experiment 2 was identical to Experiment 1 except for the following notes. Participants completed 12 experimental blocks with equal numbers of trials (96 each); 1 block each of the single-task conditions, 2 blocks each of the dual-task conditions, and 3 blocks of the triple-task condition. This equates to the same number of trials that were in each condition in Experiment 1, but blocks were made to be of equal durations in Experiment 2. Participants completed 50% more practice trials than in Experiment 1 (i.e., 48 trials per block). This was done in order to ensure that participants were trained sufficiently on the harder tasks. The entire experiment took approximately three hours. Despite the long duration of the experiment, fatigue was determined to not be systematically related to performance for included participants (see Appendix A).

Task

For all phases of Experiment 2, participants were asked to start each trial by pressing the spacebar with their left hand, and use two fingers on their right hand to respond to the prompt using the left/right arrow keys. Participants took a mandatory break of 1 minute after each experimental block, and participants were also allowed to take a break after a practice block. They were also allowed to take short breaks prior to starting a trial, should they need to rest their eyes or adjust position. There was also a mandatory short (~5-minute) break after Phase 2 where the participant was allowed to leave the testing room for a short period. Participants were instructed about the procedure for each phase of the experiment. They were told to make their best guess when responding if they were unsure about the majority colour, size, or orientation presented in the stimulus. Finally, participants were asked to refrain from moving their eyes or blinking during the presentation of the stimulus.

Results

The main analyses of Experiment 2 were conducted similarly to those of Experiment 1. Some additional analyses were also undertaken to describe the additional features of Experiment 2, such as discussion of excluded participants and the results of the thresholding procedure. A further analysis was conducted to examine the effect of task switching on accuracy and RT performance. Each analysis was aimed at answering the question of whether forming summary representations across multiple dimensions produces an attentional cost under the constraint of constant stimulus exposure duration. After the characteristics of included versus excluded participants and the results of the thresholding procedure were described, the analysis proceeded from coarse to fine-grained, as in Experiment 1. Performance at the level of the number of tasks within a condition was analyzed first so as to show if there was an overall decrease in performance as the number of attended visual dimensions increased. A secondary descriptive analysis examined the results at the level of each of the seven conditions in order to describe the effects of performing more tasks within a condition more closely. A tertiary analysis examined the effect of the presence of other tasks on task performance within a condition (e.g. the effect of adding the orientation task on colour task performance). A fourth analysis examined whether dual- and triple-task decreases in performance could be accounted for by task-switching costs.

Accuracy - Included versus excluded participants

A total of 15 participants were excluded from the analysis due to near-chance performance (< 60%) in one or more blocks of the experiment. A descriptive comparison of the accuracy of these 15 excluded participants versus the 11 included participants for each of the seven conditions reveals that included participants tended to have greater accuracy in most conditions, with the size single-task condition and the colour and orientation condition perhaps being exceptions to this trend (see Table 15, Figure 15). More than half of the excluded participants had performance near chance during the colour single-task condition, indicating that the thresholding procedure failed to find the appropriate value for this dimension in these participants (see the discussion of Experiment 2 for a possible explanation). The orientation and size tasks do not appear to suffer from this issue, as their performance was near the 85% accuracy target of the thresholding procedure.

The 15 excluded participants also tended to have greater variance than the 11 included participants. This is perhaps unsurprising, given that the 11 included participants represent a high-performing group, and thus a restricted subset of the population. The subsequent analyses using only the included 11 participants should therefore be interpreted with the caveat that they represent a high-performance group.

Table 15.

Descriptive Statistics of Accuracy Across Seven Condition Types, for Excluded and Included Participants

	Excluded				Included		
Condition	*N*	*M*	*SD*	% At Chance [a]	*N*	*M*	*SD*
C	15	65.35	10.54	53.33%	11	70.17	6.49
O	15	84.38	7.84	0.00%	11	90.34	5.39
S	15	87.85	4.65	0.00%	11	87.78	6.96
CO	15	66.39	3.29	13.33%	11	69.74	2.56
CS	15	62.64	6.21	53.33%	11	68.99	4.76
OS	15	76.88	4.24	0.00%	11	83.29	6.16
COS	15	67.20	6.01	40.00%	11	71.18	3.17

Note. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size. Standard deviation is not corrected for between-groups variance.

[a] This column represents the percentage of excluded participants who had chance performance during one or more blocks within the given condition type.

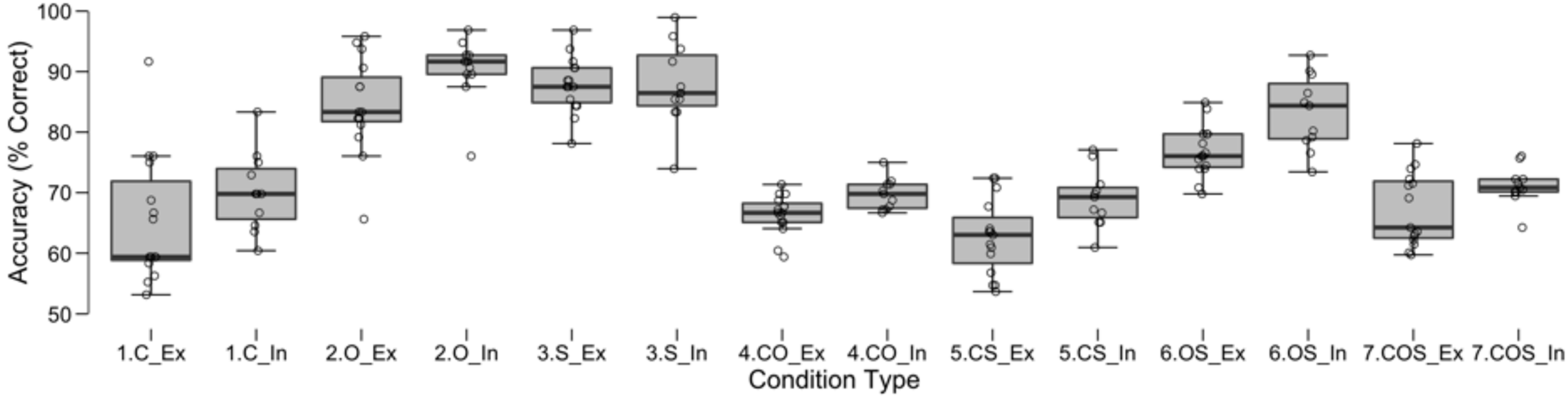

Figure 15. Accuracy boxplot of excluded (Ex; $N = 15$) and included (In; $N = 11$) participants in each of the seven condition types. *Note*. The black central bar indicates the group median, the surrounding grey box indicates first and third quartiles, lines represent 1.5 of the interquartile range, circles indicate individual cases. Variance was not corrected for between-subjects error. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Thresholding results

Prior to the main experiment consisting of seven conditions, participants engaged in a thresholding procedure designed to determine the 85% threshold for the size, orientation, and colour tasks. The descriptive statistics of the 11 participants included in the full analysis are reported below (Table 16) and displayed graphically in Figure 16. For these analyses, variance was not corrected for between-subjects error, as there are no within-subjects comparisons to be made, and each threshold is measured on a different scale.

Table 16.

Descriptive Statistics of Thresholds and Slopes for Colour, Orientation, and Size

Parameter	*M*	*SD*	95% CI	95% BCI
Colour 85% Threshold	0.08	0.02	[0.07, 0.10]	[0.07, 0.10]
Colour Slope	2.33	1.22	[1.51, 3.14]	[1.51, 3.14]
Orientation 85% Threshold	8.50	3.62	[6.07, 10.94]	[6.07, 10.94]
Orientation Slope	0.88	0.33	[0.65, 1.10]	[0.65, 1.10]
Size 85% Threshold	4.51	1.22	[3.68, 5.33]	[3.68, 5.33]
Size Slope	1.51	0.60	[1.11, 1.92]	[1.11, 1.92]

Note. $N = 11$ for all conditions. CI = Confidence interval (NHST), BCI = Bayesian credibility interval. Confidence intervals are not constructed using within-subjects corrected variance and are not corrected for multiple comparisons. Colour threshold is in units of proportion of white light added to pure red, with 0 representing zero white light. Orientation and size are measured in pixels (1 pixel = 0.016°).

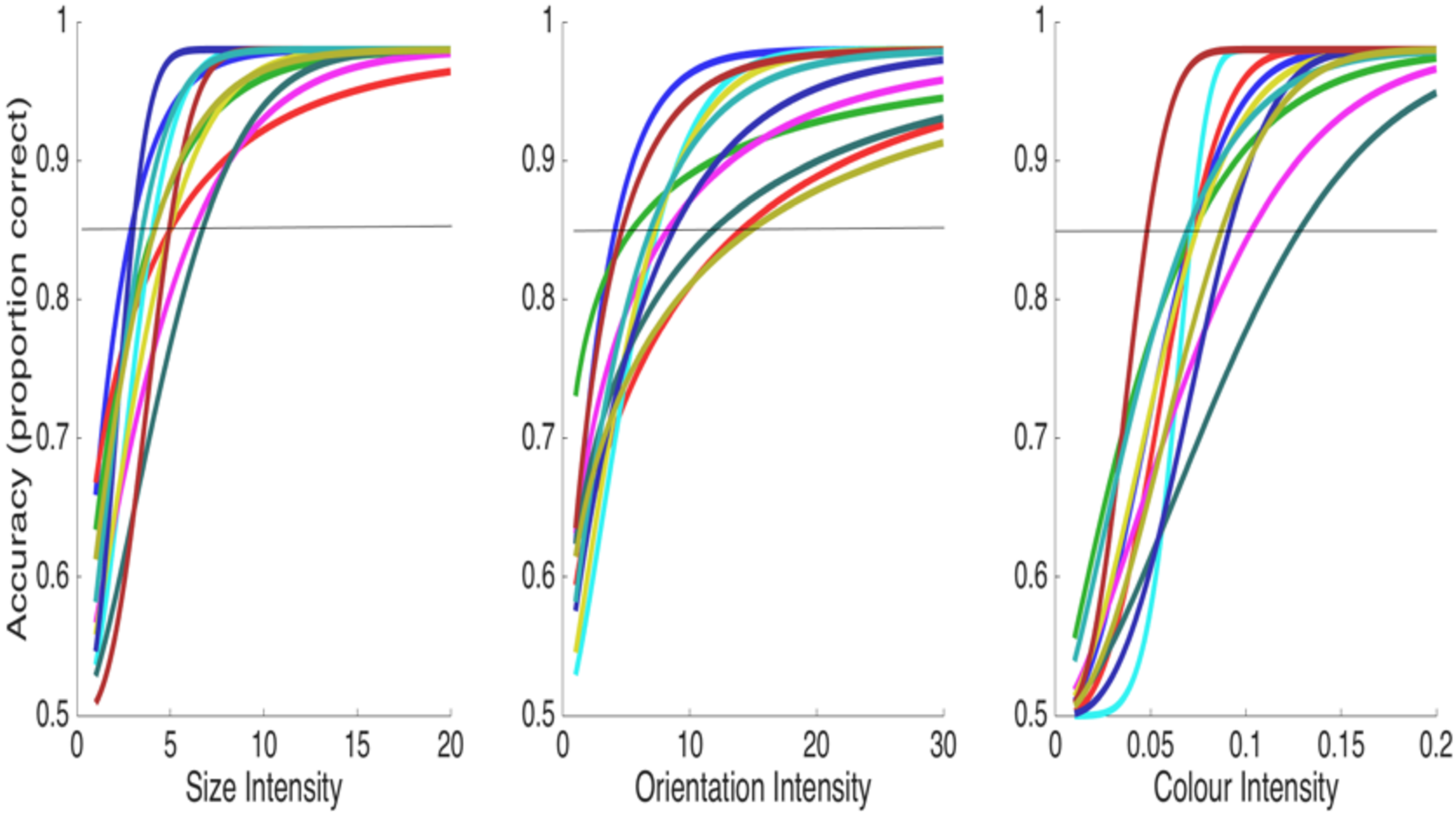

Figure 16. Maximum-likelihood fits for size, orientation, and colour ($N = 11$).

Note. Size intensity is measured in the change of width and length in pixels (1 pixel = 0.016°), orientation intensity is length of rectangle in pixels, colour intensity is in proportion white light in a pure red signal.

Accuracy - Primary analysis

Over the course of the experiment, participants engaged in seven conditions composed of the colour, orientation and size tasks. They engaged in three single-tasks, three dual-tasks and one triple-task. The main research question of this study is whether forming representations across multiple dimensions shows processing limitations under conditions of constant-but-difficult stimulus exposure duration. The results are analyzed here in terms of the number of tasks within a condition and its effect on accuracy performance.

A 1 x 3 repeated-measures ANOVA across the mean accuracy of the three single-task conditions, the mean of the three dual-task conditions, and the triple-task condition was conducted. The assumption of sphericity was not violated, Mauchly's $W = .64$, $p = .135$, nor was the assumption of normality, Shapiro-Wilk $W = 0.97$, $p = .477$. The ANOVA revealed that accuracy decreased as the number of tasks within a condition increased, $F(2,20) = 70.20$, $p < .001$, $\eta^2 = 0.875$, (see Table 17; Figure 17). Follow-up post-hoc Bonferroni-corrected *t*-tests (Table 18) revealed that single-task accuracy was higher than both dual-task and triple-task performance. Average accuracy in dual-tasks was superior to that of the triple-task, although to a lesser degree than the difference from single-task to dual task conditions. This would seem to indicate that accuracy does decrease as the number of attended dimensions increases, although perhaps in a non-linear fashion, as was found in Experiment 1, with less relative decrease as more dimensions are attended to. The effect of of task-switching may be able to partially account for this non-linearity (discussed further below).

Table 17.

Descriptive Statistics of Accuracy Across Number of Tasks Within a Condition

Number of Tasks	*M*	*SD*	95% *CI*	95% BCI
1 Task	82.77	2.36	[81.18, 84.35]	[81.18, 84.35]
2 Tasks	74.01	2.05	[72.63, 75.38]	[72.63, 75.38]
3 Tasks	71.18	1.29	[70.31, 72.05]	[70.31, 72.05]

Note. $N = 11$ for all conditions. CI = Confidence interval (NHST), BCI = Bayesian credibility interval. Confidence intervals are constructed using within-subjects corrected variance and are not corrected for multiple comparisons (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

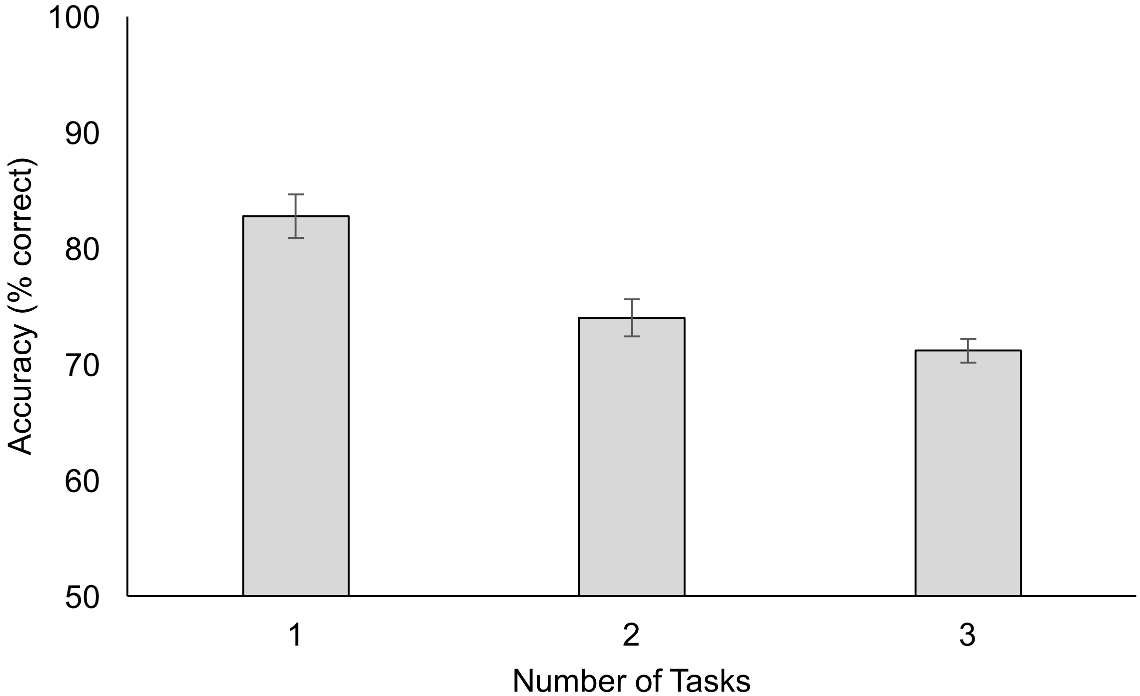

Figure 17. Mean accuracy across the number of tasks within a condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

Table 18.

Pairwise Comparisons of Accuracy Across Number of Tasks Within a Condition

Contrast	t^a	*df*	$p_{bonf}{}^b$	W^c	*p*	M_{diff}	SE_{diff}	r_{ab}	95% CI[d]	*g**	95% CI[e]
1 Task – 2 Tasks	6.88	10	< .001	.95	.670	8.76	1.27	-.84	[5.92, 11.60]	3.64	[1.68, 5.59]
1 Task – 3 Tasks	11.97	10	< .001	.88	.107	11.59	0.97	-.50	[9.43, 13.74]	5.58	[4.91, 8.97]f
2 Tasks – 3 Tasks	3.78	10	.005	.96	.763	2.83	0.75	-.05	[1.16, 4.50]	1.51	[0.43, 2.60]

[a] These are one-tailed Student's *t*-tests and assume that more dimensions lead to lower accuracy.

[b] *p*-values are corrected for three comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 24.42% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of *g** takes into account the correlation between the dependent samples, r_{ab}.

[f] Difference between *SD*s was significant ($p < .05$); Lower and upper boundaries of 95% CI represent effect size (Glass' Δ) based on treatment group (more tasks) *SD*, and control group (fewer tasks) *SD*, respectively.

Bayesian analyses revealed similar findings to the NHST results. A Bayesian repeated-measures ANOVA revealed a BF_{10} = 163,500,000,000, Error % = 0.93, indicating that the alternative hypothesis is approximately 163 billion times more likely to explain these data than the null hypothesis. Follow-up Bayesian *t*-tests (Table 19) revealed a similar pattern to the NHST tests in that there is very strong evidence (BF_{+0} > 1000) in favour of the hypothesis that average task accuracy in single tasks was greater than that of dual or triple-tasks. There is still strong evidence (BF_{+0} = 27.77) that accuracy was lower in the triple-task condition compared to dual-task conditions.

Table 19.

Bayesian Analysis of Comparisons of Accuracy Across Three Conditions

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{+0}[a]	Error %	Median	95% BCI
1 Task – 2 Tasks	1186.35	8.00×10^{-8}	1.82	[0.82, 2.95]
1 Task – 3 Tasks	93621.00	3.04×10^{-9}	3.27	[1.70, 5.02]
2 Tasks – 3 Tasks	27.77	5.88×10^{-7}	0.97	[0.27, 1.78]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be greater than zero.

Accuracy - Seven condition analysis

A descriptive examination of accuracy across each of the seven condition types (Table 20, Figure 18) revealed that accuracy in the size, orientation, and colour conditions was ~88%, ~90% and ~70%, respectively. For the high-performing ($N = 11$) group, it appears that the thresholding procedure was relatively accurate in determining the 85% thresholds for size and orientation, but not for colour. However, the thresholding procedure was able to bring performance below ceiling, which was one of its main aims.

The experiment was based on the hypothesis that attending to more visual dimensions would reduce accuracy, providing contrary evidence to the claim that forming multiple between-dimension summaries is a cost-free process. As in Experiment 1, the results are difficult to interpret due to the unusual behaviour of the colour task. It appears that conditions which contained the colour task were close to each other in accuracy performance (i.e., they were all near 70%, +/- 2%). Examination of the performance of conditions that contained the orientation and size tasks revealed a trend more in line with the hypothesis that attending to multiple visual dimensions is limited by the processes of selective attention. Performance in the orientation or size conditions decreased from ~88% in single-task conditions to ~83% in the dual-task condition, and to ~70% in the triple-task condition.

As a consequence of the colour task being more difficult than the orientation and size tasks, participant strategizing may have played a role in the results of multi-task conditions. Participants could have adopted the strategy of maximizing performance on non-colour tasks when the colour task was present in the condition. A more fine-grained analysis examining task performance within each condition could help elucidate this hypothesis (see below). To anticipate, it appears that many participants (~ 50%) may have adopted this strategy, as their performance in the colour task was near chance in multi-task blocks. It is worth reiterating that each of these participants was above chance performance in each block of the experiment, so these data do not exhibit a floor effect, strictly speaking. Participants were not screened out on the basis of their adopting a particular strategy.

Table 20.

Descriptive Statistics of Accuracy Across Seven Conditions

Condition	M	SD	95% CI[a]	95% BCI
C	70.17	6.36	[65.90, 74.44]	[65.90, 74.44]
O	90.34	2.97	[88.34, 92.34]	[88.34, 92.34]
S	87.78	5.09	[84.37, 91.20]	[84.37, 91.20]
CO	69.74	2.98	[67.74, 71.75]	[67.74, 71.75]
CS	68.99	3.90	[66.37, 71.61]	[66.37, 71.61]
OS	83.29	3.96	[80.62, 85.95]	[80.62, 85.95]
COS	71.18	1.66	[70.06, 72.30]	[70.06, 72.30]

Note. N = 11 for all conditions. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

[a] Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) and are not adjusted for multiple comparisons.

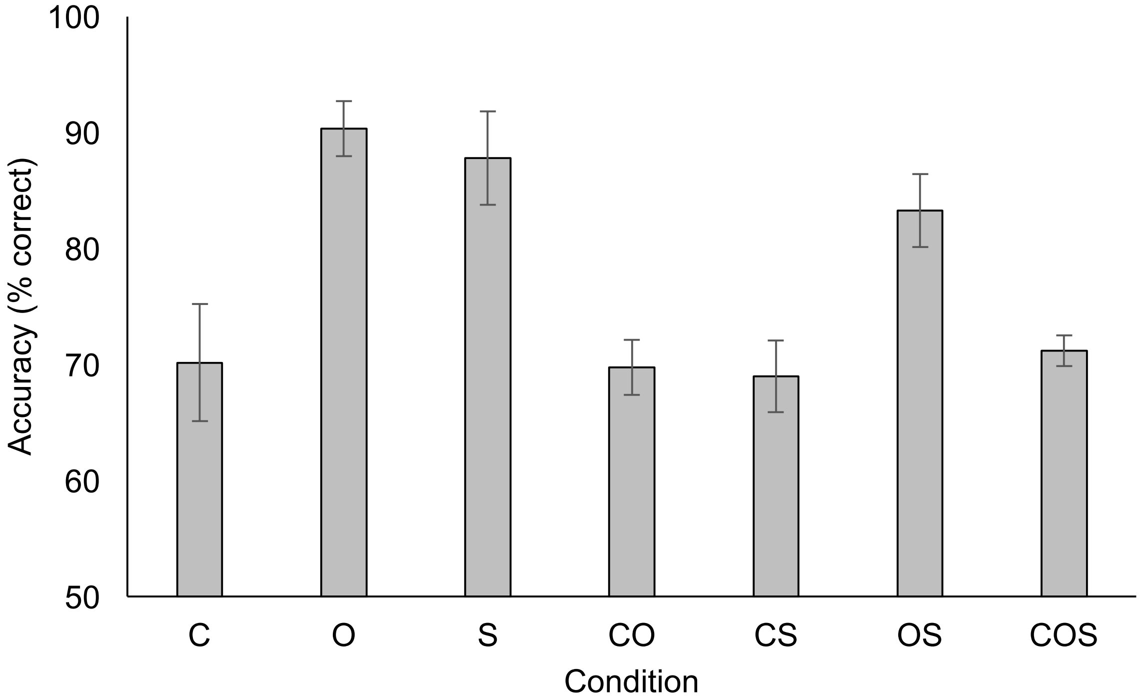

Figure 18. Accuracy across seven condition types.

Note. Error bars represent 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) and are not corrected for multiple comparisons. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Accuracy - Task within condition analysis

A third analysis investigating task performance within each condition revealed that performance in each task decreased from its respective baseline when an additional task was added (Table 21, Figures 19, 20, 21). A series of paired-samples *t*-tests comparing the accuracy on each task in a multi-task condition compared to its respective single-task baseline (Table 22) largely confirmed this general observation. All comparisons showed a decrease of over 5 accuracy percentage points in accuracy compared to baseline (M_{dif} range -5.68 – -13.07; Hedge's g^* range: 1.32 - 3.29), and all but one test was significant after Bonferroni-correction. The Bayesian analyses showed the same pattern, with BF_{-0} ranging from 7.78 to 1639 (Table 23). However, the relationship between the number and type of tasks in each condition and subsequent accuracy on the task of interest was not straightforward. Task accuracy tended to be lowest when the colour task was one of the additional tasks (i.e., the colour and size; colour and orientation; and colour, orientation, and size conditions have lower accuracy compared to the orientation and size condition), with performance being near chance for each colour task of a multi-task condition. In the two-feature conditions, 45% of participants were indistinguishable from chance in the colour and orientation condition, and the same number were indistinguishable from chance in the colour and size condition. Further, 72% of participants were statistically indistinguishable from chance performance in the colour task within the colour, orientation, and size condition. This may have been due to participant strategizing; many participants may have opted to focus on the other task(s) present in the condition when colour was included. The presence of an unusually difficult colour task and possible systematic participant strategizing makes interpretation of the results difficult. However, it is clear that performance tended to decrease as there were more tasks present within the condition, consistent with the hypothesis that attending to more visual dimensions is an attentionally demanding task, which occurs a cost of selective attention.

It also appears that the decrease in accuracy associated with the number of tasks present in the condition was non-linear. The non-linearity of the decrease in accuracy might be attributable to the fact that in single task conditions, participants did not have to switch tasks, whereas in multi-task conditions, participants must have switched tasks at unexpected times. The presence of a task-switching cost may be able to account for much of the non-linearity of the decrease in accuracy. It could also perhaps even explain the total effect of the decrease in accuracy associated with the number of tasks present within a set, which was hypothesized to be due to the limits of perception associated with selective attention.

Table 21

Descriptive Statistics of Accuracy for Tasks Within Each Condition

Task Within Condition	*M*	*SD*	95% CI	95% BCI
C	70.17	4.80	[66.95, 73.39]	[66.95, 73.39]
C W/IN CO	58.05	3.55	[55.66, 60.43]	[55.66, 60.43]
C W/IN CS	62.41	4.33	[59.50, 65.31]	[59.50, 65.31]
C W/IN COS	57.10	3.12	[55.01, 59.20]	[55.01, 59.20]
O	90.34	3.25	[88.16, 92.53]	[88.16, 92.53]
O W/IN CO	81.44	3.03	[79.40, 83.48]	[79.40, 83.48]
O W/IN OS	84.66	3.31	[82.43, 86.88]	[82.43, 86.88]
O W/IN COS	80.78	1.92	[79.49, 82.06]	[79.49, 82.06]
S	87.78	4.17	[84.99, 90.58]	[84.99, 90.58]
S W/IN CS	75.57	4.48	[72.56, 78.57]	[72.56, 78.57]
S W/IN OS	81.91	4.01	[79.22, 84.60]	[79.22, 84.60]
S W/IN COS	75.66	3.08	[73.59, 77.73]	[73.59, 77.73]

Note. $N = 11$ in all conditions. Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of each condition type (colour, orientation, or size), and are not corrected for multiple comparisons. If correcting for nine comparisons, confidence intervals would be 49.57% larger. W/IN = within, C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

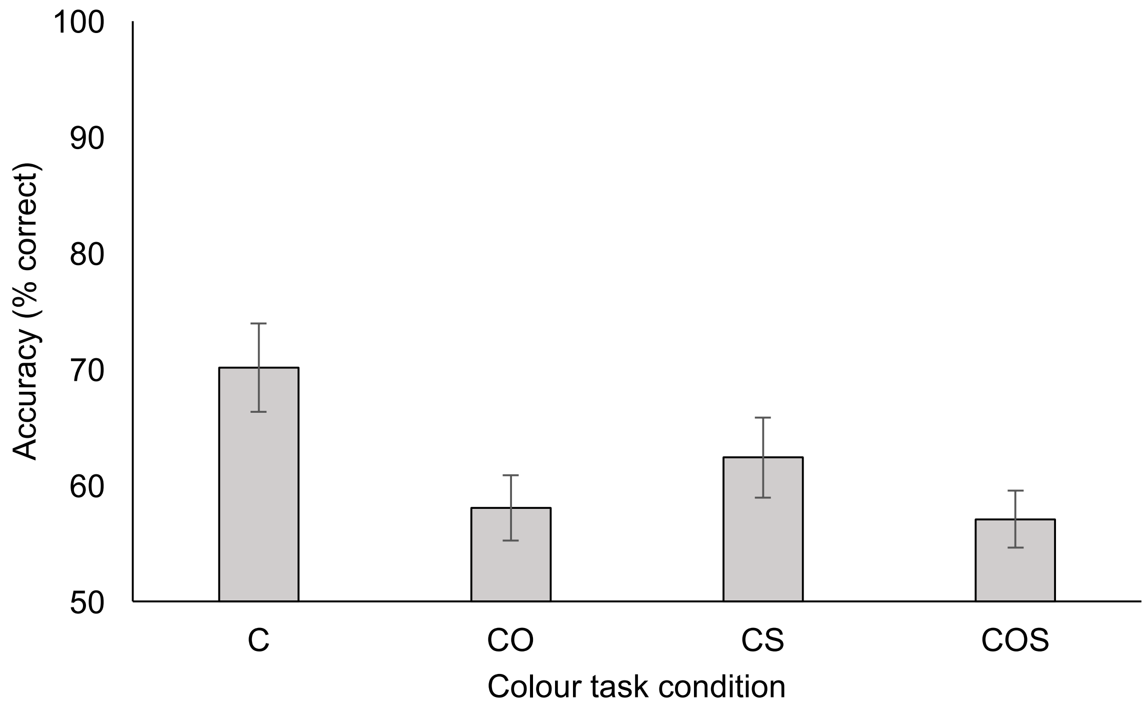

Figure 19. Accuracy on the colour task within each colour-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of colour task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. C = colour, CO = colour and orientation, CS = colour and size, COS = colour, orientation, and size.

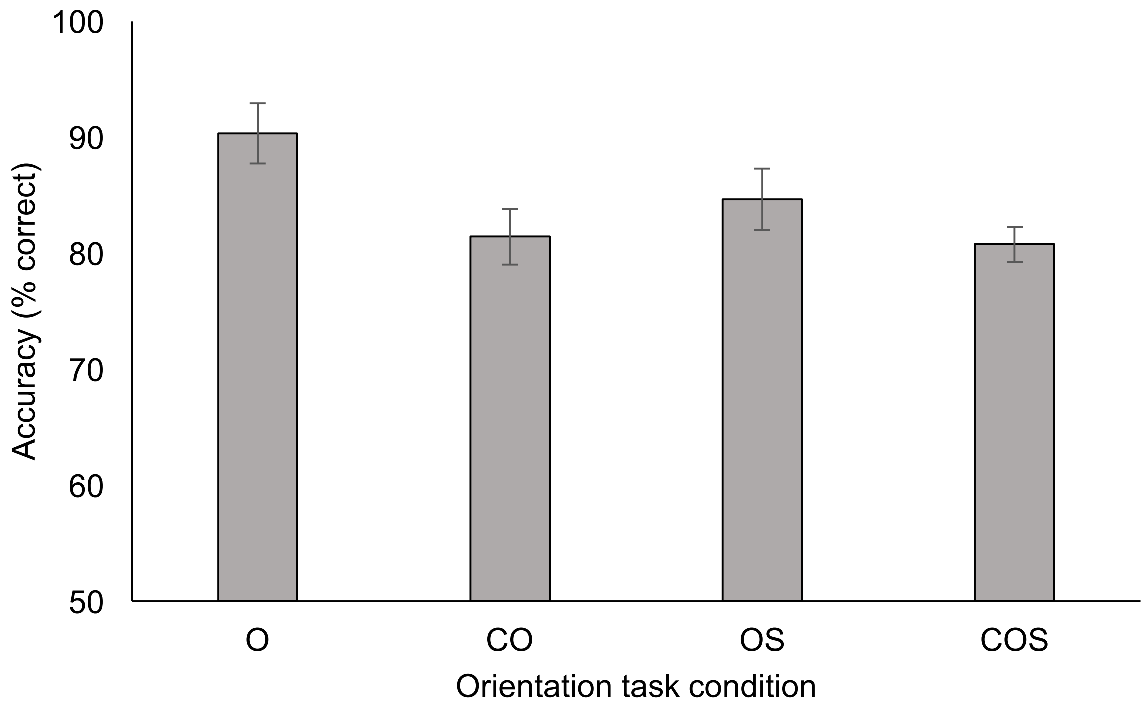

Figure 20. Accuracy on the orientation task within each orientation-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of orientation task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. O = orientation, CO = colour and orientation, OS = orientation and size, COS = colour, orientation, and size.

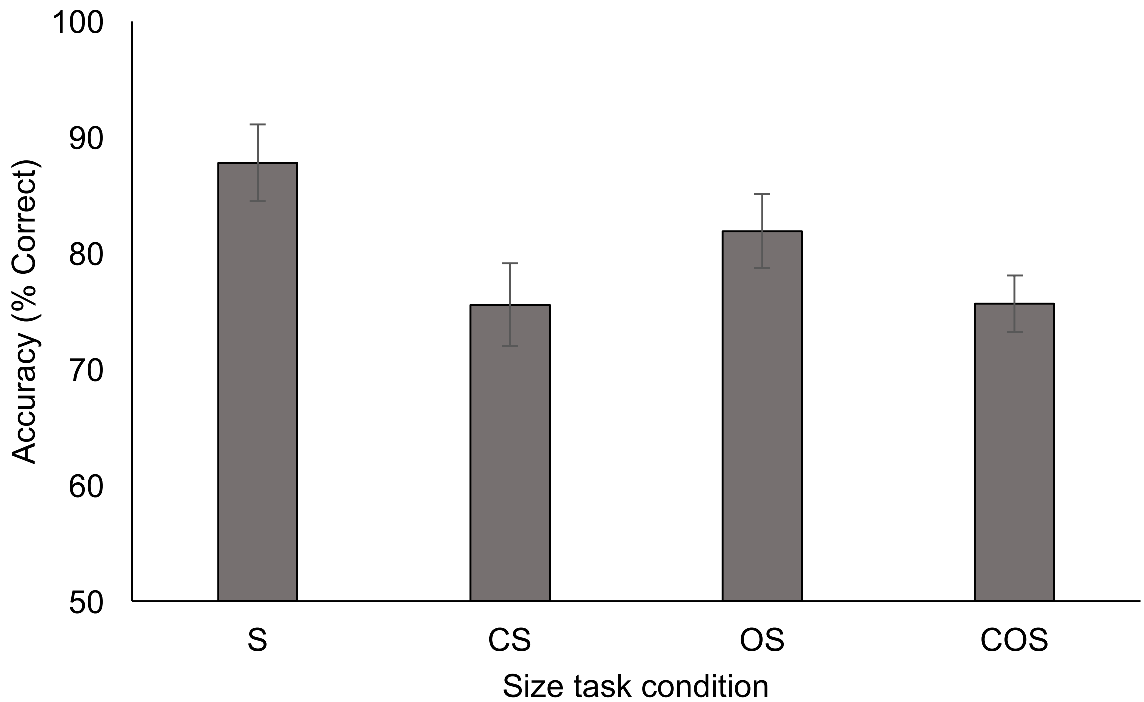

Figure 21. Accuracy on the size task within each size-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of size task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. S = size, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Table 22.

Accuracy Difference from Baseline for Multi-Task Conditions

Contrast	t^a	df	p_{bonf}^b	W^c	p	M_{diff}	SE_{diff}	r_{ab}	95% CI[d]	g^*	95% CI[e]
C W/IN CO - C	-5.48	10	.001	.91	.240	-12.12	2.21	-.54	[-7.19, -17.05]	-2.63	[-4.19, -1.08]
C W/IN CS - C	-3.55	10	.024	.94	.542	-7.77	2.19	-.26	[-2.89, -12.64]	-1.56	[-2.72, -0.40]
C W/IN COS - C	-6.16	10	.001	.94	.549	-13.07	2.12	-.56	[-8.34, -17.80]	-2.96	[-4.63, -1.30]
O W/IN CO - O	-5.45	10	.001	.94	.477	-8.90	1.63	-.49	[-5.26, -12.54]	-2.60	[-4.12, -1.07]
O W/IN OS - O	-3.57	10	.023	.95	.625	-5.68	1.59	-.29	[-2.14, -9.23]	-1.59	[-2.76, -0.41]
O W/IN COS - O	-7.19	10	<.001	.96	.718	-9.56	1.33	-.42	[-6.60, -12.53]	-3.29	[-5.04, -1.54]
S W/IN CS - S	-5.70	10	.001	.96	.705	-12.22	2.14	-.35	[-7.44, -16.99]	-2.59	[-4.09, -1.10]
S W/IN OS - S	-2.86	10	.077	.95	.681	-5.87	2.06	-.39	[-1.29, -10.45]	-1.32	[-2.46, -0.17]
S W/IN COS - S	-6.77	10	<.001	.97	.874	-12.12	1.79	-.33	[-8.14, -16.11]	-3.03	[-4.68, -1.39]

[a] These are one-tailed Student's *t*-tests and assume that accuracy would be lower in multi-task conditions.

[b] *p*-values are corrected for nine comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 49.57% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of g^* takes into account the correlation between the dependent samples, r_{ab}.

Table 23.

Bayesian Analysis of Accuracy Difference from Baseline for Multi-Task Conditions

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{-0}[a]	Error %	Median	95% BCI
C W/IN CO - C	245.09	1.59×10^{-7}	-1.45	[-2.42, -0.55]
C W/IN CS - C	20.31	7.90×10^{-7}	-0.91	[-1.68, -0.22]
C W/IN COS - C	541.02	8.93×10^{-8}	-1.63	[-0.65, -2.68]
O W/IN CO - O	236.49	1.19×10^{-7}	-1.45	[-2.42, -0.54]
O W/IN OS - O	20.85	7.61×10^{-7}	-0.91	[-1.69, -0.24]
O W/IN COS - O	1639.73	1.68×10^{-8}	-1.92	[-3.09, -0.85]
S W/IN CS - S	318.23	3.78×10^{-7}	-1.51	[-2.52, -0.59]
S W/IN OS - S	7.78	1.79×10^{-7}	-0.72	[-1.43, -0.13]
S W/IN COS - S	1066.30	9.37×10^{-8}	-1.80	[-2.93, -0.79]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be less than zero.

Accuracy - Task switch cost analysis

The present experiment required participants to engage in a single-, dual-, or triple-task across seven different condition types. In each of the multi-task conditions, participants were required to switch from one task to another at unpredictable times. This design was used so as to require attention to be paid to multiple visual dimensions within the stimuli. As mentioned above, switching tasks leads to a reduction in accuracy and an increase in RT, generally referred to as a *switch cost* (Monsell, 2003). The switch cost can be measured in terms of a difference in performance between *switch trials*, where trial *t*-1 was a different task than the current trial *t,* to that of *non-switch trials*, where trial *t*-1 was of the same task.

There is also an effect known as the *mixing cost* where the effect of having to switch tasks within a block produces greater RT and more errors than merely performing the same task for the entire block, over and above the trial-by-trial switch cost. Both of these terms together are termed the *global switch cost* in the present report. The global switch cost was calculated for each participant by examining the difference in performance between switch trials and non-switch trials across all conditions of the experiment. The average global switch cost for accuracy across all participants was determined to be -9.99% (SD = 3.07%).

Here, the relevant hypothesis to be tested was that the decrease in accuracy from single-task conditions to multi-task conditions was greater than could be accounted for by the global switch cost alone. For each participant, their respective global switch cost was subtracted from their performance in each of the multi-task conditions (see descriptive statistics Table 24, Figure 22). This had the effect of bringing accuracy on the dual and triple-tasks conditions up by 9.99%. This method of analysis had the same statistical effect of factoring out the switch cost from the difference between the conditions, but the present approach was easier to display graphically. Analyses on the change in accuracy as the number of tasks increased were then performed on the switch cost-corrected data.

Two paired-samples *t*-tests examined the difference between average single-task accuracy and average dual-task and triple-task accuracy, after accounting for the global switch cost (Table 25). The analyses revealed that there was no reliable decrease in accuracy as the number of tasks increases, after accounting for global switch cost. Indeed, it appears that there was a slight increase in accuracy from single- to dual-task conditions (M_{dif} = -1.23%), although this difference was not statistically significant (p_{bonf} = 1.0). A Bayesian analysis (Table 26) revealed moderate evidence in favour of the null hypothesis that there was no difference in accuracy between single-task conditions and switch-cost corrected dual-task conditions (BF_{01} = 6.24).

An NHST analysis found no credible evidence that the slight decrease in accuracy from single-tasks to the triple task (M_{dif} =1.59%), was reliable, p_{bonf} = 0.144. The Bayesian analysis revealed insufficient evidence to favour either the null or alternative hypothesis, BF_{+0} = 1.44. Thus it appears that the decrease in accuracy associated with increasing the number of tasks within a condition can be accounted for by the global task switching cost alone. This finding would not support the hypothesis that decreases in accuracy associated with more tasks present within a condition would be caused by the perceptual limitations of selective attention.

Table 24.

Descriptive Statistics of Switch Cost-Corrected Accuracy Across Number of Tasks Within a Condition

Number of Tasks	*M*	*SD*	95% CI	95% BCI
1 Task	82.77	2.14	[81.33, 84.20]	[81.33, 84.20]
2 Tasks - GSC	84.00	1.72	[82.84, 85.15]	[82.84, 85.15]
3 Tasks - GSC	81.17	1.55	[80.13, 82.22]	[80.13, 82.22]

Note. *N* = 11 for all conditions. GSC = global switch cost, CI = Confidence interval (NHST), BCI = Bayesian credibility interval. Confidence intervals are constructed using within-subjects corrected variance and are not corrected for multiple comparisons (Loftus & Masson, 1994). If correcting for two comparisons, confidence intervals would be 15.36% larger.

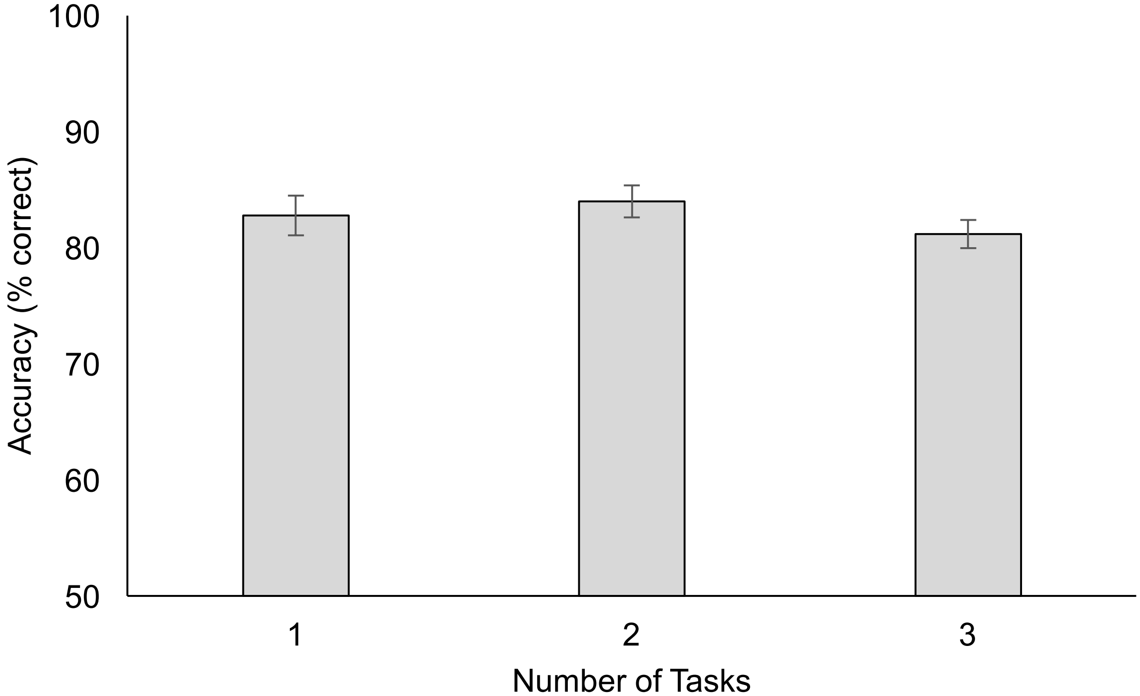

Figure 22. Switch cost-corrected mean accuracy across the number of tasks within a condition. *Note*. The global switch cost in accuracy of 9.99% has been added to the average dual-task accuracy and triple-task accuracy. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994). If correcting for two comparisons, confidence intervals would be 15.36% larger.

Table 25.

Pairwise Comparisons of Switch Cost-Corrected Accuracy Across Number of Tasks Within a Condition

Contrast	t^a	df	$p_{bonf}{}^b$	W^c	p	M_{diff}	SE_{diff}	r_{ab}	95% CI[d]	$g^{*\,e}$	95% CI
1 Task – (2 Tasks - GSC)	-1.15	10	1.00	.96	.793	-1.23	1.07	-.70	[-3.62, 1.16]	-0.55	[-1.66, 0.57]
1 Task – (3 Tasks - GSC)	1.59	10	.144	.93	.429	1.59	1.00	-.61	[-0.64, 3.83]	0.73	[-0.38, 1.84]

Note. GSC = global switch cost

[a] These are one-tailed Student's *t*-tests and assume that more dimensions lead to lower accuracy.

[b] *p*-values are corrected for two comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for two comparisons confidence intervals would be 15.36% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of g^* takes into account the correlation between the dependent samples, r_{ab}.

.

Table 26.

Bayesian Analysis of Pairwise Comparisons of Switch Cost-Corrected Accuracy Across Number of Tasks Within a Condition

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{+0} [a]	Error %	Median	95% BCI
1 Task – (2 Tasks - GSC)	0.16	1.50×10^{-4}	0.11	[0.01, 0.46]
1 Task – (3 Tasks - GSC)	1.44	5.80×10^{-5}	0.42	[0.04, 1.00]

Note. GSC = global switch cost, BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be greater than zero.

Response time - Primary analysis

Response time was analyzed as another means of assessing the effect of attending to multiple feature dimensions. It was assumed that the presence of multiple tasks within a condition would require attention to more visual dimensions, leading to increased RT. Response time was analyzed in a similar manner to accuracy, above. The primary analysis examined RT as a function of the number of tasks within a condition (i.e, the average RT of the three single task conditions versus the average RT of the dual task conditions versus the RT of the triple-task condition).

A 1 x 3 repeated measures ANOVA was conducted to examine the effect of the number of tasks in a condition on RT. The assumption of sphericity was not violated, Mauchly's $W = .74$, $p = 0.250$. The Shapiro-Wilk test showed a significant departure of normality, $W = .92$, $p = .020$. Results should therefore be interpreted with caution, and taken in conjunction with the more robust Bayesian analysis. The ANOVA showed that the number of tasks in a condition had an effect on RT, $F(2,20) = 107.5$, $p < .001$, $\eta^2 = .915$, with a tendency of increasing RT as the number of tasks increases (Table 27, Figure 23). Follow-up post-hoc Bonferroni-corrected *t*-tests confirmed this observation (Table 28), with the average RT of the dual- and triple tasks being higher than the average single-task RT. Triple-task RT was also higher than the average dual-task RT, although to a lesser extent than the single- versus dual-task comparison. There was therefore a non-linear relationship between the number of tasks present in a condition and the increase in RT, when not accounting for task-switching, as was found in Experiment 1.

A Bayesian 1 x 3 repeated measures ANOVA found similar results to the NHST analysis. The alternative hypothesis was found to be approximately 3.9 trillion times more likely than the null, $BF_{10} = 3.847 \times 10^{13}$, Error % = 1.10. Bayesian paired-samples *t*-tests (Table 29) found very strong evidence ($BF_{-0} > 60{,}000$) in favour of the hypothesis that average dual- and triple-task RTs were greater than the average single-task RT. There was also strong evidence ($BF_{-0} = 85.77$) that the triple-task condition had higher RT than the average of the dual-task conditions.

Table 27.

Descriptive Statistics of Response Time Across Number of Tasks Within a Condition

Number of Tasks	*M*	*SD*	95% CI	95% BCI
1 Task	594.30	83.69	[538.10, 650.50]	[538.10, 650.50]
2 Tasks	988.80	51.14	[954.50, 1023.20]	[954.50, 1023.20]
3 Tasks	1136.10	80.30	[1082.20, 1190.10]	[1082.20, 1190.10]

Note. $N = 11$ for all conditions. CI = Confidence interval (NHST), BCI = Bayesian credibility interval. Confidence intervals are constructed using within-subjects corrected variance and are not corrected for multiple comparisons (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

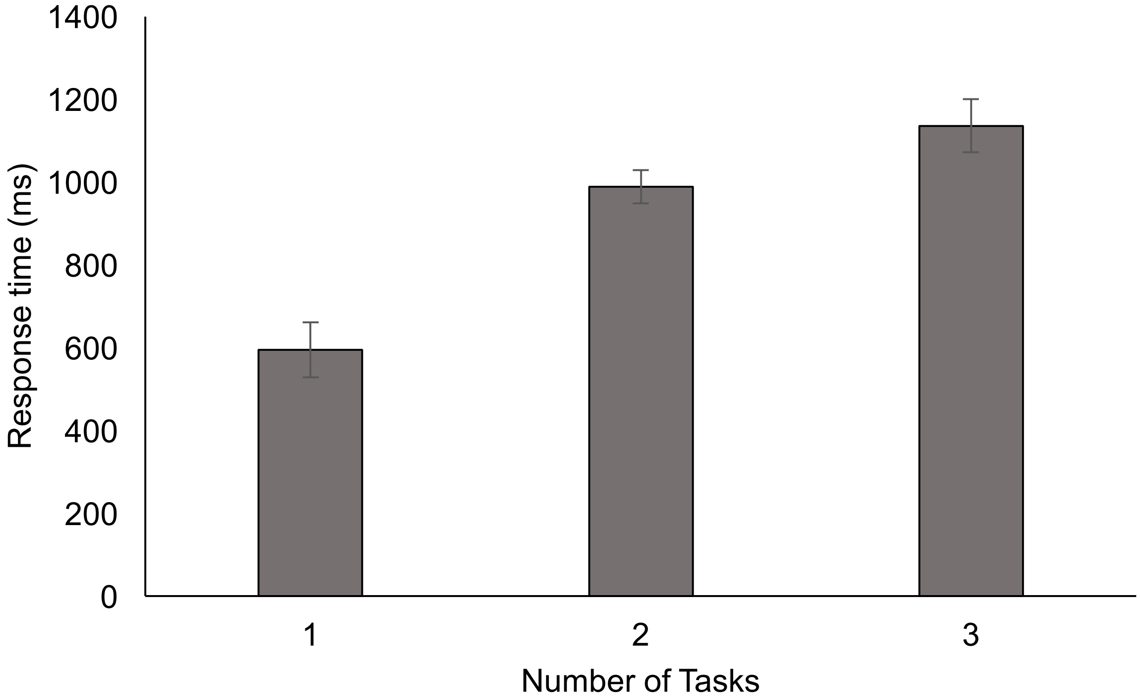

Figure 23. Mean response time across the number of tasks within a condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994). If correcting for three comparisons, confidence intervals would be 24.42% larger.

Table 28.

Pairwise Comparisons of Response Time Across Number of Tasks Within a Condition

Condition	t^a	df	p_{bonf} [b]	W^c	p	M_{diff}	SE_{diff}	r_{ab}	95% CI[d]	g^* [e]	95% CI
1 Task – 2 Tasks	-11.57	10	<.001	.79	.007	-394.50	34.10	-.37	[-318.52, -470.50]	-5.22	[2.73, 7.70]
1 Task – 3 Tasks	-11.53	10	<.001	.97	.878	-541.80	46.99	-.81	[-437.11, -646.50]	-6.06	[3.18, 8.94]
2 Tasks – 3 Tasks	-4.63	10	.001	.92	.299	-147.30	31.80	-.25	[-76.46, -218.20]	-2.01	[0.72, 3.29]

[a] These are one-tailed Student's *t*-tests and assume that more dimensions lead to higher response time.

[b] *p*-values are corrected for three comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 24.42% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of g^* takes into account the correlation between the dependent samples, r_{ab}.

Table 29.

Bayesian Analysis Pairwise Comparisons of Response Time Across Number of Tasks Within a Condition

	Bayes Factor		Effect Size (δ)	
Condition	BF_{+0} [a]	Error %	Median	95% BCI
1 Task – 2 Tasks	70470.08	~ 6.98 x 10^{-11}	3.15	[1.64, 4.89]
1 Task – 3 Tasks	68563.58	~ 5.73 x 10^{-11}	3.13	[1.60, 4.86]
2 Tasks – 3 Tasks	85.77	~ 4.62 x 10^{-7}	1.21	[0.40, 2.08]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be greater than zero.

Response time - Seven condition analysis

An analysis of RT within each condition revealed a trend consistent with the hypothesis that RT increases as the number of tasks increases. The data showed less of an effect of the unusually difficult colour task, as was found with the accuracy data, and showed fairly consistent performance as a function of the number of tasks within a condition, regardless of whether the colour task was present in the condition (Table 30, Figure 24).

Table 30.

Descriptive Statistics of Response Time Across Seven Conditions

Condition	*M*	*SD*	95% CI[a]	95% BCI
C	599.50	87.96	[540.40, 658.60]	[540.40, 658.60]
O	587.10	58.84	[547.60, 626.70]	[547.60, 626.70]
S	596.30	117.68	[517.20, 675.30]	[517.20, 675.30]
CO	957.90	100.62	[890.30, 1025.50]	[890.30, 1025.50]
CS	1018.60	87.86	[959.60, 1077.60]	[959.60, 1077.60]
OS	989.90	54.80	[953.10, 1026.70]	[953.10, 1026.70]
COS	1136.10	103.24	[1066.80, 1205.50]	[1066.80, 1205.50]

Note. $N = 11$ for all conditions. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

[a] Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) and are not adjusted for multiple comparisons.

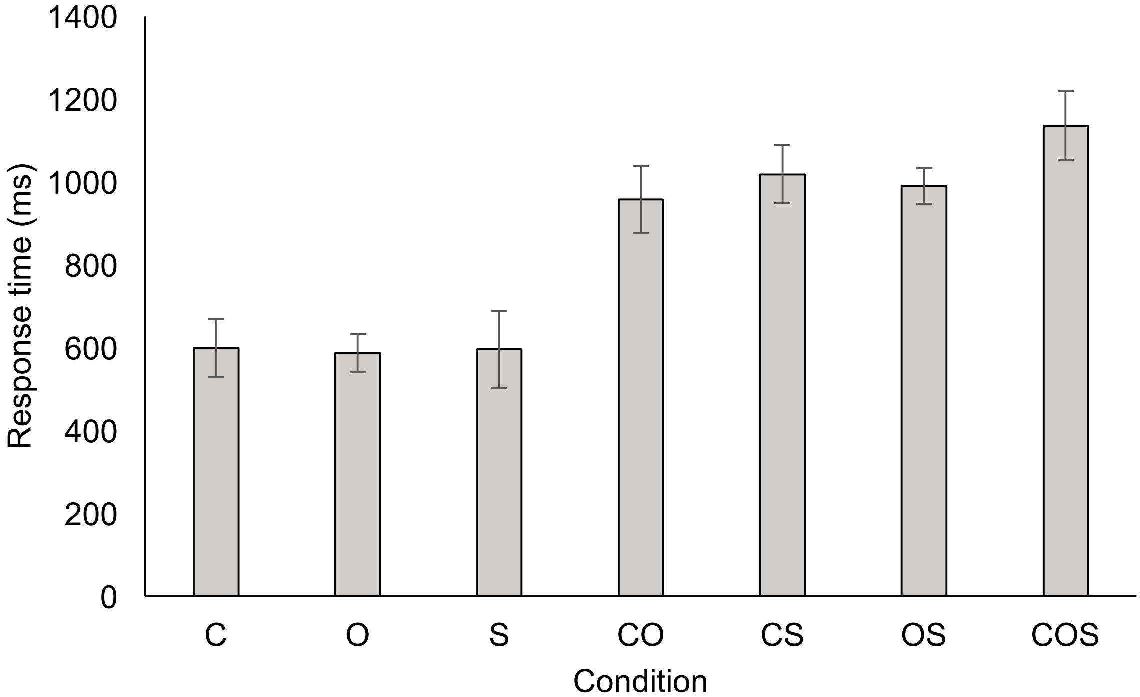

Figure 24. Response time across seven condition types.

Note. Error bars represent 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) and are not corrected for multiple comparisons. C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Response time - Task within condition analysis

Additional analysis investigating RT on each task within a condition confirmed the observation that RT increased in the presence of other tasks, and followed a more straightforward trend than the accuracy data. Mean RT for tasks within dual-task conditions tended to be consistent, with mean RT in tasks within the triple-task condition being consistently higher than the dual-tasks (see Table 31, Figures 25, 26, 27). Each task RT in the context of multi-task conditions was statistically significantly higher than its respective single-task baseline (Table 32). The Bayesian results (Table 33) all showed substantial evidence in favour of the hypothesis that task RT in multi-task conditions was greater than their single-task baselines ($BF_{+0} > 900$). Mean differences from baseline were all consistently larger when there were three tasks in the condition (M_{dif} range: 511 – 559ms), as compared to two (M_{dif} range: 318 – 422ms). The effect sizes were reflective of this trend, where Hedge's g^* ranged from 3.30 to 5.14 for dual-task versus baseline comparisons, and from 4.15 to 5.70 for triple-task versus baseline comparisons. The variability of standard deviations within each task obscures the more linear trend found in the raw difference scores. Bayesian measures of effect size showed an even less consistent pattern than the raw difference scores, where median δ ranged from 1.77 to 4.15 in dual-task versus baseline comparisons and from 2.26 to 4.20 in triple-task versus baseline comparisons. The Bayesian analyses were perhaps more sensitive to differences in variance between tasks than the NHST results. Nevertheless, the results were still consistent with the hypothesis that attending to multiple visual dimensions simultaneously is subject to the constraints of selective attention.

Table 31.

Descriptive Statistics of Task Response Time Within Each Condition

Task Within Condition	*M*	*SD*	95% CI	95% BCI
C	599.50	96.50	[534.70, 664.30]	[534.70, 664.30]
C W/IN CO	1010.90	103.21	[941.60, 1080.30]	[941.60, 1080.30]
C W/IN CS	1022.20	93.22	[959.50, 1084.80]	[959.50, 1084.80]
C W/IN COS	1154.70	90.61	[1093.80, 1215.50]	[1093.80, 1215.50]
O	587.10	75.08	[536.70, 637.60]	[536.70, 637.60]
O W/IN CO	905.00	66.89	[860.00, 949.90]	[860.00, 949.90]
O W/IN OS	983.40	66.08	[939.00, 1027.80]	[939.00, 1027.80]
O W/IN COS	1098.30	88.79	[1038.70, 1158.00]	[1038.70, 1158.00]
S	596.30	135.96	[504.90, 687.60]	[504.90, 687.60]
S W/IN CS	1015.00	92.34	[952.90, 1077.00]	[952.90, 1077.00]
S W/IN OS	996.50	49.88	[963.00, 1030.00]	[963.00, 1030.00]
S W/IN COS	1155.40	110.03	[1081.50, 1229.30]	[1081.50, 1229.30]

Note. $N = 11$ for each condition. Confidence intervals are constructed using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of each condition type (colour, orientation, or size), and are not corrected for multiple comparisons. If correcting for nine comparisons, confidence intervals would be 49.57% larger. W/IN = within, C = colour, O = orientation, S = size, CO = colour and orientation, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size, CI = Confidence interval (NHST), BCI = Bayesian credibility interval.

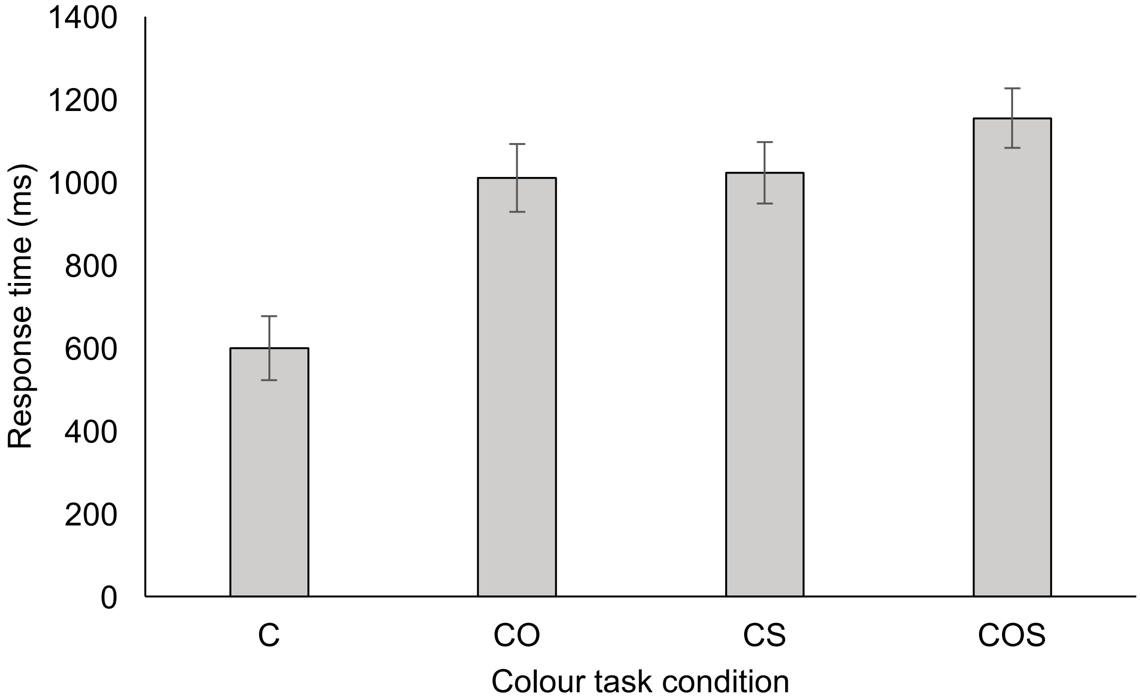

Figure 25. Response time on the colour task within each colour-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of colour task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. C = colour, CO = colour and orientation, CS = colour and size, COS = colour, orientation, and size.

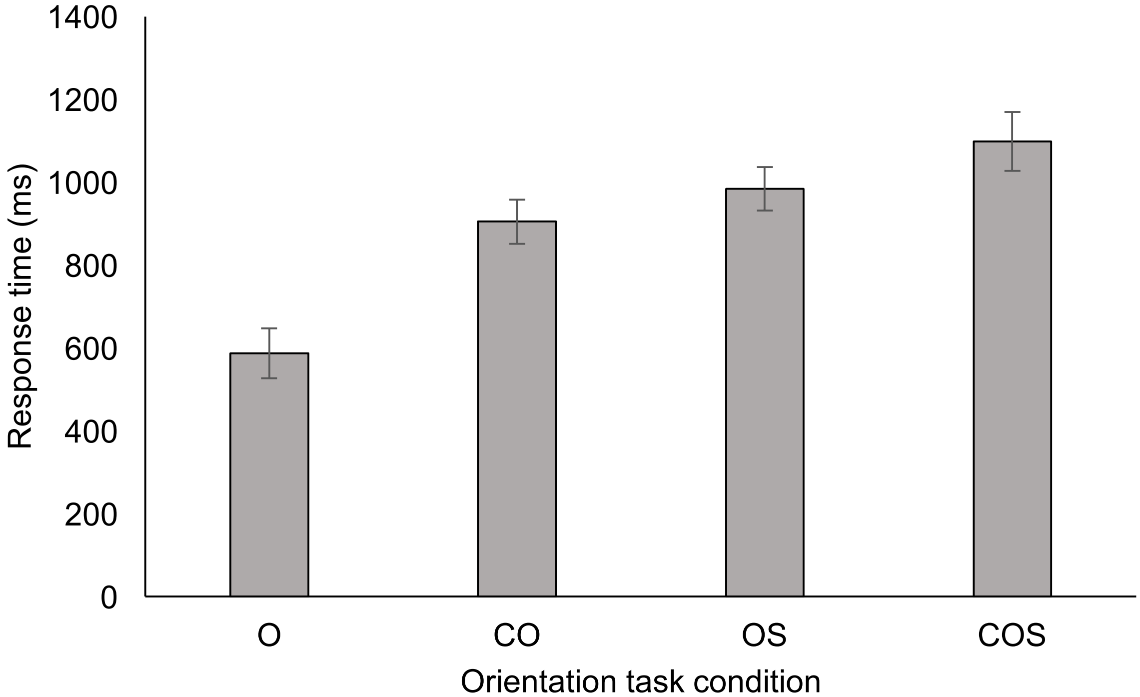

Figure 26. Response time on the orientation task within each orientation-containing condition. *Note*. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of orientation task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. O = orientation, CO = colour and orientation, OS = orientation and size, COS = colour, orientation, and size.

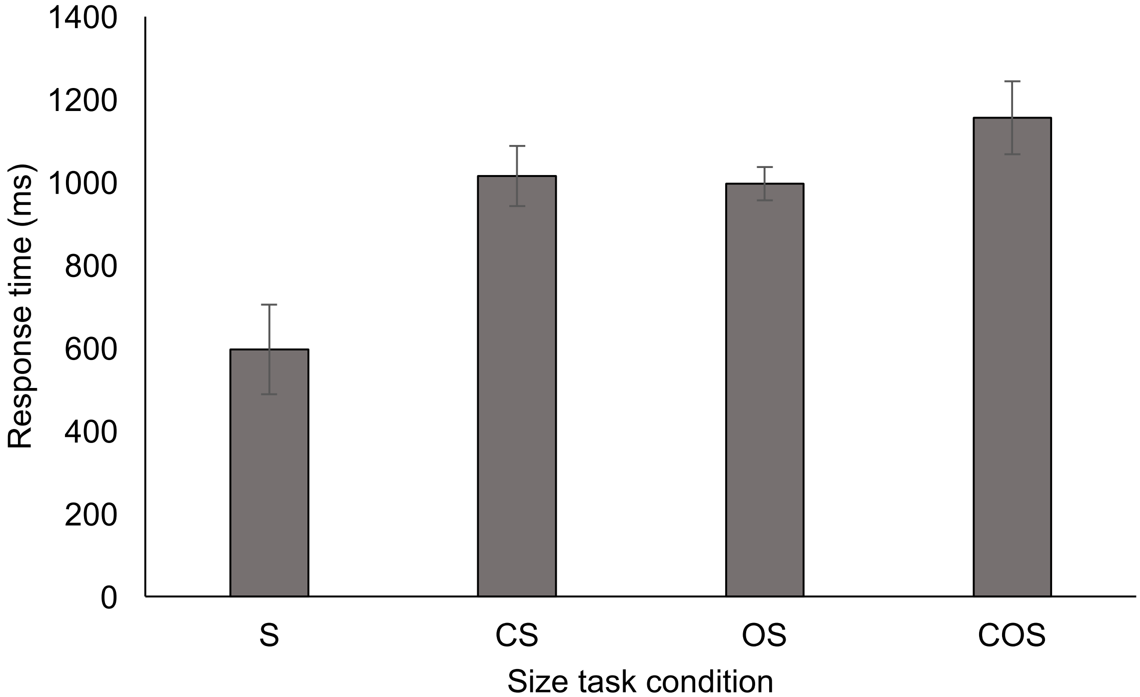

Figure 27. Response time on the size task within each size-containing condition.

Note. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994) using the grand mean of size task performance across four conditions. If correcting for nine comparisons, confidence intervals would be 49.57% larger. S = size, CS = colour and size, OS = orientation and size, COS = colour, orientation, and size.

Table 32.

Response Time Difference from Baseline for Multi-Task Conditions

Contrast	t^{a}	*df*	$p_{bonf}{}^{b}$	W^{c}	*p*	M_{diff}	SE_{diff}	r_{ab}	95% CI[d]	$g^{*\,e}$	95% CI[e]
C W/IN CO - C	7.23	10	<.001	.95	.595	411.40	56.91	-.79	[284.60, 538.20]	3.78	[1.78, 5.77]
C W/IN CS - C	9.06	10	<.001	.94	.553	422.70	46.65	-.33	[318.70, 526.60]	4.09	[2.05, 6.12]
C W/IN COS - C	15.26	10	<.001	.90	.167	555.20	36.37	.17	[474.10, 636.20]	5.44	[2.94, 7.94]
O W/IN CO - O	9.71	10	<.001	.99	.994	317.80	32.74	-.17	[244.90, 390.80]	4.10	[2.09, 6.11]
O W/IN OS - O	15.15	10	<.001	.94	.569	396.20	26.16	.25	[338.00, 454.50]	5.14	[2.77, 7.50]
O W/IN COS - O	10.60	10	<.001	.95	.578	511.20	48.23	-.91	[403.70, 618.70]	5.70	[2.95, 8.45]
S W/IN CS - S	6.67	10	<.001	.91	.267	418.70	62.74	-.65	[278.90, 558.50]	3.30	[1.50, 5.11]
S W/IN OS - S	8.83	10	<.001	.88	.110	400.20	45.30	-.12	[299.30, 501.10]	3.58	[2.94, 8.02][f]
S W/IN COS - S	8.32	10	<.001	.95	.609	559.10	67.19	-.64	[409.40, 708.80]	4.15	[2.04, 6.25]

[a] These are one-tailed Student's *t*-tests and assume that response times would be higher in multi-task conditions.

[b] *p*-values are corrected for nine comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for three comparisons confidence intervals would be 49.57% larger.

[e]. The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of g^* takes into account the correlation between the dependent samples, r_{ab}.

[f] Difference between *SD*s was significant ($p < .05$); lower and upper bounds of the 95% *CI* represents effect size (Glass' Δ) based on treatment group (multi-task conditions) *SD*, and control group (single-task condition) *SD*, respectively.

Table 33.

Bayesian Analysis of Response Time Difference from Baseline for Multi-Task Conditions

	Bayes Factor		Effect Size (δ)	
Contrast	BF_{+0}[a]	Error %	Median	95% BCI
C W/IN CO - C	1715.90	1.06×10^{-8}	1.92	[0.83, 3.13]
C W/IN CS - C	9704.50	1.49×10^{-9}	2.43	[1.19, 3.82]
C W/IN COS - C	733052.00	9.33×10^{-10}	4.20	[2.28, 6.41]
O W/IN CO - O	16809.20	2.73×10^{-9}	2.63	[1.32, 4.14]
O W/IN OS - O	687437.90	3.31×10^{-10}	4.15	[2.23, 6.37]
O W/IN COS - O	34246.70	3.73×10^{-9}	2.90	[1.47, 4.47]
S W/IN CS - S	956.50	9.42×10^{-8}	1.77	[0.77, 2.91]
S W/IN OS - S	7945.80	1.18×10^{-9}	2.38	[1.13, 3.71]
S W/IN COS - S	4981.90	2.47×10^{-10}	2.26	[1.08, 3.56]

Note. BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be greater than zero.

Response time - Task switch cost analysis

As with the accuracy data, the fact that RT increased with the number of tasks within a condition might be explained in terms of the cost of switching tasks alone (Monsell, 2003), and not the attentional cost of attending to multiple feature dimensions. In order to examine whether the increase in RT could be accounted for through task-switching alone, the global switch cost for RT was calculated for each participant in a manner similar to that of accuracy. The average global switch cost for RT across participants was determined to be 350.68ms (SD = 95.67). Each participant's global switch cost was subtracted from the average RT for dual-tasks and the triple task (see Table 34 for descriptive statistics, and Figure 28 for a graphical representation). Dependent samples t-tests (Table 35) then examined whether the increase in RT associated with having more tasks within a condition could be accounted for entirely due to the effect of task switching. The results showed that average RT for dual-task conditions was marginally statistically significantly higher than the average single-task RT, $p_{bonf} = .091$, with a relatively large effect size in favour of the hypothesis ($M_{dif} = -43.82$, $g^* = 0.81$). This finding suggests that there may have been a meaningful difference from baseline for dual-tasks, but the difference may have been obscured by a lack of power. There appears to have been a more robust difference between average RT in single-task versus triple-task conditions, with triple-task RT being statistically significantly higher than average single-task RT, $p_{bonf} < .001$, with a substantially larger effect size ($M_{dif} = -191.13$, $g^* = 3.18$). Thus it appears that the effect of higher RT as a result of more tasks within a condition cannot be entirely accounted for by the effect of task switching, providing evidence in favour of the conclusion that there is an attentional cost associated with perceiving multiple visual dimensions simultaneously over and above the cost of task switching.

Table 34.

Descriptive Statistics of Response Time Across Number of Tasks Within a Condition.

Number of Tasks	*M*	*SD*	95% CI	95% BCI
1 Task	594.30	46.80	[562.90, 625.80]	[562.90, 625.80]
2 Tasks - GSC	638.10	52.68	[602.70, 673.50]	[602.70, 673.50]
3 Tasks - GSC	785.40	62.30	[743.60, 827.30]	[743.60, 827.30]

Note. $N = 11$ for all conditions. GSC = global switch cost, CI = Confidence interval (NHST), BCI = Bayesian credibility interval. Confidence intervals are constructed using within-subjects corrected variance and are not corrected for multiple comparisons (Loftus & Masson, 1994). If correcting for two comparisons, confidence intervals would be 15.36% larger.

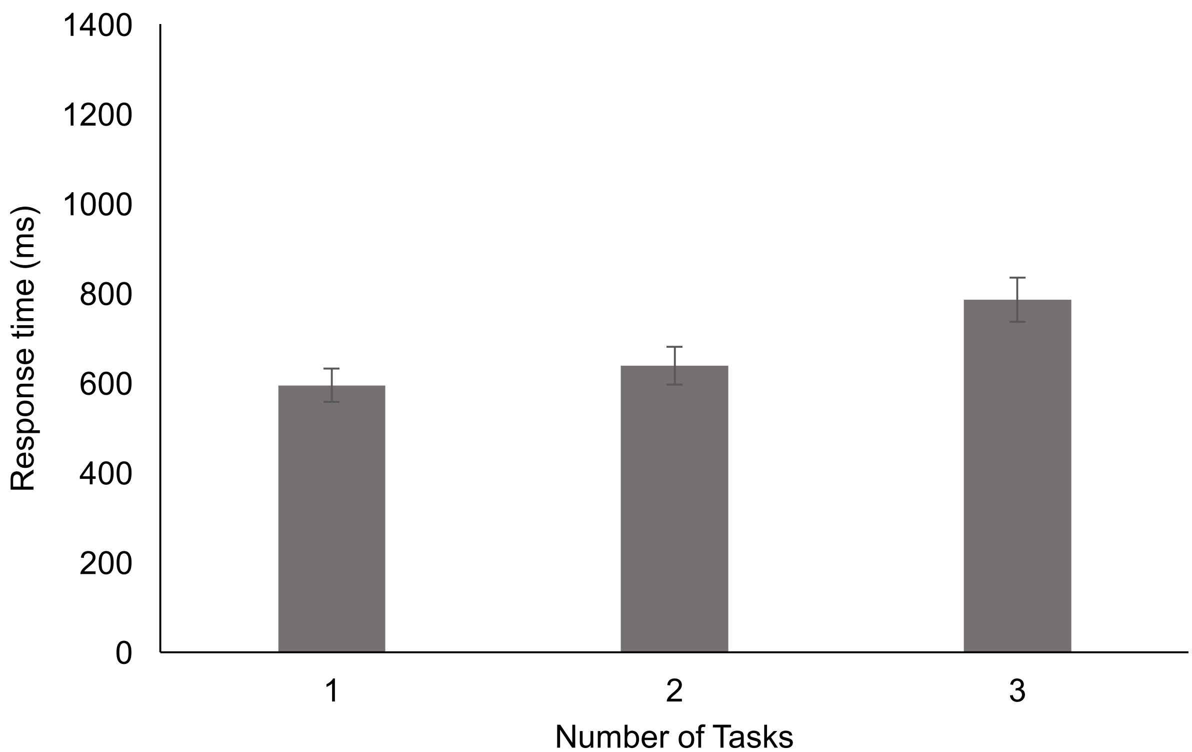

Figure 28. Switch cost-corrected mean response time (RT) across the number of tasks within a condition.

Note. The global switch cost in RT of 350.68ms has been added to the average dual-task RT and triple-task RT. Error bars represent uncorrected 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994). If correcting for two comparisons, confidence intervals would be 15.36% larger.

Table 35.

Pairwise Comparisons of Switch Cost-Corrected Response Time Across Number of Tasks Within a Condition

Contrast	t^{a}	*df*	p_{bonf}^{b}	W^{c}	*p*	M_{diff}	SE_{diff}	r_{ab}	95% CI[d]	g^{*} [e]	95% CI
1 Task – (2 Tasks - GSC)	-1.87	10	.091	.97	.883	-43.82	23.45	-.22	[-96.08, 8.43]	-0.81	[-1.80, 0.18]
1 Task – (3 Tasks - GSC)	-6.55	10	< .001	.85	.043	-191.13	29.18	-.57	[-256.16, -126.11]	-3.18	[-4.92, -1.44]

Note. GSC = global switch cost.

[a] These are one-tailed Student's *t*-tests and assume that more dimensions lead to higher response time.

[b] *p*-values are corrected for three comparisons using Bonferroni adjustment.

[c] *W* = Shapiro-Wilke test of normality; significant results suggests a departure from normality.

[d] The 95% confidence interval represent a non-directional, non-corrected estimate around the mean difference. If using Bonferroni correction for two comparisons confidence intervals would be 15.36% larger.

[e] The 95% confidence interval is approximate and based on the *z*-distribution. The standard error of g^{*} takes into account the correlation between the dependent samples, r_{ab}.

The results of the analogous Bayesian analyses (Table 36) were largely in agreement with the NHST results. The comparison between single- and switch cost-corrected dual-task RT failed to find conclusive evidence in favour of either the null or alternative hypotheses (BF_{-0} = 2.053), although a small to moderate effect size in favour of the alternative hypothesis was still found, median δ = 0.48. The analysis found strong evidence (BF_{-0} = 836.38) that switch-cost corrected triple-task RT was higher than average single-task RT, with a large effect size, median δ = 1.74. These analyses therefore provide converging evidence that the effect of increasing RT as a result of the number of tasks in a condition was not entirely due to the cost of task switching, and lends evidence in favour of the hypothesis that the increase in RT also results from the perceptual cost of attending to multiple visual dimensions.

Table 36.

Bayesian Analysis of Comparisons of Switch Cost-Corrected Response Time Across Three Conditions

	Bayes Factor		Effect size (δ)	
Contrast	BF_{-0}	Error %	Median	95% BCI
1 Task – (2 Tasks - GSC)	2.05	~ 3.24 x 10 $^{-5}$	-0.48	[-1.07, -0.05]
1 Task – (3 Tasks - GSC)	836.38	~ 7.15 x 10 $^{-8}$	-1.74	[-2.83, -0.74]

Note. GSC = global switch cost, BCI = Bayesian credibility interval.

[a] Bayes Factors (*BF*) are directional and assume the difference would be less than zero.

Discussion – Experiment 2

The experiment sought to show that attending to multiple visual dimensions simultaneously under conditions of constant and difficult stimulus exposure duration produces a cost associated with selective attention. The experiment was designed to tax the attentional system so as to elicit lower accuracy and greater RT while attending to an increased number of visual dimensions. It was assumed that requiring participants to respond to more types of features within a block would force them to attend to more feature dimensions in the stimuli. However, requiring participants to respond to different queries within a block also created a potential for a task-switching cost, which may be able to account for some of the observed decrements in performance.

The experiment provided partial support for the hypothesis that perceiving multiple visual dimensions induces a cost of selective attention. When the cost of task-switching is not taken into account, there is clear evidence that accuracy decreases as the number of tasks increases, consistent with the hypothesis that attending to multiple visual features under fixed exposure duration incurs a cost of selective attention. Similarly, RT increases as the number of tasks within a condition increases, indicating greater task difficulty. However, the effect of task switching alone may be able to account for all of the observed decreases in accuracy in Experiment 2. Thus, the evidence from the accuracy data does not clearly support the hypothesis of the study, although it is worth highlighting that the switch cost estimated in Experiment 2 is likely an overestimation of its true effect on participant behaviour (discussed further below).

Response time data showed decrements in performance that are not entirely due to task switching, given that the increase in RT associated with completing multiple tasks within a condition was greater than that accounted for by the global switch cost. There is therefore some evidence that the requirement to attend to multiple dimensions simultaneously has an attentional cost over and above that of a task switching cost. These results are consistent with the predictions of Boolean map theory and inconsistent with a broad interpretation of the claim that forming multiple between-dimension summaries is an unlimited capacity process, as Attarha and Moore (2015b) had claimed. It also does not appear likely that forming summary representations is an "intention free" process that "precedes the limited capacity bottleneck that forces selective attention" (Chong & Treisman, 2005, p 899). To the contrary, it appears that forming summary representations, especially across multiple dimensions, is an attentionally demanding and limited process, at least under the constraints of Experiment 2.

The main constraints of the experiment were the fixed stimulus presentation duration of 200ms for each task and the perceptual difficulty of the single-task conditions, achieved through the thresholding procedure preceding the main experiment. The fixed stimulus presentation duration was used to test whether stimulus exposure duration is a critical variable in forming set summary representations across multiple visual dimensions, as Boolean map theory would predict. Attarha and Moore (2015b) did not control this variable in their experiments, which likely biased their results in favour of finding no difference between processing a single visual dimension versus multiple visual dimensions.

The thresholding procedure of Experiment 2 was aimed at avoiding ceiling and floor effects as well as equalizing the difficulty of each task both within and between participants by determining the stimulus intensity eliciting 85% correct performance on each task. Although the current iteration of the experiment was not entirely successful in achieving these aims, with single-task colour accuracy being near 70%, the results still point in the direction of there being an attentional limit to perceiving summary statistics across multiple visual dimensions. While it is still unclear why it was so difficult to estimate the true 85% accuracy threshold for the colour task, it may be that the true slope of the psychometric function for colour is quite steep. If this were the case, small inaccuracies in estimating the 85% threshold resulted in large differences in performance. Further refinements to the colour thresholding procedure should be undertaken to address this potential confound in the design (discussed further below).

An additional confound of the present study is the cost of task-switching. As participants were instructed to attend to an increasing number of visual dimensions to complete the tasks in each multi-task condition, they also were required to switch between tasks. This might have incurred a cognitive cost due to the change in task, resulting in a decrease in accuracy and an increase in RTs. While also potentially present in Experiment 1, the effect of task switching was estimated in Experiment 2, and showed that it can account for all of the decrease in accuracy and much of the increase in RT associated with increasing the number of tasks in each condition. However, it should be noted that the estimate of the global switch cost found in Experiment 2 is likely an overestimation. This is due to the fact that the switch cost is systematically related to the increase in attentional/perceptual difficulty of representing multiple visual features simultaneously. Factoring out all accuracy and response time information that can be attributed to task switching will also factor out information related to perceptual/attentional difficulty. The result is an over-inflated estimation of global task switching cost, which leads to anomalous findings. For example, after the global task switch cost for accuracy had been factored into the results of Experiment 2, it appears that accuracy *increased* slightly when participants were performing a dual-task versus a single task. It seems unlikely that the dual-task actually facilitated greater accuracy, but rather that the true task-switching cost was estimated inaccurately. A design which could estimate the cost of switching tasks independently of the effect of attending to multiple visual dimensions would likely find a global switch cost much less than the current design was able to.

General Discussion

The present study sought to explore the role of selective attention in the perception of set summary statistics across multiple visual dimensions. Previous researchers have claimed that the representation of set summary statistics is a process that occurs “without intention”, “in parallel”, and “precedes the limited capacity bottleneck that forces selective attention” (Chong & Treisman, 2005, p. 899). Further, the findings of other researchers (Attarha & Moore, 2015b) suggest that there may not be an attentional cost of forming set summary representations across multiple visual dimensions. Attarha and Moore (2015b) claimed that the representation of summary statistics between different visual dimensions an “unlimited capacity process” (Attarha & Moore, 2015b, p. 2). If interpreted narrowly, this claim suggests that the extraction of summary statistics is a process that unfolds equally efficiently, regardless of the number of elements that must be averaged over (i.e., it is set-size independent). If interpreted more broadly, the claim suggests that there is no limitation of selective attention when forming set summary representations across multiple visual dimensions. The current thesis did not test the narrow claim about set-size independence, and this claim is not in conflict with Boolean map theory (Huang & Pashler, 2007). However, the current thesis does shed light on the broad claim about selective attention. The results appear to show a cost of attending to multiple visual dimensions when forming representations of set summary statistics under the constraint of a fixed stimulus exposure duration. Accuracy tends to decrease and RT tends to increase as more visual dimensions are attended to. These results would suggest that there is a limitation of selective attention in perceiving multiple visual dimensions simultaneously.

According to Boolean map theory (Huang & Pashler, 2007), the attentional process of selection is responsible for combining information across multiple visual dimensions through the processes of union or intersection. These computational processes must unfold serially and require a finite amount of time to complete. The theory would predict that stimulus presentation duration is a critical variable that allows for the construction of representations of multiple dimensions. This prediction has been borne out by experiments that show that single-dimension visual search (feature search) elicits greater accuracy than two-dimension visual search (conjunction search) under short exposure durations (100 or 200ms), but conjunction search elicits greater accuracy under longer exposure durations (400ms). The results of the present experiment appear to be consistent with the prediction that a fixed and difficult stimulus exposure duration would interfere with selective attention's ability to compute the relevant representations across multiple dimensions.

Attarha and Moore (2015b) had found that observers were able to form summary representations simultaneously across the dimensions of orientation and size with little to no cost. They found that performance on the size and orientation single-tasks was comparable to that of the orientation and size dual-task, in that sequential presentation of half of the stimuli did not result in greater accuracy than simultaneous presentation of all of the stimuli. Further, a sensitivity analysis of the simultaneous conditions of their orientation, size, and dual tasks, reveals a *d'* equal to 0.95, 1.05, and 0.99, respectively. However, in an effort to prevent floor and ceiling effects, the stimulus presentation durations differed between their three tasks, with average presentation durations of 60, 90, and 230ms, for the orientation, size, and dual-tasks, respectively.

The fixed presentation time of the experiments in this thesis was aimed at testing whether the lack of difference between single- and dual-task performance found by Attarha and Moore (2015b) could be accounted for by the different stimulus presentation durations between the single- and dual-tasks of their experiments. Experiment 1 showed an effect consistent with the hypothesis that stimulus exposure duration may be a critical factor in forming summary representations across multiple dimensions, as would be predicted by Boolean map theory. However, Experiment 1 suffered from the shortcoming of a ceiling effect in the colour task, which makes the results difficult to interpret. Experiment 2 attempted to bring the accuracy on the three single-tasks to near 85%. The thresholding procedure achieved its aim with the orientation and size tasks, but the colour single-task accuracy was near 70%. Nevertheless, a similar pattern emerged showing that accuracy decreases and RT increases as more dimensions are attended to in order to complete the tasks. The results appear to demonstrate a constraint on the ability to form set summary representations across multiple dimensions. At least one constraint is that there must be sufficient stimulus exposure duration in order for the relevant computations combining information across multiple dimensions to be performed.

Perceptual difficulty

The perceptual difficulty of the task may also be a critical variable to consider when discussing the role of attention in forming summary representations across even a single visual dimension. Both experiments of the current thesis attempted to avoid ceiling and floor effects by controlling the perceptual discriminability of the two modes within each visual dimension. Experiment 1 used pilot experiments to coarsely estimate the average 85% threshold across size and orientation dimensions, but did not control for colour discriminability. It was found that the addition of the colour task in multi-task conditions in Experiment 1 did not decrease performance much in comparison to the addition of the orientation or size tasks. Within performance on the colour task alone, accuracy decreased by ~5% and RT increased by ~305ms in the triple-task condition, whereas in the orientation and size tasks, the decrease in accuracy was ~10% and 12%, respectively, and the increase in RT was 424 and 383ms, respectively. While not explicitly calculated in Experiment 1, the cost of task switching may be able to explain some of the decrease in performance observed for the relatively easy colour task, but it appears that the performance decrements were diminished under the relatively easy perceptual demands of Experiment 1's colour task.

It may be that like the 'pop-out' effect observed in visual search, there is a threshold of perceptual discriminability that must be achieved before set-size independent summary representations are elicited. In visual search, a spectrum of search efficiencies are elicited by varying degrees of perceptual salience of a target singleton compared to the distractors (Wolfe & Horowitz, 2004). A similar phenomenon may be at work for summary statistics whereby summaries can be formed efficiently and effortlessly under easy perceptual conditions, but may be inefficient under conditions of perceptual difficulty. Perceptual difficulty of the task might be an additional constraint on the claim that summary statistics are done 'without intention' and 'prior to the bottleneck of selective attention'. This claim would need to be tested in future experiments.

Task switching

A confound of both experiments, although only explicitly discussed in Experiment 2, is the effect of task switching (Monsell, 2003). In Experiment 2, the difference in performance between switch trials and non-switch trials (i.e. the global switch cost) was determined to be approximately a 10% decrease in accuracy and a 350ms increase in RT. When the switch cost is factored into the analysis, the decrease in accuracy associated with having more tasks within a condition can be accounted for in terms of switch cost alone. An effect still remains for RT even after the switch cost is taken into account.

Given that the perceptual difficulty associated with the task of attending to multiple dimensions of visual stimuli was systematically related to the effect of the global switch cost, factoring in the total effect of the switch cost likely removes much of the effect associated with the perceptual difficulty of the task. Recall that the global switch cost was calculated by comparing the task performance on all non-switch trials to all switch trials. Every trial of the three single-task conditions is by definition a non-switch trial, so the inclusion of these relatively perceptually easy trials in the estimate of total non-switch trial performance gives a heavy weight to the easiest conditions in the estimation of non-switch trial performance.

The problem can be conceptualized as an issue of shared variance. Both the switch cost and perceptual difficulty due to attentional load can explain some of the variance in the decrease of performance associated with the number of tasks within a condition. By factoring out all of the variance that can be accounted for by the switch cost, the remaining unique variance associated with the perceptual difficulty of the task explains relatively little of the total. A better study design would estimate the effect of task switching independently of the perceptual difficulty of the task, so that a more accurate estimate of its unique effect could be made.

It should be noted that the task-switching cost is not a unique feature of the present study; many paradigms intended to manipulate attention include an element of task switching, yet its effect is rarely calculated or discussed. The design of the present study's experiments is akin to that of a pre-/post-cueing paradigm. In a standard pre-/post-cueing design (e.g., Chong & Treisman, 2005; Emmanouil & Treisman, 2008; Huang, 2015), a pre-cue is given on some trials where the task (e.g. feature to be reported) is specified prior to the presentation of the stimulus so that participants can selectively attend to this aspect of the stimulus. On post-cue trials, the task is specified after the stimulus has been presented, so selective attention cannot be systematically deployed to the task under investigation. The difference in performance between pre-cue and post-cue trials is taken as an indication of the effect of attention on task performance. In the present experiment, cues were given at the start of a block, so that participants could selectively attend to the relevant features of the set during each trial of the block. The single-task conditions are analogous to a pre-cue, where the participant can anticipate what feature will be queried. The multi-task conditions are analogous to a post-cue condition of a standard design, where there is some uncertainty in which feature will be queried.

Emmanouil and Treisman's (2008) study is an example of a standard pre-/post-cueing paradigm, where they studied the effect of pre- versus post-cueing on the ability of participants to report mean size, speed of motion, or orientation across a set of stimuli that varied across two of these dimensions. In the post-cue blocks, participants had to switch from reporting the mean size to the mean speed of motion, or mean orientation, at unexpected times. Like the experiments in this thesis, their tasks would have incurred a switch cost. However, the effect of the switch cost was not analyzed or discussed in their study. It is therefore possible that the differences in performance they found between pre-cue and post-cue blocks might be (at least partly) explained by the presence of a task-switch cost, and not the attention manipulation of their study.

This criticism applies to a number of the previous studies reviewed in this article (Chong & Treisman, 2005; Emmanouil & Treisman, 2008; Huang, 2015). It is therefore unclear whether observed differences or non-differences found in these studies are accounted for by a task-switch cost, or the attention manipulation. One notable exception is Attarha and Moore's (2015b) study, which used a 2-AFC / 4-AFC simultaneous/sequential design which would not have had a task-switching component.

Working memory and attention

A conceptual issue with the present study is determining whether the performance decrements associated with multi-task interference are related to limitations of visual attention or visual working memory. Attention is commonly defined as a process which selects some information to be processed over other competing information in the environment. Working memory is a process which encodes, stores, and manipulates information in a highly accessible state (Fougnie, 2008). The present experiments required that participants selectively attend to some of the information present in the stimuli (e.g. the average colour and size), sometimes to the exclusion of other features in the stimuli (e.g., the average size). Participants also had to encode this information to report the majority colour, orientation, or size when prompted. Recall that on each trial, a perceptual mask was present for 600ms between the offset of the stimulus and the question prompt. Some representation of the stimulus must therefore have been maintained in a short-term store for at least 600ms in order to complete the task. As the number of tasks within a condition increased from one to two to three, the number of features to both attend to and encode also increased.

While the conceptual, behavioural, and neural interrelations between attention and working memory are beyond the scope of this report, it bears mentioning that the two concepts are distinct, but share many behavioural manifestations and neural substrates (Feng, Pratt, & Spence, 2012; Fusser et al., 2011; Kane, Conway, Hambrick, D.Z., & Engle, 2007; Mayer et al., 2007). Boolean map theory is primarily a theory about visual attention, but it does consider the process of creating Boolean maps through the operations of union and intersection to be a type of working memory with limited capacity (Huang & Pashler, 2007). That is to say, there is no clear distinction in the theory between attention and working memory in representing the visual world, as they are assumed to be working in tandem. The key point of Boolean map theory is in defining the type of data structure the visual system uses to represent the world (Boolean maps) and the processes that operate on them (e.g., the union and intersection operations, among others).

For the purposes of the present discussion, it can be granted that there is a component of working memory being tasked in the present experiments in addition to attention. Many visual attention-related tasks also involve some component of working memory, especially those involving dual-task methodologies. For example, the work of Attarha and Moore (2015b) required participants to explicitly encode both the mean size and mean orientation of a set of Gabors in their 4-AFC dual-task. Emmanouil and Treisman (2008) also required two separate representations of mean size and orientation (or size and speed) to be represented in the post-cue condition, (as opposed to only one in the pre-cue condition). It should be noted that most accounts of visual working memory contend that the average human observer can hold between 3 and 5 ‘chunks’ of information in memory (Feng et al., 2012; Mayer et al., 2007), so even the triple-task of the present experiments likely did not exceed the working memory capacity of the average participant. The primary limiting factor of working memory encoding in the present study may be the relatively short stimulus exposure duration of 200ms (Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011). Future studies can examine this constraint by varying the stimulus exposure duration to determine the upper limit of performance. If sufficiently long exposure durations (e.g. > 2s) produce near-perfect performance, it would be an indication that encoding the three visual dimensions is not beyond the capacity of visual working memory.

Future studies may be able to tease apart the relative contributions of working memory and attention in the ability to represent multiple-dimension summaries. For example, a task similar to Experiments 1 and 2 of the current study, but with very easy perceptual discriminability might be able to estimate the effect of working memory absent the constraints on attention related to perceptual difficulty. It is worth noting that Experiment 1 used a colour task with very little perceptual difficulty, and the decreases in performance within the colour task were much less compared to those in the more difficult size and orientation tasks when multiple tasks were present within a condition. A future experiment could determine whether there is an upper limit to the ability to encode the three features in working memory, independent of the effect of perceptual discriminability.

Future directions

Future studies may be able to overcome some of the shortcomings of the current thesis. Firstly, a more precise method of estimating the 85% colour threshold should be sought. A different procedure such as the psi-marginal adaptive method (Prins, 2013) could be used for this purpose. Alternatively, the proper solution may simply be to collect more data from each participant during the colour thresholding portion of the experiment. An even more complex design which would systematically estimate the thresholds for colour, orientation, and size simultaneously would also be possible, although perhaps unnecessarily complicated for the purposes of keeping performance within a reasonable range.

Second, a more precise estimate of the true effect of task switching would allow for clearer conclusions to be drawn regarding the role of attention related to the simultaneous perception of multiple visual dimensions. One way to accomplish this would be to implement a design similar to Experiment 2, but with easy perceptual discriminability on each task. This might allow for the estimation of global switch cost independent of the effect of perceptual discriminability. This estimate of the global switch cost could then be factored into a replication of Experiment 2 with properly thresholded perceptual discriminability.

Likewise, this sort of easy/hard discriminability design might be able to shed light on the relative role of working memory in performance decrements. Given that the same working memory constraints would be involved regardless of the perceptual discriminability of the task, differences in performance between the easy and hard task would likely reflect constraints on perceptual attention, and not working memory.

Another design perhaps more pertinent to the question of whether stimulus exposure duration poses a constraint on the ability to perceive multiple visual dimensions simultaneously would be to systematically vary stimulus exposure durations and measure the relevant thresholds of each condition. This design could be combined with an easy/hard discriminability condition so that the effect of task switching and working memory encoding on stimulus exposure duration thresholds could be measured independently of the effect of stimulus intensity.

Further, the effect of task switching could be avoided altogether by using a 2AFC/4AFC/8AFC design for the single-, dual- and triple-tasks, respectively. This design would be more akin to Attarha and Moore's (2015b) approach, which did not suffer from the same kind of task-switching concerns of the present thesis or other past research. It should be noted that this design would not overcome the working memory confound discussed above, as the number of features to be explicitly encoded would increase with the number of dimensions to attend to. Perhaps this design in combination with the easy/hard discriminability conditions would be the best solution to solve both issues.

Conclusions

The present study sought to examine the role of selective attention in the perception of set summary statistics across multiple visual dimensions. According to Boolean map theory (Huang & Pashler, 2007), the perception of multiple visual dimensions is set-size independent, but the process requires a series of computational steps which must be performed serially. Stimulus exposure duration is therefore a limiting factor in the ability to form summary representations across multiple visual dimensions, according to the theory. The current study found evidence to support the conclusion that a fixed and difficult stimulus exposure duration causes interference in the ability to form set summary representations across multiple visual dimensions. The study points to the need for careful control of stimulus properties and procedures to ensure the phenomenon under question is being isolated. In particular, the effects of task switching, and working memory load should be taken into account as potential confounding factors in attention-related studies. Future studies should estimate these particular costs independently of the attention/perceptual manipulation. Further, future research should examine more systematically the role of stimulus exposure duration on the perception of multiple visual dimensions simultaneously to shed light both on Boolean map theory and the phenomena of set summary representations more generally.

References

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, *12*(2), 157–162. http://doi.org/10.1111/1467-9280.00327

Attarha, M., & Moore, C. M. (2015a). The capacity limitations of orientation summary statistics. *Attention, Perception & Psychophysics*, (March), 1–16. http://doi.org/10.3758/s13414-015-0870-0

Attarha, M., & Moore, C. M. (2015b). The perceptual processing capacity of summary statistics between and within feature dimensions. *Journal of Vision*, *15*(4), 9. http://doi.org/10.1167/15.4.9

Bach, M. (2007). The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*(7), 965–971. http://doi.org/10.1007/s00417-006-0474-4

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*(10), 6–6. http://doi.org/10.1167/11.10.6

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychological Science*, *22*(3), 384–392. http://doi.org/10.1177/0956797610397956

Brainard, D. H. (1997a). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176952

Brainard, D. H. (1997b). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. http://doi.org/10.1163/156856897X00357

Brand, J., Oriet, C., & Sykes Tottenham, L. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(1), 63–69. http://doi.org/10.1037/a0026950

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. http://doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. http://doi.org/10.1016/j.visres.2004.10.004

Cole, B. L., Lian, K.-Y., & Lakkis, C. (2006). The new Richmond HRR pseudoisochromatic test for colour vision is better than the Ishihara test. *Clinical and Experimental Optometry*, *89*(2), 73–80. http://doi.org/10.1111/j.1444-0938.2006.00015.x

Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Palgrave Macmillan. Retrieved from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/index.htm

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. http://doi.org/10.1177/1745691611406920

Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, *56*(293), 52–64. http://doi.org/10.2307/2282330

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics*, *70*(6), 946–954. http://doi.org/10.3758/PP.70.6.946

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. http://doi.org/10.3758/BF03193146

Feng, J., Pratt, J., & Spence, I. (2012). Attention and visuospatial working memory share the same processing resources. *Frontiers in Psychology*, *3*(APR), 1–11. http://doi.org/10.3389/fpsyg.2012.00103

Fougnie, D. (2008). *The relationship between attention and working memory*. *New Research on Short-Term Memory*. http://doi.org/10.3389/conf.fnhum.2011.207.00576

Fusser, F., Linden, D. E. J., Rahm, B., Hampel, H., Haenschel, C., & Mayer, J. S. (2011). Common capacity-limited neural mechanisms of selective attention and spatial working memory encoding. *European Journal of Neuroscience*, *34*(5), 827–838. http://doi.org/10.1111/j.1460-9568.2011.07794.x

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446. http://doi.org/10.1037/xge0000053

Healey, C. G., & Enns, J. T. (2012). Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics*, *18*(7), 1170–1188. http://doi.org/10.1109/TVCG.2011.127

Hesselmann, G., Hebart, M., & Malach, R. (2011). Differential BOLD Activity Associated with

Subjective and Objective Reports during "Blindsight" in Normal Observers. *Journal of Neuroscience*, *31*(36), 12936–12944. http://doi.org/10.1523/JNEUROSCI.1556-11.2011

Huang, L. (2015). Statistical Properties Demand as Much Attention as Object Features. *PLOS ONE*, *10*(8), e0131191. http://doi.org/10.1371/journal.pone.0131191

Huang, L. (2015). Statistical Properties Demand as Much Attention as Object Features. *Plos One*, *10*(8), e0131191. http://doi.org/10.1371/journal.pone.0131191

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*(3), 599–631. http://doi.org/10.1037/0033-295X.114.3.599

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the Limits Of Human Visual Awareness. *Science*, *317*, 823–825.

Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, *48*, 149–160. http://doi.org/10.1016/j.concog.2016.11.007

JASP Team. (2017). JASP (Version 0.8.12). Retrieved from https://jasp-stats.org/

Kane, M. J., Conway, A. R. A., Hambrick, D.Z., &, & Engle, R. W. (2007). Variation in Working Memory Capacity as Variation in Executive Attention and Control. *Variation in Working Memory*, 21–48.

Kingdom, F. A., & Prins, N. (2010). *Psychophysics: A Practical Introduction* (First). London: Elsevier Ltd.

Kleiner, M., Brainard, D., & Pelli, D. (2007). "What's new in Psychtoolbox-3?" *Perception 36 ECVP Abstract Supplement.* Retrieved from http://psychtoolbox.org/credits/[http://www.perceptionweb.com/abstract.cgi?id=v070821

Kline, R. B. (2004). *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research* (1st ed.). Washington DC: American Psychological Association.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737. http://doi.org/10.1016/S0042-6989(98)00285-5

Kruschke, J. K. (2016). *Doing Bayesian Data Analysis* (Second). London: Elsevier.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490. http://doi.org/10.3758/BF03210951

Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second). Cambridge, Massachusetts: MIT Press.

Macmillan, N. a, & Creelman, C. D. (2005). Detection Theory: A User's Guide. In *Detection theory: A user's guide (2nd ed.).*

Mauchly, J. W. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209. http://doi.org/10.1214/aoms/1177731915

Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4), 6, 1–18. http://doi.org/10.1167/15.4.6.doi

Mayer, J. S., Bittner, R. A., Nikolić, D., Bledowski, C., Goebel, R., & Linden, D. E. J. (2007). Common neural substrates for visual working memory and attention. *NeuroImage*, *36*(2), 441–453. http://doi.org/10.1016/j.neuroimage.2007.03.007

Monsell, S. (2003). Task switching, *7*(3), 134–140. http://doi.org/10.1016/S1364-6613(03)00028-7

Moore, C. M., & Egeth, H. (1998). How Does Feature-Based Attention Affect Visual Processing? *Journal of Experimental Psychology: Human Perception & Performance*, *24*(4), 1296–1310. http://doi.org/10.1037/0096-1523.24.4.1296

Morey, R. D., & Rouder, J. N. (2015). BayesFactor (Version 0.9.11-3).

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176953

Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, *13*(7), 3–3. http://doi.org/10.1167/13.7.3

Prins, N., & Kingdom, F. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rentzeperis, I., Nikolaev, A. R., Kiper, D. C., & van Leeuwen, C. (2014). Distributed processing of color and form in the visual cortex. *Frontiers in Psychology*, *5*(AUG), 1–14. http://doi.org/10.3389/fpsyg.2014.00932

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. http://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. http://doi.org/10.3758/PBR.16.2.225

RStudio Team. (2015). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Scharff, A., Palmer, J., & Moore, C. M. (2011). Extending the simultaneous-sequential paradigm to measure perceptual capacity for features and words. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 813–833. http://doi.org/10.1037/a0021440

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. http://doi.org/10.1093/biomet/52.3-4.591

Shih, S., & Sperling, G. (1996). Is There Feature-Based Attentional Selection in Visual Search ? *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 758–779.

Simmons, J., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. *SSRN Electronic Journal*, 1–4. http://doi.org/10.2139/ssrn.2160588

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. http://doi.org/10.1016/0010-0285(80)90005-5

Viviani, P., & Aymoz, C. (2001). Colour, form, and movement are not perceived simultaneously. *Vision Research*, *41*(22), 2909–2918. http://doi.org/10.1016/S0042-6989(01)00160-2

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science*, *6*(3), 291–298. http://doi.org/10.1177/1745691611406923

Whitney, D., Haberman, J., & Sweeny, T. (2014). From Textures to Crowds: Multiple Levels of Summary Statistical Perception. *The New Visual Neuroscience*.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it?, *5*(June), 1–7.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, *274*(5670), 423–428. http://doi.org/10.1038/274423a0

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *The Journal of Neuroscience*, *11*(March), 641–649. http://doi.org/10.1.1.211.5726

Appendix A

Effect of Fatigue on Performance in Experiment 2

In Experiment 2, participants ($N = 11$) engaged in 12 experimental blocks of 96 trials. Each block consisted of one of the seven conditions of the experiment. An analysis was undertaken to determine if performance was systematically related to the temporal order of the blocks (e.g., if performance systematically increases or decreases over the course of the 12 blocks). A 1 x 12 Bayesian repeated-measures ANOVA was conducted to examine the relationship between accuracy and block order. The results showed strong evidence in favour of the null hypothesis, $BF_{01} = 13.39$, Error % = 0.325, indicating that accuracy was not systematically related to the temporal order of the blocks (see Figure A1).

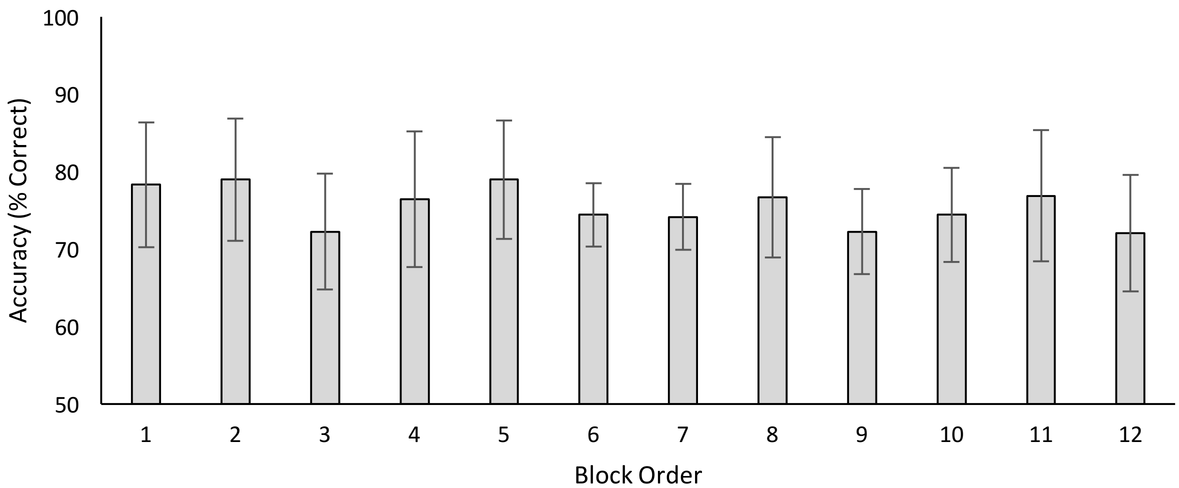

Figure A1. Accuracy across temporal block order. Error bars represent 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994).

Additionally, a 1 x 12 Bayesian repeated measures ANOVA was conducted to examine the effect of temporal block order on RTs. The analysis revealed strong evidence in favour of the null hypothesis, $BF_{01} = 15.38$, Error % = 0.30, indicating that RTs are not systematically related to the temporal order of the blocks (see Figure A2).

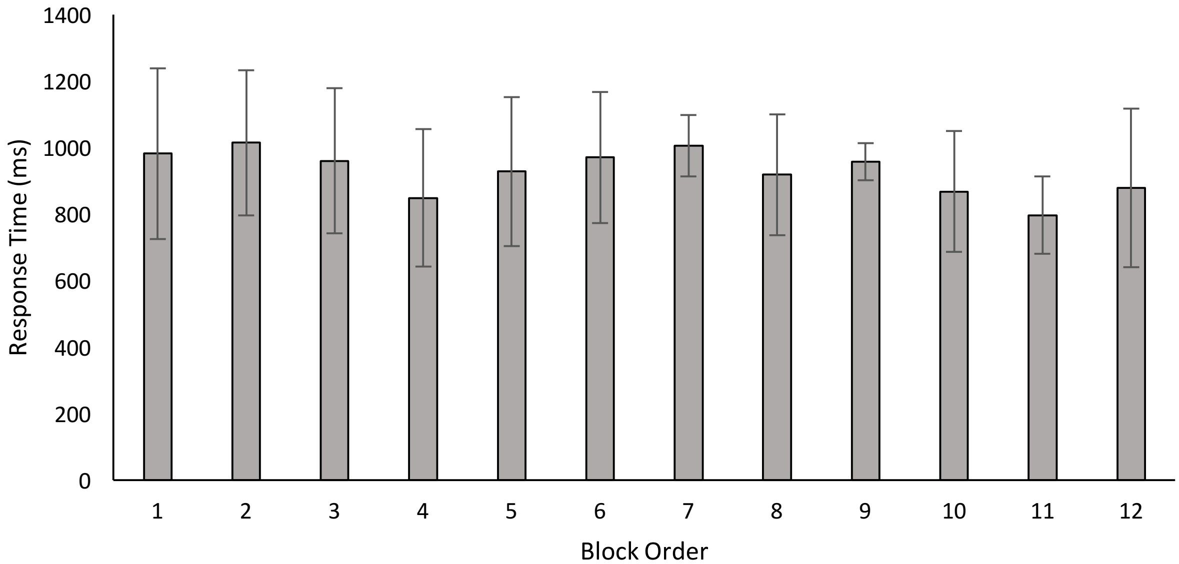

Figure A2. Response time across temporal block order. Error bars represent 95% confidence intervals using within-subjects corrected variance (Loftus & Masson, 1994).