Manuscript accepted for publication in *Attention, Perception, & Psychophysics,* and is available at DOI: 10.3758/s13414-020-02062-7

Individual differences in multiple object tracking, attentional cueing, and age account for variability in the capacity of audiovisual integration

Jonathan M. P. Wilbiks

Annika Beatteay

University of New Brunswick – Saint John

Address : Department of Psychology
University of New Brunswick – Saint John
100 Tucker Park Road
P.O. Box 5050
Saint John, New Brunswick, Canada
E2L 4L5

Telephone : +1 (506) 648-5658

E-mail : jwilbiks@unb.ca

Abstract

There has been a recent increase in individual differences research within the field of audio-visual perception (Spence & Squire, 2003), and furthering the understanding of audiovisual integration capacity with an individual differences approach is an important facet within this line of research. Across four experiments, participants were asked to complete an audiovisual integration capacity task (cf. Van der Burg et al., 2013; Wilbiks & Dyson, 2016; 2018), along with differing combinations of additional perceptual tasks. Experiment 1 employed a multiple object tracking task and a visual working memory task. Experiment 2 compared performance on the capacity task with that of the attention network test. Experiment 3 examined participants' focus in space through a Navon task and vigilance through time. Having completed this exploratory work, in Experiment 4 we collected data again from the tasks that were found to correlate significantly across the first three experiments and entered them into a regression model to predict capacity. The current research provides a preliminary explanation of the vast individual differences seen in audiovisual integration capacity in previous research, showing that by considering an individual's multiple object tracking span, focus in space, and attentional factors, we can account for up to 34.3% of the observed variation in capacity. Future research should seek to examine higher-level differences between individuals that may contribute to audiovisual integration capacity, including neurodevelopmental and mental health differences.

The processes of sensation, perception, attention, and higher levels of cognitive processing have previously been found to be related to one another on numerous different time scales and sensory modalities. The process of audiovisual integration refers to the combination of sight with sound, whereas the capacity for audiovisual integration refers to an individual's ability to appropriately integrate visual stimuli with sound (Welch & Warren, 1980). The capacity of audiovisual integration can be established by presenting participants with several trials in which varying numbers of dots change from white to black (or vice versa) repeatedly. On one of these changes, a tone is presented, which serves to increase the perceptibility of the dots that changed (cf. the 'pip and pop' effect; Van der Burg et al., 2008). Later, a dot is probed, and a participant identifies whether that dot changed (or not) in synchrony with the tone. Through a data-modelling procedure, estimates of an individual's capacity are established. That is: how many visual stimuli are they successfully able to integrate with a single tone.

In recent years, research into audiovisual integration has considered two different perspectives on its capacity. One line of research provides evidence that the capacity of integration is strictly limited to a single item, regardless of stimulus or environmental factors, or any other conditions that have been known to influence unimodal perceptual capacities (Van der Burg et al., 2013; Olivers et al., 2016). Conversely, research from our lab has shown that capacity can exceed one item, and that it is influenced by similar factors as are unimodal processes (Wilbiks & Dyson, 2016; 2018) such as the level of visual load (Lavie, 2005), the speed of presentation (Marois & Ivanoff, 2005), and temporal predictability (Wasserman, et al., 1983). There are important differences in the methodological features between these two sets of experiments; namely, the level of visual load present in the study (Lavie, 2005), the speed of presentation (Marois & Ivanoff, 2005), environmental factors such as temporal predictability

(Wasserman, Chatlosh, & Neunaber, 1983) and proactive interference (Kane & Engle, 2000). The version of the task being used in the current research is high in temporal predictability (the critical presentation is always in the same frame), and high in proactive interference (there are a relatively high number of pre-critical frames, as compared to other previous conditions). This combination of stimulus parameters had been previously established to be of moderate difficulty, and to yield relatively high estimates of audiovisual integration capacity. While these factors can be used to explain the differences between the capacities found by work in these two research groups, a phenomenon that is clear in the data across all experiments is the large range of capacity measures observed across participants.

In considering the differing perspectives of having either a one-to-one audiovisual integration system or one in which numerous stimuli in one modality may be integrated with another, one may wish to appeal to ecological validity. That is to say, it seems to be more parsimonious to have a system in which a single auditory stimulus integrates with a single visual stimulus, because this is more common in the world in which we live (cf. Olivers et al., 2016). However, there are also instances in which it may be adaptive to integrate numerous visual candidates with a single sound. For example, if you find yourself in a situation where you hear a hungry roar that may be a lion or a tiger, and see those two animals somewhere nearby, you would be best served to integrate the sound with each of them and hope for some *post hoc* information to arise that will help you disambiguate which one roared (and therefore, is more likely to approach). In a more modern (and more likely) example, somebody working in a healthcare context may hear an alarm sounding on a patient’s monitors while seeing multiple visual alerts flashing on a screen. Again, this is a situation where the first step is to note all possible binding candidates and then to progress through them to see which one needs attention.

One potential criticism of the experimental design employed in these studies would be to say that participants are not truly integrating auditory and visual information, but rather that they are attending selectively to visual stimuli while receiving an auditory cue. If this was the case, then using any other type of cue would also lead to similar performance. In previous research (Wilbiks & Dyson, 2016; Wilbiks et al., 2020), we have addressed this criticism by including versions of the task where visual cues are employed in place of auditory cues and have found that there is no facilitatory effect from visual cues. In the interest of time and simplicity for participants, we did not include this type of control experiment in the current research. However, while we believe that these previous studies provide evidence for the fact that we are measuring audiovisual integration, it is important to explore alternative interpretations of the data. Based on the parameters of the task, it is possible that participants are able to complete the task without strictly integrating auditory and visual information. Most of the dynamic phase of the paradigm involves tracking a number of visual objects that change polarity from black to white (or vice versa) a number of times. When a tone is presented on the critical presentation, participants are asked to note which location(s) changed at the same time as that tone, and then respond to whether a single probed location changed in synchrony with the tone. Considering this in the context of audiovisual integration, we can conceptualize the task in the same way as Van der Burg et al.'s (2008) 'pip and pop' effect, wherein an auditory tone boosted the perceptibility of a synchronous visual stimulus in a busy display. In this work, it was found that a visual cue did not aid in perception, which seemingly rules out an interpretation related to crossmodal attentional alerting. Based on this, as well as other work from Van der Burg et al. (2011; 2013), we conceptualized our task in the same way, and have conducted a series of studies using a similar paradigm.

While, in our opinion, audiovisual integration remains a valid explanation for this effect (in the same way as the ‘pip and pop’ effect described by Van der Burg et al., 2008), it is also possible to explain the findings as a crossmodal attentional capture effect (Matusz & Eimer, 2011). In this conceptualization, the tone serves as a simultaneously presented attentional cue that leads the participant to note the location(s) that changed at that moment. Previous research has shown that visual attention can be successfully enhanced through cues presented simultaneously with (e.g. Green & Woldorff, 2012) or even slightly *after* (e.g. Sergent et al., 2013) the target stimulus. Within this conceptualization of the task, the reason that the spatially non-specific visual-only versions we employed in previous research were ineffective could be because they drew attention away from the dots themselves. In Van der Burg et al. (2013), they included a condition with a specific visual cue (i.e. target dots flashed green shortly before the critical presentation) and found that this facilitated responding. We do this not to assert that this is definitively not audiovisual integration, nor that it is, but rather as a way of acknowledging that the findings may be driven by two different processes. According to Matusz and Eimer (2011), crossmodal attentional capture can occur as an isolated cueing effect, but may also be boosted further by multisensory integration. This novel conceptualization allows for more flexibility in interpretation, which future studies should seek to disambiguate. Additionally, given that it is possible that alternatives to audiovisual integration can drive this effect, it is also interesting to consider that monitoring multiple inputs is something that can be conceptualized in the healthcare environment described above, wherein one monitors numerous vital sign indicators while noting any that stray into dangerous ranges. In this instance, the ability being measured is the individual’s capacity for tracking the state of multiple objects while awaiting a cue.

Factors Affecting Audiovisual Integration

There are several factors that influence the integration of auditory and visual information, including spatial (e.g. Calvert et al., 2004) and temporal (e.g. Spence & Squire, 2003) coincidence of the individual stimuli, as well as the crossmodal congruency of the stimuli (Spence, 2011). In addition to these factors, there are others that are important to the successful integration of auditory and visual information. One such concept relevant to audiovisual integration research is that of attentional freeze (Vroomen & de Gelder, 2000). Attentional freeze refers to an illusion that occurs when an abrupt noise happens while watching rapidly changing visual stimuli. When this phenomenon occurs, the visual stimulus appears to momentarily freeze in place, as if stopped by the sound (Vroomen & de Gelder, 2000). Given the types of methodologies typically employed in audiovisual integration research, the illusion of attentional freeze may well be akin to the pip and pop effect described by Van der Burg et al. (2008), wherein the auditory stimulus is shown to boost a target out of a busy visual display through the process of audiovisual integration. Individual differences in the degree to which a participant experiences the attentional freeze (and/or pip and pop effect) may impact how well an individual is able to judge the simultaneity of auditory and visual stimuli.

Brand-D’Abrescia and Lavie (2008) explored the effects of distraction on task coordination between and within sensory modalities with the hypothesis that the demand of coordinating two separate tasks would reduce the ability of executive control to prevent interference from irrelevant distractors for either visual, or audio and visual tasks. It was found that when the two tasks employed were between the modalities, the effects of distraction were significantly increased as the coordination of the two tasks across modalities exerts a greater demand of executive control. Within the same modality, the tasks did not greatly increase the effects of distraction. Brand-D'Abrescia and Lavie (2008) suggested that the difficulty of the task is an additional factor to consider; however, when they reduced the difficulty of the auditory task, the effects of coordinating the two tasks continued to impact the distractibility of the subjects.

The findings of Brand-D'Abrescia and Lavie (2008) provide additional support for load theory (Lavie, 2005) and suggest that task coordination across sensory modalities is more demanding of executive control processes, thus more prone to distraction. Eayrs and Lavie (2018) examined the limits of capacity for visual perception resulting in "inattentional blindness" and established a method to predict whether an individual has superior perceptual abilities through a series of four studies involving subitizing, MOT, load-induced blindness, and change detection. In further considering multisensory integration, Hagmann and Russo (2016) examined bi-sensory and tri-sensory integration. They found that there seemed to be two qualitatively different groups of people – those who were assisted by redundant signals within auditory, visual, and somatosensory modalities, and others who did not show improvements in performance. They proposed engaging in individual differences research to consider the possibility that multisensory integration is not a universal process.

Individual Differences in Audiovisual Integration

In the work of Van der Burg and colleagues (2013), capacities found across 5 sets of conditions (and 2 experiments) ranged from .48 to 1.56. Wilbiks & Dyson (2016) found an even wider range of capacities from .18 to 2.50 for their 4 experiments, and Figure 1 illustrates the wide range of values observed in individuals. As such, it is of interest to consider whether these differences in capacity can be predicted by considering individuals' abilities on underlying perceptual abilities. Alternatively, we may find that audiovisual integration capacity is a functional unit of its own that individuals can excel at or not, independently of other abilities. There has been a recent increase in individual differences research within the field of audiovisual perception. In the concluding paragraph of their review of multisensory integration, Spence and Squire (2003) discuss the existence of significant individual differences in audiovisual perception, and the fact that there was a relative dearth of research examining factors that may lead to these differences. As an example, they discuss Stone et al. (2001) who found that their participants' point of subjective simultaneity between auditory and visual stimuli ranged from a 20 ms auditory lead to a 150 ms auditory lag – a 170 ms range. Examining the underlying mechanisms that may cause this kind of individual difference is important so as to ascertain how individuals can work to tune their perceptual abilities more precisely.

After Spence and Squire (2003), researchers began to examine individual differences within multisensory research. Miller and D'Esposito (2005) found that the temporal binding window – the window within which individuals can bind auditory and visual stimuli – varies in width, asymmetry, and location between people. Stevenson, Zemtsov, and Wallace (2012) summarized a number of findings with regard to the temporal binding window as it pertains to susceptibility to perceptual illusions such as the McGurk effect (McGurk & McDonald, 1976). First, they found that there is a clear relationship between the temporal binding window and multisensory integration processes, as these two processes develop in the individual at the same time (Hillock, Power, & Wallace, 2011). Stevenson and colleagues (2012) presented participants with a flash-beep task, employing varying SOAs to ascertain participants' temporal binding windows, while multisensory integration and susceptibility to audiovisual illusions were examined using sound-induced-flash tasks and McGurk illusions. They found a significant correlation between temporal binding window asymmetry and multisensory integration ability, such that preference for an auditory lag was associated with successful integration (and reduced susceptibility to illusions). Additionally, the narrowness of one's temporal binding window correlated positively with precision of multisensory integration. Stevenson and colleagues (2012) additionally noted a connection between the width of the TBW and the ability to perceive multisensory illusions in populations that experience multisensory dysfunction such as individuals with autism spectrum disorder (ASD). They suggest that individuals with ASD often have an impaired ability to integrate multisensory stimuli into a single unit and illustrate significant difficulty in perceiving the McGurk effect, which may be due to their atypically wide TBW (Foss-Feig et al., 2010; Kwakye et al., 2011).

Additionally, the research conducted by Kawakami and colleagues (2018) suggests that the narrowness of the temporal binding window and the accompanying deficiency in multisensory integration is associated with higher levels of autistic traits. Specifically, multisensory integration deficiency appears to be linked to problems with social skills, suggesting that multisensory integration and the temporal binding window are important for appropriate social interactions, especially when these interactions occur as a continuous flow of events (Kawakami et al., 2018). Zmigrod and Zmigrod (2016) found that a narrower temporal binding window was associated with increased sensitivity (and thus, improved performance) on Raven's APM and RAT tests. Eayrs and Lavie (2018) provide support to the notion that individuals with greater subitizing abilities have better sensitivity and accuracy when searching for changes in a task, including the detection of peripheral stimuli while attending a central task. This perceptual ability also proves to be beneficial during MOT tasks. Subitizing was also found to be an important predictor for capacity even when controlling for the effects of working memory. Overall, Eayrs and Lavie (2018) established subitizing as a measure of an individual's potential perceptual capacity, while also demonstrating more generally that certain perceptual and cognitive abilities are connected.

In a further examination of individual differences across temporal binding windows, Donohue and colleagues (2010) compared video game players with non-video game players to determine whether players experience an advantage when integrating multisensory information. In two separate discrimination tasks, video gamers appear to have the upper hand. In a simultaneity judgement task, they performed better than non-gamers at distinguishing whether audio and visual stimuli were occurring at the same time, or not. Additionally, in a temporal-order judgement task, it appears video game players have an increased ability to ascertain the temporal sequence of multisensory stimuli. These findings illustrate the importance of the temporal binding window to multisensory integration, as well as the idea that our individual experiences may alter our capacity for temporal integration (Donohue et al., 2010). Meyerhoff and Gehrer (2017) utilized an individual differences approach to investigate the relationship between visuo-perceptual capabilities and audiovisual integration. They compared the useful field-of-view of an individual with their ability to detect synchronous audiovisual events within a number of visual distractors. They found that visual capabilities predicted performance on the audiovisual synchrony task.

Current Research and Hypotheses

In light of recent research on the capacity of audiovisual integration (Van der Burg et al., 2013; Wilbiks & Dyson, 2016), we could extend this argument to considering individual differences in capacity. If it is the case that certain individuals do not show facilitation when presented with multiple redundant stimulus modalities, perhaps it is also the case that certain individuals do not experience audiovisual integration via a pip-and-pop mechanism (as per Van der Burg et al., 2008). To take a less polarized approach, perhaps there are certain underlying factors that contribute to the degree that one experiences audiovisual integration, ranging from almost non-existent (e.g. *K* = .180 at the bottom of the range observed in Wilbiks & Dyson, 2016) to exemplary (e.g. *K* = 4.00 in Wilbiks & Dyson (2018); the maximum achievable capacity occurring after training).

The current research aims to address some of these potential contributing factors to variation in audiovisual integration capacity. Experiments 1, 2, and 3 represent exploratory research wherein we identify important factors that may contribute to capacity, while Experiment 4 is a confirmatory study with the goal of building a predictive model of capacity. Experiment 1 examines performance on tasks known to index an individual's unimodal visual perceptual and/or cognitive ability, comparing performance on a visual working memory task and a multiple object tracking task to performance on an audiovisual integration capacity task. Experiment 2 considers the potential for the role of different attentional factors in establishing the capacity of audiovisual integration. Experiment 3 examines measures of focus – both in terms of vigilance over time and narrowness of an individual's field of focus, and their connection to audiovisual integration. After assessing correlations found between measures on the first three experiments, Experiment 4 identifies factors that are most likely to play a role in the capacity of audiovisual integration in an individual and seeks to build a predictive model that accounts for a large portion of the variation seen in this measure.

Experiment 1

In the first experiment of this series, we examined whether the ability to engage in multiple object tracking (MOT), as well one's visual working memory span, can account for some of the variation in the wide range of differences found in audiovisual integration capacity. Both processes have been studied in their own right and have had measures of their respective capacities determined. When objects are moving at a relatively low speed (<9.4°/s), findings show that individuals are able to track around 4 items (Pylyshyn & Storm, 1988). Other research shows that MOT capacity varies based on experimental factors (Bettencourt & Somers, 2009), and that there are large individual differences in capacity (Oksama & Hyönä, 2004). Oksama and Hyönä (2004) also show that there is a link between MOT capacity and other cognitive functions (including, not unremarkably, visual working memory). Drew and Vogel (2008) separate out two underlying processes in multiple object tracking – selection of targets (Yantis, 1992) and tracking itself (Cavanagh & Alvarez, 2005). In considering situations where stimuli are changing, but not moving around the display, Holcombe and Chen (2013) showed that the capacity of tracking multiple objects varies depending on the speed at which those objects are rotating. If the speed of rotation is less than 250 ms per rotation, a maximum of one object can be tracked, while at a speed of rotation less than 385 ms per rotation two objects can be tracked. Wilbiks and Dyson (2016) discuss the fact that knowledge of how many targets exist on each trial is a prerequisite of successful tracking, and that it would be important to have this information at the beginning of the trial in order to select targets successfully. In addition to this, it seems that one's ability to track multiple objects should contribute to successful enumeration of targets. It is also possible that the task used here draws on an individual's attentional resources in the same way that multiple object tracking draws on attentional resources (for a review, see: Meyerhoff, Papenmeier, & Huff, 2017). This gives us additional interest in the potential for correlations between an individual's multiple object tracking span, and their capacity for audiovisual integration. As such, a variant of Pylyshyn and Storm's (1988) multiple object tracking task will be employed to determine participants' MOT span.

Visual working memory has been found to have a capacity of between 3 and 4 items in most individuals (Cowan, 2001). However, Vogel and Machizawa (2004) found that around this average, there exists a wide range of capacities from 1.5 to 6, and that neural activity in the form of contralateral delay activity (CDA) serves as a significant predictor of capacity. Further, visual working memory has been shown to be predictive of abilities in visual perception (Irwin, 1992), attention (Woodman, Vogel, & Luck, 2001; Woodman, Luck, & Schall, 2007), and even higher-level abilities such as reasoning (Kyllonen & Christal, 1990). To this end, it is likely that the capacity of visual working memory will also be correlated with capacity of audiovisual integration.

The capacity of visual working memory has been successfully quantified by using a Corsi task (Corsi, 1972), wherein participants are asked to track a number of blocks which change colour in a certain sequential order, and then respond by tapping the blocks in the same order. More recently, this task was modernized and standardized, with a finding that the average score on the task is 6.2, with 68% of individuals falling between 5 and 7 (Kessels et al., 2000). This task will be employed to determine the visual working memory span of participants, and scores will be compared to measures of capacity that are determined through the main task.

We expected that both multiple object tracking and working memory capacities would be positively correlated with capacity as measured by the audiovisual integration task.

Method

All experimental and recruitment practices were approved by the Research Ethics Board at Mount Allison University and at the University of New Brunswick Saint John.

Participants

An *a priori* calculation was conducted to determine an adequate sample size, using the effect size of the highest level interaction found in Wilbiks & Dyson (2018), which was η_p^2 = .165. Using this effect size and a desired power (1 - β) of 0.9, it was found that a sample size of 45 participants would be adequate for correlational research. Knowing that not all participants would produce usable data sets, a decision was made to recruit 55 participants from an undergraduate research participant pool, and to compensate them for their participation with partial class credit in an introductory psychology course. After data collection was complete, we calculated a 95% confidence interval (CI) around 50% (chance responding) on the audiovisual integration capacity task and removed the 7 participants who performed within that CI, on average and across all conditions (after Wilbiks & Dyson, 2016). The final sample consisted of 48 participants with an average age of 19.2 years (SD = 2.1), with 30 females and 43 right-handed individuals.

Audiovisual Integration Capacity Task

The main audiovisual integration capacity task followed what has been employed previously in our lab (Wilbiks & Dyson, 2016; 2018). Visual stimuli were presented on a Dell 2407WFP monitor at a viewing distance of approximately 57 cm. Stimulus generation and delivery were controlled by Neurobehavioral Systems Presentation software. 16 individual stimulus combinations were created, by orthogonally varying the display duration of visual stimuli (200, 400, 600, or 800 ms), and the number of visual stimuli changing on each alternation (1, 2, 3, or 4). These 16 conditions were each presented twice (1 valid probe, 1 invalid probe) to create an experimental block with 32 trials. Each participant completed one practice block of 16 trials, and 8 experimental blocks consisting of 32 trials each, for a total of 256 experimental trials.

The visual stimuli employed consisted of dots 1.5° in diameter displayed in either black (0, 0, 0) or white (255, 255, 255) against a mid-grey background (128, 128, 128). Eight dots at a time were presented along an implied circle, which had a diameter of 13°, the center of which was marked by a 0.15° fixation dot. A single, smaller probe dot was overlaid on a target dot at the end of each trial, and was red (255, 0, 0) with a diameter of 1°. Auditory stimuli were created using SoundEdit 16 (MacroMedia) and consisted of a 60 ms sine tone with 5 ms linear on-set and off-set ramps, presented at a frequency of 500 Hz. Sounds were presented binaurally via Sennheiser HD 202 headphones at an intensity of approximately 74 dB(C).

Each trial began with the fixation point displayed in the center of the screen for 500 ms. Independent sets of 8 (black or white) dots were generated for each frame and presented for one of: 200, 400, 600, or 800 ms (dependent on display duration condition), for a total of 10 presentations (see Figure 2 for a visual representation). These sets of dots were generated for each trial and for each participant based on the following guidelines. The initial state was determined by randomly setting each location to be black or white. On a given trial, anywhere from 1 to 4 locations were to be changing between presentations, and these locations were randomly determined as well. On the penultimate (9th) presentation, the onset of the dots was accompanied by an auditory tone. Following a 1000 ms retention interval, the final array of dots was displayed again, along with an overlay of a red probe dot on one of the dots. Participants were asked to respond to whether the dot at the probed location had changed or not on the critical frame using a computer keyboard. On valid trials (where the correct response was ‘yes’), the probe dot was randomly assigned to one of the locations that had changed on the frame change with the tone, and on invalid trials it was randomly assigned to one of the locations that did not change on that frame. No feedback was provided, and the subsequent trial began shortly after a response was entered. Trial order was randomized in practice and in experimental trials.

Visual Working Memory Task

The visual working memory task employed was a variant of the Corsi task (Corsi, 1972), and was run using PsyToolKit (Stoet, 2010; Stoet, 2017). Participants were presented with an array of ten magenta squares measuring 3° x 3°, which were randomly arranged on the screen. A sequence of these squares then changed to yellow and back again at 300 ms intervals, after which an audiovisual cue of "GO" indicated that participants should respond. Participants then clicked on squares in the same sequence as they believed they had seen change, before clicking a green "DONE" button. Feedback was provided, and the next trial began after 1000 ms. The first trial involved a sequence of two squares. The task used a modified staircase procedure wherein subsequent trials would have the same length of sequence (if the response was incorrect), or a sequence one square longer (if the response was correct). Once three consecutive incorrect responses were given, the program was halted, and the longest correctly responded to sequence was considered to be the individual's visual working memory score.

Multiple Object Tracking Task

The multiple object tracking task involved presentation of a number of visual 'targets', which were blue circles 1.5° in diameter. The first trial started with five targets being presented to participants, and then the targets began to move in random directions at a speed of 3.0°/s. Once the targets began to move, a number of identical non-target stimuli equal to the number of targets appeared and moved around the screen at the same rate of speed as the targets. Participants were asked to keep track of the movement of the targets, and once the targets stopped moving (after 6 seconds), to click the targets. This task also used a modified staircase procedure – in this instance, providing two consecutive correct responses led to an increase in difficulty (e.g. from 5 to 6 targets), while providing two consecutive incorrect responses led to a decrease in difficulty (e.g. from 5 to 4 targets). Participants completed ten trials altogether. Multiple object tracking scores were calculated by taking the maximum value of targets tracked successfully.

Results

Estimates of audiovisual capacity (*K*) were derived in the same manner as in previous research (Van der Burg et al., 2013; Wilbiks & Dyson, 2016; 2018). This process involves calculating the raw proportion for each participant for each combination of display duration and number of objects tracked. The raw proportion correct scores are then fitted to a model based on Cowan's (2001) *K*, wherein if the number of visual elements changing is less than or equal to an individual's capacity, their performance on those trials should approach certainty (i.e. if $n \leq K$, $p = 1$), but if an individual's capacity is less than the number of visual elements changing (i.e. $n > K$), then their performance can be modelled as follows: $p = K/2n + .5$. This fitting procedure optimizes the value of *K* by minimizing the root-mean-square error between the raw proportion correct and the ideal model. An initial analysis compared capacity estimates at different display durations by means of a one-way ANOVA, the results of which are displayed in Figure 3. This analysis revealed a significant main effect of duration: $F(3, 141) = 46.63$, MSE = 0.255, $p < .001$, $\eta_p^2 = .498$. Post-hoc comparisons using Bonferroni corrections revealed that capacity was significantly different (that is, increased with increasing display duration) at each level except for a non-significant difference between capacity at 600 and 800 ms ($p_{bonf} = .054$).

The average working memory span demonstrated by participants was 6.33, with a standard deviation of 1.02, a minimum of 4, and a maximum of 8. The average multiple object tracking span was 4.83, with a standard deviation of 1.04, a minimum of 3, and a maximum of 7. Individual differences between participants were assessed by comparing participants' scores on multiple object tracking and visual working memory tasks with their respective capacity for audiovisual integration. Data were entered into a series of Pearson correlations to compare capacity at each display duration with one another and with MOT and VWM spans. Full results of the correlations are displayed in Table 1, but specific pertinent effects are discussed here. Significant correlations were found between capacity estimates for all combinations of display durations (*r*s ranging from .569 - .833), which provides additional evidence in support of our perspective of integration at slow and fast durations being quantitatively different measures of the same process (as per Wilbiks & Dyson, 2016; 2018).

Comparisons of capacity estimates (*K*) to visual working memory span did not reveal any significant correlations (*r*s ranging from .088 - .180). However, multiple object tracking span was found to significantly correlate with capacity at 200 ms (*r* = .404, *p* = .004), 400 ms (*r* = .377, *p* = .008), and 600 ms (*r* = .401, *p* = .005), but not at 800 ms (*r* = .255, *p* = .080). These moderate, significant correlations (plotted in Figure 4) indicate that individuals who exhibit higher levels of multiple object tracking ability also show greater capacity for audiovisual integration at all but the slowest duration tested. This suggests that there may be a connection between an individual's ability to track movement of multiple visual objects, and their ability to integrate a greater number of visual objects with an auditory tone, especially under conditions of relatively high and moderate difficulty (as indexed by speed of presentation), but not under relatively low difficulty conditions (i.e. 800 ms). However, while these correlation coefficients vary numerically, Z-tests revealed no significant differences found between them, so we cannot conclusively state that multiple object tracking is differentially affecting audiovisual integration capacity at different speeds of presentation.

Discussion

Experiment 1 explored the connections between multiple object tracking and visual working memory to the capacity of audiovisual integration. Our results demonstrated two important findings. Firstly, the correlation between visual working memory and the capacity for audiovisual integration was not found to be statistically significant, as well as being weak in magnitude. Although Irwin (1992) successfully demonstrated that visual working memory is predictive of perceptual abilities, and others (Woodman et al., 2001; Woodman et al., 2007) found it to be vital to attention, thus far visual working memory has not shown to be conducive to predicting the capacity of audiovisual integration.

Secondly, a significant correlation does exist between the ability to track multiple objects and the capacity of audiovisual integration. Pylyshyn and Storm (1988) showed that participants were capable of tracking 4 objects when moving at a relatively low speed (<9.4°/s); and in the current experiment, participants demonstrated an average multiple object tracking span of 4.83, with a minimum of 3, a maximum of 7, and a standard deviation of 1.02. When the individual differences in scores were compared with their corresponding capacity for audiovisual integration there appears to be a moderate, yet significant correlation between the ability to track multiple objects and audiovisual integration. This was an expected finding, given that the audiovisual integration capacity task involves a requirement of tracking the states (in this case, colour) of multiple objects, while most multiple object tracking tasks involve tracking the movements of multiple objects. It is sensible to expect individuals who are better at tracking the movements of multiple objects to also be proficient at keeping track of the states of multiple objects, which are then cued by the tone. These findings are also very similar to those found by Meyerhoff & Gehrer (2017), who took an individual differences approach to studying the capacity of audiovisual integration. This work tested whether an individual's detection of visual object direction changes that may be coincident with tones could be predicted using that individual's useful-field-of-view or their short-term memory for the colours of visual stimuli (Meyerhoff & Gehrer, 2017). They found that an individual's useful-field-of-view on a visual perception task predicted their performance on an audiovisual detection task, but that visual short-term memory did not predict performance on the audiovisual task.

Overall, the results of Experiment 1 suggest that multiple object tracking capabilities are more closely predictive of the capacity for audiovisual integration than the capacity of the visual working memory. As such, multiple object tracking span will be examined further as a potential predictor of audiovisual integration capacity in Experiment 4, while visual working memory span will not be studied further.

Experiment 2

In Experiment 2, we sought to establish the potential roles of endogenous and exogenous attentional factors in determining the capacity of audiovisual integration. Visual spatial attention serves as the basis for numerous control systems (Fan et al., 2009), and as such it is likely an important element involved in audiovisual integration. To appropriately gauge the impact of attention on individual differences in audiovisual integration capacity, the Attention Network Test-Revised (ANT-R) has been used to assess three separate components of attention: the alerting network, the orienting network and the executive control. The alerting network is closely linked to higher order cognitive processes and plays a critical role in preparing to perceive stimulus (Fan et al., 2009). Perhaps most relevant to the current experiment is the orienting network, as it involves the selection of specific aspects of various sensory inputs (Fan et al., 2009). Orienting may be exogenous, such as when a sudden stimulus draws the attention reflexively, or endogenous, when the attention is shifted voluntarily (Fan et al., 2009; Posner, et al., 1984; Tang et al., 2016). Orienting involves the slow or rapid shifting of attention either within a modality, or amongst various modalities as in audiovisual integration (Fan et al., 2009). The executive control aspect of attention is often involved in planning and decision making, as well as error detection, common higher order processes. (Fan et al., 2009).

The current experiment used the ANT-R model of presenting participants with a fixation cross in the middle of a grey screen. On either side of the fixation cross were two identical rectangles wherein the stimulus would be presented throughout the trials. The participants were provided a number of cueing conditions prior to the stimulus being presented, drawing on either the orienting or alerting attention networks. These cues could be either valid or invalid. Van der Stoep and colleagues (2015) were among the first to investigate the impact of exogenous attention on multisensory integration as prior multisensory research had focused almost exclusively on the effects of endogenous attention. They found that exogenous attention plays a larger role on audiovisual integration when spatial attention is task relevant. Additionally they showed that exogenous spatial attention can increase the speed of multisensory stimuli processing, while also decreasing the overall capacity of multisensory integration. Previous studies, such as that by Talsma and Woldorff (2005), found that the influence of endogenous attention on multisensory integration was beneficial, in that endogenous attention on a specific location enhances the multisensory integration of stimuli; Van der Stoep and colleagues (2015) found that when the location of the stimulus was unpredictable, the capacity for multisensory integration was limited. As such, while exogenously attending a multisensory stimulus makes an observer capable of rapid processing, exogenous attention may actually reduce the capacity to integrate audiovisual stimuli. Therefore, it is likely that participants who can orient rapidly and disregard invalid cues will be able to integrate a higher quantity of audiovisual stimuli, as opposed to participants who experience difficulty orienting and are less able to ignore invalid cues. For Experiment 2, we hypothesized that there would be negative correlations between cue validity and audiovisual integration capacity due to the high costs of disengaging from an invalid flanker, and that individuals who experience less difficulty disengaging from invalid cues would be more likely to have a higher capacity for audiovisual integration.

Method

50 new participants were recruited in the same manner as in Experiment 1. Based on the same criteria as in Experiment 1, 8 of these participants were removed from the analysis, leaving a final sample of 42 participants with an average age of 20.0 (SD = 4.1), comprised of 30 females and 12 males, and a total of 4 left-handed individuals. The audiovisual integration capacity task used was identical to the one employed in Experiment 1. In addition to this task, a version of the Attention Network Test-Revised (Fan et al., 2009) was employed, and was run in Inquisit 5 (version 5.0.14.0).

Attention Network Test – Revised

The Attention Network Test – Revised (ANT-R) involved participants being presented with a black fixation cross in the centre of a grey screen. Centered 5° to the left and the right of the fixation cross, rectangles were presented with a black outline and empty centre, measuring 5° x 2°. The main stimuli used were sets of five arrows (1° x 1° each), arranged in a horizontal line. In congruent conditions, all five arrows were pointing in the same direction (either left of right), while in incongruent conditions, the centre arrow was pointing in the opposite direction of the other arrows (e.g. centre arrow pointing right, other arrows all pointing left). Participants were asked to maintain fixation on the cross until such time that arrows were presented in one of the rectangles. When the arrows were presented, participants were asked to respond to the direction of the centre arrow by pressing the "E" key on a computer keyboard with their left hand if the arrow was pointing left, and by pressing the "I" key with their right hand if the arrow was pointing right. As per the design of the ANT-R, participants were provided with a combination of different cueing conditions prior to the arrows being presented. They could receive no cue (the arrows appearing is the first stimulus that occurs), an alerting cue (both rectangles flash rapidly (100 ms) before targets are presented), or an orienting cue (one rectangle flashes rapidly (100 ms) before targets are presented). The orienting cue could be either valid (cues the location in which the arrows will appear) or invalid (cues the location in which the arrows will not appear), with a validity of 80%. If cues were present on a given trial, they could be presented 0, 400, or 800 ms ahead of the targets.

Results

Estimates of audiovisual capacity (*K*) were established as in Experiment 1 (means in Figure 3), and an initial analysis by means of a one-way ANOVA revealed a main effect of duration: F(3, 123) = 43.56, MSE = 0.291, p < .001, η_{p^2} = .515. Post-hoc comparisons using Bonferroni corrections (*p* < .05) revealed significant differences between each pairwise comparison of duration except from 600 to 800 ms (p_{bonf}= .450). These findings directly replicated the findings of Experiment 1 with regard to capacity estimates.

Analysis of the ANT-R involves computation of a number of scores that address an individual's abilities in terms of different attentional networks – a summary of these scores is included in the introduction of this experiment. The average scores found in our sample were as follows: Alerting (M = 46.6, SD = 56.6), Validity (M = 102.5, SD = 47.4), Orienting (M = 95.3, SD = 55.7), Flanker Conflict (M = 148.9, SD = 65.6), Location Conflict (M = -0.5, SD = 33.2). Contributions to individual differences between participants were assessed by comparing participants' individual sub-scores of the ANT-R with their scores on the audiovisual integration capacity task. Data were entered into a series of Pearson correlations to compare capacity at each display duration with one another and with attentional alerting, attentional cue validity, orienting time, flanker conflict, and location conflict. Full results of the correlations are displayed in Table 2, with significant correlations discussed here. Significant correlations were found between capacity estimates for all combinations of durations (*r*s ranging from .535 - .687), as in Experiment 1. Additionally, no significant correlations were found between the different networks tested by the ANT-R, which confirms that the respective contributions of each network are independent from one another.

Moderate negative correlations were found between cue validity and audiovisual integration capacity at 200 ms (*r* = -.331, *p* = .032), 600 ms (*r* = -.398, *p* = .009), and 800 ms (*r* = -.366, *p* = .017), with no significant correlation at 400 ms (*r* = -.289, *p* = .063). These significant negative correlations (see Figure 5) were expected, as the validity score is a difference score, subtracting reaction times on valid trials from reaction times on invalid trials. As such, a larger score here represents a greater cost of disengaging from an invalid flanker. Therefore, the negative correlations found here indicate that individuals who suffer less disengaging cost from invalid attentional cues are more likely to have greater audiovisual integration capacity estimates.

Significant correlations were also found between flanker conflict effect scores and capacity estimates at 400 ms (*r* = -.318, *p* = .040), 600 ms (*r* = -.359, *p* = .020), and 800 ms (*r* = -

.338, *p* = .029), but not at 200 ms (*r* = -.222, *p* = .157). As in the previous correlation, the negative correlations found here represent the association of reduced flanker conflict cost (a difference score of trials when the arrows were all pointing in the same direction subtracted from trials when the flanker (non-central) arrows were pointing in the opposite direction) with increased audiovisual integration capacity.

Discussion

Previously, we had hypothesized that participants who could orient rapidly and disregard invalid cues would be able to integrate a higher quantity of audiovisual stimuli, as opposed to participants who have trouble orienting and are less able to ignore invalid cues. The analysis conducted following Experiment 2 indicated a replication of the results for audiovisual capacity found in Experiment 1, showing significant differences between 200, 400, and 600 ms, but not between 600 and 800 ms. As noted by Van der Stoep and colleagues (2015), when the location of the stimulus was unpredictable due to the appearance of an invalid cue, the integration of audiovisual information was significantly worse. We have found similar results, but by examining audiovisual integration capacity, rather than multisensory spatial integration, which provides an additional connection between audiovisual integration capacity and unimodal attentional factors. The current study supported this theory by illustrating that the ability to deflect false cues differs between participants, and those who are unable to ignore the invalid cue have a noticeably lower capacity for audiovisual integration than those who were capable of ignoring the invalid cues, thus performing at a higher level of audiovisual integration capacity. Although flanker conflict and cue validity were found to correlate significantly with capacity, unexpectedly, orienting of attention did not appear to play as large a role as originally anticipated. As we observed several significant correlations between subscores on the ANT-R and audiovisual integration capacity, we will include an examination of the attention network test as a potential predictor of audiovisual integration capacity in Experiment 4.

Experiment 3

Experiment 3 examined the potential connection between audiovisual integration capacity and an individual's ability to maintain focus: both focus over time (i.e. vigilance) and focus in space (i.e. global vs. local focus). The Mackworth clock task has previously been employed for testing vigilance, the readiness to act through an extended period of time (Mackworth, 1948; Hancock, 1986; Lichstein, Riedel & Richman, 2000). This task assesses the participant's level of vigilance by having them monitor a white circle with a single clock arm that makes one small movement every second. The participant is then told that occasionally the arm will make a movement that is twice the length of the standard move. When this occurs, they are to press the response key. Mackworth (1948) illustrated that the longer the participants watched the movement of the clock, the higher the percentage of missed signals became. A two-hour span of testing was measured in four half hour increments and the percentage of missed cues rose from 15.7% within the first half hour, up to 28.0% by the last half increment. This experiment suggested that when testing is conducted for a significant length of time, visual perception decreases. Additionally, the results indicate that when one has recently perceived a visual stimulus and is not anticipating another one to arrive so soon, the participant may have reduced levels of readiness (Mackworth, 1948). The audiovisual integration capacity task being employed requires a participant to maintain vigilance over a period of between 2000 and 8000 ms (depending on display duration), looking for instantaneous changes in stimulus polarity. As such, we expect that an ability to remain vigilant over time will be associated with successful performance on the capacity task.

Experiment 3 also makes use of the Navon (1977) task to test how participants interpret visual stimuli. In this task, participants view a series of letters created from smaller versions of letters and they must indicate if they see a specific letter. In the current experiment, Hs and Os were the target letters used. Participants were instructed to respond as to whether an H or and O was present in a visual display, regardless of whether it was large or small. Participants who responded more quickly to large letters composed of smaller ones indicated that they maintain an initially wider lens of focus when examining a visual stimulus, while participants who respond more quickly to small letters versus large ones tend to maintain a narrower focus (Navon, 1977). The increased speed that comes with observing big picture stimuli likely plays a role in the capacity for audiovisual integration, although it is not uncommon for individuals to spend more time attending to the more complex aspects of a figure, rather than observing the stimulus as a whole. Navon (1977) further notes that individuals typically perceive a stimulus as a whole first, but then draw more details from the stimulus the longer they are able to observe it, from global features to localized details. We expected participants who tend to have a global precedence in perception (i.e. a wider field of focus) to have a greater capacity of audiovisual integration as compared to participants whose initial focus is more attentive to the smaller details of the display.

Method

50 new participants were recruited in the same manner as in previous experiments, and 9 of them were removed from the data set as per the criteria outlined in Experiment 1. The final sample consisted of 41 individuals with an average age of 20.0 (SD = 3.2), 27 females and 14 males, and 3 left-handed individuals. Again, the identical audiovisual integration capacity task was employed, in addition to tasks measuring participants' ability to maintain focus over an

extended period of time (300,000 ms; Mackworth, 1948), and participants' tendency to focus on global or local features (Navon task; Navon, 1977).

Mackworth Clock Task

This task involves presentation of a single arrow, representing a hand on an analog clock face. This arrow was presented in green, was 10° in length, and began pointed at the "12:00" position. Every 1000 ms, the arrow changed orientation by moving the outside point of the arrow in a circular direction, simulating the movement of a clock. On most of these movements, it shifted by 6°, but on a randomly selected 5% of the movements, it shifted by 12°. Participants were asked to maintain focus on the movement of the arrow, and to press the spacebar on a computer keyboard whenever the movement was to a greater degree than the 'normal' movements. This task continued for 300 seconds.

Navon Task

The Navon task involved the presentation of a large (approx. 10° height on average, but with variation between specific letters) letter, which was composed of a number of identical smaller letters (approx. 1° each). These letters were presented in white font, on a black background. Participants were asked to respond to whether they saw any Hs or Os, regardless of whether they appeared as the large letter or the small letters. Participants completed a total of 100 trials, of which 25 had an H or O present as the large (global) feature, 25 had an H or O present as the small (local) feature, and 50 had no Hs or Os present.

Results

Estimates of audiovisual capacity (*K*) were established as in earlier experiments (means in Figure 3), and an initial analysis by means of a one-way ANOVA revealed a main effect of duration: F(3, 120) = 35.57, MSE = 0.271, p < .001, η_p^2 = .471. Post-hoc comparisons using

Bonferroni corrections revealed significant differences between capacity estimates at each display duration (all *p*s < .031). For analysis of the Mackworth clock task, we employed signal detection calculations, finding that the average d' in our sample was 3.09 (SD = .87). Analysis of the Navon task found that there was an average Navon score of -2.70 (SD = 24.44). This means that, on average, our sample tended to attend to the global and local features of the display with relatively equal weighting.

Individual differences between participants were assessed by comparing participants' scores on the Mackworth Clock task and the Navon task with their scores on the audiovisual integration capacity task. Data were once again entered into a series of Pearson correlations to compare capacity at each display duration with one another and with d' in the Mackworth task, and with global precedence in the Navon task. Full results of the correlations are displayed in Table 3, with significant correlations discussed here. Significant correlations were found between capacity estimates for all combinations of durations (*r*s ranging from .363 - .818), again replicating and supporting previous research findings. No significant correlation was found between vigilance on the Mackworth task and capacity estimates (all *r*s < .229). However, significant correlations (plotted in Figure 6) were found between global precedence scores and capacity estimates at 600 ms (*r* = .311, *p* = .048) and 800 ms (*r* = .324, *p* = .039), although not for capacity at 200 ms (*r* = .214, *p* = .179) or 400 ms (*r* = .185, *p* = .246). This suggests that, when the capacity task was in the easier range (slower SOAs), taking a global perspective on the visual display may be associated with greater capacity estimates, while the same was not true when the capacity task was more difficult. As in Experiment 1, a z-test did not reveal significant differences in this relationship across different speeds of presentation.

Discussion

Unexpectedly, we found weak, non-significant correlations between the level of vigilance ascertained by the Mackworth clock test and the capacity for audiovisual integration. This indicates that vigilance may not have as large a role in audiovisual integration as we expected. However, the Navon task produced significant correlations between global precedence and the capacity for audiovisual integration during the 600ms and 800ms time intervals. As noted both behaviourally and with electrophysiological evidence in previous experiments (Wilbiks & Dyson, 2016), the ability to perceive and integrate increases when using slower intervals, whereas during the 200ms and 400ms intervals, there were no significant correlations between global precedence scores and capacity estimates. As we did observe significant correlations between Navon scores and audiovisual integration capacity, we will employ this task again in Experiment 4.

Experiment 4

Having completed the exploratory analyses described in the preceding three experiments, we have been able to identify measures of unimodal perception, attention, and focus that correlate significantly with an individual's capacity for integrating auditory and visual stimuli. This provides us with important information regarding the nature of audiovisual integration capacity. Namely, it suggests that the ability to track multiple objects, an ability to maintain attention on valid cues (and to disengage from invalid cues), an ability to avoid being adversely affected by incongruent flanker stimuli, and a tendency to focus on global, rather than local features, are necessary skills in elevating one's capacity of audiovisual integration. Having established these through correlational research, Experiment 4 represents an attempt to establish a model with which we can predict an individual's audiovisual integration capacity through comparison of their scores on these tasks. This process will also elucidate the amount of the variability we see in audiovisual integration capacity that can be explained by these underlying processes, and how much of the variability we see is due to something specific to audiovisual integration itself.

This analysis was completed through the use of a hierarchical linear regression, which we planned based on the basis of theoretical concerns as well as previous research findings. At the first step of the regression, we wished to include demographic factors that were predicted to play a role, and in this case that was the age of the participant. Previous research has shown that with increase in age comes decreases in audiovisual integration abilities (e.g. de Dieuleveult et al., 2017). At the second step, we added in multiple object tracking, which was observed to be the strongest factor in the correlational analyses, and which has also been previously shown to draw on similar attentional resources as audiovisual integration (e.g. Meyerhoff et al., 2017). Step 3 involved the addition of attentional factors, which have previously been shown to be related to audiovisual integration (e.g. Van der Stoep et al., 2015). However, based on our earlier findings in Experiment 2, only the Alerting and Orienting subscores were added at this step, as they were the scores that correlated significantly with audiovisual integration capacity. At Step 4, the Navon score was added to the model, as this was observed to play a role in our correlational analyses, and at Step 5 all remaining predictors were added to the model.

It is important to note that at the outset of the analysis we were ambivalent with regard to the findings of the regression model. Based on the findings of the first three experiments, we certainly *expected* to find that the individual variability in audiovisual integration capacity can be predicted to some extent by the measures that have been discussed. This is in alignment with other research into individual differences in perception (e.g. Eayrs & Lavie, 2018; Robison & Unsworth, 2017). However, we do not expect the model to be able to account for all variability in audiovisual integration capacity, as we expect that there is variability between individuals on their ability to track and integrate information from different modalities as well as other, as yet untested perceptual and cognitive abilities. As such, the intent of this experiment is to establish how much of the variability can be explained through the sources tested, as well as which abilities matter more (or less) than others. If this investigation is successful, then we will have established a predictive model which should be further examined. If, however, we are unable to account for significant amounts of variation in audiovisual integration, this would suggest that audiovisual integration capacity itself may be an independent construct. That having been said, we hypothesized that multiple object tracking, attentional cue validity, flanker conflict, and global precedence would each contribute significantly to the capacity differences observed between individuals.

Method

59 participants were tested, with an average age of 21.2 (SD = 6.1). The group consisted of 13 males and 46 females, and 3 of them were left-handed. Participants were asked to complete the following tasks from the earlier experiments: multiple object tracking task, attentional network test – revised, Navon task, and audiovisual integration capacity task. The details of each task were identical to the way they were described in earlier experiments.

Results

Estimates of audiovisual capacity (*K*) were established as in the earlier experiments (means in Figure 3), and an initial analysis by means of a one-way ANOVA revealed a main effect of duration: F(3, 174) = 46.296, MSE = 0.272, p < .001, η_{p^2} = .444. Post-hoc comparisons using Bonferroni comparisons (*p* < .05) revealed that all differences were significant from one another, with the exception of capacity at 600 and 800ms (p_{bonf} = .292). Mean scores for the other measures were calculated (see Table 4) and were largely in line with what was found in the earlier experiments. For the multiple object tracking task, we completed an additional calculation to account for the effects of guessing in our task, due to variation in the number of targets and distractors from trial to trial. For this, we used the formula provided by Koldewyn and colleagues (2013), based on that of Hulleman (2005). According to this formula, quantifying an individual's multiple object tracking capacity (*K*) can be achieved through the following formula:

$$K = \frac{(oc - t^2)}{(o + c - 2t)}$$

where o is the number of total objects (targets + distractors), t is the number of targets to be tracked, and c is the number of correctly identified targets on each trial. We used this formula to assess the capacity estimate for each participant for each trial, and then took the average of these estimates to produce an individual's capacity score.

Data were then entered into hierarchical linear regression models for each display duration, beginning with the predictor that had the highest correlation with capacity in earlier experiment, and progressing through subsequently weaker correlations. Based on Experiments 1-3, the order in which elements were added to the hierarchical correlation is as follows: Step 1 - Age; Step 2 - MOT; Step 3 - Alerting, Orienting; Step 4 - Navon; Step 5 - all other factors. At each stage, the model was assessed as to the overall predictive power of the model, as well as of the strength of each individual predictor. The steps of each regression model and strength of each predictor in those models are available as supplementary materials (Tables S1-S4), and the strongest regression model for each duration is discussed here. It is important to note that we were seeking the regression model for each duration that accounts for the greatest amount of variability in the data (i.e. the highest R-squared value), and we were less concerned with specific predictors that reach, or do not reach, standard levels of statistical significance. Given that we had *a priori* hypotheses (and preliminary results from Experiment 1-3) in support of these factors, we decided to proceed to the higher steps of the regression to eventually include all predictor variables.

At 200 ms, the strongest statistically significant regression equation can account for 12.1% of the variation in audiovisual integration capacity. The strongest contributor to this is multiple object tracking (MOT), ($\beta = .312$, $p < .05$), with an additional contribution from participant age. MOT was expected to play an important role in the regression equation based on what was found in the first 3 experiments, and this was observed again in Experiment 4. While age was not specifically predicted to play a role, it is also not entirely unexpected that increasing age would be associated with lower capacity, based on prior research into multisensory integration (for a full review see: de Dieuleveult, Siemonsma, van Erp, & Brouwer, 2017).

Strangely, none of the models at 400 ms reached statistical significance, but the final model did account for 22.1% of the variation in the data, with MOT as the only significant contributor ($\beta = .287$, $p < .05$). The strongest significant model at 600 ms ($R^2 = .161$, $p < .05$) featured MOT as a significant predictor ($\beta = .404$, $p < .01$), along with Navon ($\beta = -.142$, n.s.), Orienting ($\beta = -.117$, n.s.), and age ($\beta = -.036$, n.s.). However, at 800 ms, we find a model that includes MOT ($\beta = .401$, $p < .01$), along with all other factors, that can account for over a third of the variability present in capacity measures ($R^2 = .343$, $p < .05$). This model indicates that, as predicted, the ability to track multiple objects is strongly associated with the capacity for audiovisual integration. It also suggests that the successful use of alerting signals in an attentional task, as well as an ability to disengage from conflicting cue locations, are associated with capacity increases, albeit non-significantly. Finally, it is interesting that as the display duration of the audiovisual integration capacity task increases from 200, through 600, to 800 ms, there is a concomitant increase in the predictive value of the regression model, from accounting for 12.1%, to 16.1%, to 34.3% of the observed variability. Our previous work (Wilbiks & Dyson, 2016) showed that at relatively fast (e.g. 200 ms) speeds of presentation, visual perceptual areas are unable to process the difference in incoming signals, and perhaps this is what is being manifested here as well.

In order to establish the unique contribution of each individual predictor, we calculated semipartial correlations for each predictor for each speed of presentation, with the results displayed in Table 5. Semipartial correlations can be interpreted in the same way as ΔR^2, as they indicate the amount of predictive value each predictor would give if they were the last ones added to a given regression model, which eliminates the conditional nature of the order in which items were entered into a model. Notably, across all four speeds of presentation, multiple object tracking was found to be a significant predictor, and no other predictors reached statistical significance. This is in agreement with what was found in the overall regression analyses, which showed that MOT was by far the strongest predictor at any stage. Additionally, this may suggest that the link between multiple object tracking and the main capacity task is more closely linked than had previously been thought. While no other predictors reached significance, Navon score and location conflict were shown to have semi-partial correlations that were not negligible, suggesting that they may be involved in predicting an individual's capacity. However, as they were not statistically significant, we must be cautious in our interpretation of these findings.

Discussion

The findings from Experiment 4 indicate that it is possible to account for a significant portion of the variability in audiovisual integration capacity between individuals by using a combination of other perceptual metrics. Multiple object tracking was found to be the most important predictor across all speeds of presentation, while other factors such as susceptibility to location conflict in an attentional cueing task, attentional alerting, and local precedence on a Navon task played smaller, and statistically non-significant roles in the regression equations. Additionally, we note that there is an increase in the amount of variability accounted for by the regression as we slowed the rate of presentation in the audiovisual integration capacity task. While at relatively fast speeds of presentation (200 and 600 ms), only 12.1% or 16.1% of the variation was accounted for, over a third of the variation can be explained at the slowest speed of presentation (34.3% at 800 ms). While this difference should be studied further, it is fair to relate this to the fact that the audiovisual integration task at the fastest speeds of presentation is exceedingly difficult, which led to previous research showing that capacity is strictly limited to one item (Van der Burg et al., 2013). More recently, work from our lab (Wilbiks & Dyson, 2016) showed that both behaviourally and electrophysiologically, there is an inability to successfully quantify incoming visual information at a speed of 200ms. If this is the case, it seems likely that a large amount of the variability in capacity is a function of strictly being able to process an incoming visual stream at such a high speed. Multiple object tracking is similar to the audiovisual integration task, such that it is highly sensitive to speed of presentation (Holcombe & Chen, 2013; Tombu & Seiffert, 2008; Liu et al., 2005), which is likely why it is one of the significant predictors here. We also see age play a non-significant role as a negative predictor, which is logical given the general decline of visual perception with age (cf. de Dieuleveult et al., 2017).

When the speed of visual presentation is decreased as far as 800ms, our model is able to account for a larger proportion of the variation in capacity. An additional finding of interest here is the difference in model that was found to be the strongest at each presentation speed. At 200 ms, the only significant model was the one that included age and multiple object tracking, which is in line with the extremely high level of perceptual load present in the task at that SOA. At slower speeds of presentation, the strongest models also included factors such as attentional factors and other measures. This fits well with our understanding of the task at slower speeds, which has a lower level of perceptual load, while still measuring the same core construct (audiovisual integration capacity). Across all presentation speeds, however, multiple object tracking remains the strongest predictor, along with local precedence on the Navon task, a lack of susceptibility to location conflict, and alerting. As far as this research shows, the ideal combination of traits to maximize one's capacity of audiovisual integration include: a high level of ability to track multiple objects, a tendency to focus on details before focusing on generality in a visual scene, an ability to disengage from misleading information within an attention task, and an ability to shift one's attention in response to exogenous attentional cues.

These findings can be taken forward in future research to further the understanding of the theoretical underpinnings of audiovisual integration. For example, in order to be able to integrate a number of visual stimuli with a tone, we must first be able to track a number of those stimuli successfully. Furthermore, we need to be able to disengage from earlier presentations of a display to then focus on subsequent displays that have stimuli changing in different locations. This whole task is also modulated by an individual's ability to monitor specific elements within a composite display, as well as by an individual's general audiovisual integration abilities, which may be indexed by age. These findings are mainly in agreement with previous research, as well as with theoretical perspectives on audiovisual integration and crossmodal attentional capture.

Reliability Assessment

With a recent increase in research examining individual differences on perceptual and/or cognitive tasks, there is a need to ascertain the reliability of the tasks being used. While reliability has long been an important factor in the design of questionnaires and similar evaluation tools, this has been less of a focus in cognitive paradigms. Recent work has revealed that even well-known, robust cognitive tasks may not be highly reliable, and that this may be an issue in terms of using these tasks in correlational research (Hedge, Powell, & Sumner, 2018). Some of the tasks employed in the current research have been tested and found to have high reliability (e.g. multiple object tracking *r* = .96; Huang, Mo, & Li, 2012), while others have been found to be less reliable (e.g. Navon task *r* = .17; Hedge et al., 2018). In order to examine the test-retest reliability of the tasks in the specific iterations used in the current research, this additional experiment was conducted. We wish to be completely transparent in stating that this reliability analysis was conducted *post hoc* based on a reviewer's constructive criticism - that is to say, we had already conducted, analysed, and interpreted Experiments 1-4 *before* we ran this assessment, and that any discussion of reliability in the interpretation of the results is based on this timeline.

Method

Each of the tasks used in the current research, with the exception of the audiovisual integration capacity task, were employed again in this reliability analysis experiment. 58 new participants were recruited, none of whom had participated in the earlier experiments in this series. The average age of these participants was 21.6 years (SD = 7.7), with 46 females and 12 males, and 6 participants reporting being left-handed. Each of these participants was asked to complete the Attention Network Test, the Mackworth Clock Task, the Corsi task, the Navon task, and the multiple object tracking task, based on the same experimental details as listed above. Each participant completed the battery of tasks twice, within a single testing session. We also analyzed the reliability of the audiovisual integration capacity task by splitting the first 128 trials from the last 128 trials as completed by participants in Experiment 4 and analyzing them in the same way as described below.

Results

Scores for each of the measures were tabulated in the same way as in the experiments described above. Scores for the two repetitions of each task were analysed for reliability by obtaining the intraclass correlation coefficients, using a two-way random effects model for absolute agreement (ICC (2,1); as per Hedge et al., 2018). The ICCs, along with 95% confidence intervals and tests of significance against a norm of 0, are in Table 6. Using the typical interpretations of ICC values put forth by Koo & Li (2016), we evaluate each of our measures as having poor (ICC < .5), moderate (ICC > .5), good (ICC > .75), or excellent (ICC > .9) reliability. Multiple object tracking, ANT Flanker Congruency, and capacity at 600 ms were the only tasks found to have good reliability. ANT Validity, the Corsi task, and capacity at 200, 400, and 800 ms were found to have moderate reliability. All our other measures were found to have poor reliability. However, while their ICC value put them in the ‘poor’ category, we note that ANT Alerting, ANT Orienting, Mackworth task, and the Navon task all had ICC values that were significantly greater than 0.

In considering whether the reliability of the tasks may have affected our findings, it is important to examine the relative reliability for the factors that played an important role in our regression models. While we find multiple object tracking (ICC = .852) and ANT Flanker Conflict (ICC = .782) to have good reliability, the Navon task (ICC = .343) and the Location conflict effect (ICC = .113) both have poor reliability. The implications of this finding are discussed in detail in the following section.

Discussion

Our analysis of the reliability of the tasks we used in our study echoes the findings of Hedge and colleagues (2018) - namely, that perceptual and cognitive tasks that have been previously found to be robust in group-level analyses may have widely varying levels of test-retest reliability, which may make given tasks less appropriate for using within an individual differences context. While our analysis of reliability has taken place *post hoc*, future work in this field should work to establish which tasks are appropriate for use in this type of analysis, and efforts should be made to develop novel tasks for testing these constructs that have moderate to high levels of reliability.

In an individual differences study such as this one, it would be ideal to have a number of tasks being used that have similar levels of reliability. Having a higher level of reliability is related to the potential for a task to correlate with others, and as such there may be a confound between the reliability of a task and its correlation with other tasks being employed. According to Hedge and colleagues (2018), the observed correlation between two measures is attenuated by their reliabilities. That is to say, with lower reliability measures, the correlation that we observe is actually less than the true correlation between the two measures being compared. Due to this phenomenon, it is possible that correlations (and therefore also regressions) conducted with reliable measures are more likely to reach statistical significance because those measures are more likely to yield higher effect sizes. Indeed, the MOT task that was found to be the strongest contributor to all of the regression models was one of the highest reliability tasks (ICC = .852), which could call into question what is driving this effect - true correlation, or higher reliability.

Conversely, a task such as the Navon may actually have a relatively high true correlation with capacity, which is not observed due to the limiting factor of reliability. Hedge and colleagues (2018) describe a method for producing disattenuated correlation coefficients based on the calculations of Spearman (1904). By following the equation:

$$\text{“}true\text{” } correlation\ (x, y)\ =\ \frac{Sample\ correlation\ (x, y)}{\sqrt{reliability(x) \cdot reliability(y)}}$$

one can produce an estimate of correlation if both tasks were reliable. While this is cannot be used inferentially, it is interesting to note that certain predictors in this research have relatively high correlations when the effect of reliability is removed. For example, Navon scores which had observed *r*s of between .214 and .324 in Experiment 4 (based on semi-partial correlations) have disattenuated *r*s between .393 and .698. As such, future research must follow Hedge and colleagues’ (2018) instructions to use only highly reliable measures in individual differences research.

While attenuation through low reliability may be implicated in some of the unexpected results we observed, we also note that we found several tasks with moderate or good reliability that did NOT correlate significantly with audiovisual integration capacity (i.e. ANT Flanker Conflict: ICC = .782; Corsi task: ICC = .576). Given that these tasks had relatively high reliability and did not correlate suggests that a simple relationship between reliability and correlation is not able to explain what we are observing. While there is still the danger that some tasks that we had expected to correlate may not be correlating due to their *low* reliability, this is something that could be followed up on in future research by first establishing reliable measures of these constructs, and then conducting a similar study to the one reported in Experiment 4.

General Discussion

The current research used a series of four experiments to determine which tests of perceptual and cognitive processes could be associated with the capacity of audiovisual integration. The tests that elicited significant correlational results were then combined in a fourth experiment to further investigate their ability to account for individual variability in capacity. Experiment 1 established support for the connection between multiple object tracking and the capacity for audiovisual integration. As Pylyshyn and Storm (1988) had previously found that objects moving at a relatively low speed (<9.4°/s) led to the ability to track four or more items, when we applied a similar MOT task in our research, the results that we found were supportive of their findings. Contrary to our expectations, there did not appear to be as strong a connection between visual working memory and AVI capacity as initially predicted, despite Irwin's (1992) study linking visual working memory with perceptual abilities and Oksama and Hyönä's 2004 research exploring the link between MOT capacity and other visual working memory. Although Kyllonen and Christal (1990) discovered a link between visual working memory and higher-level abilities such as reasoning, visual working memory did not appear to correlate with the capacity of audiovisual integration.

Fan and colleagues (2009) highlighted the importance of attention as the basis for numerous control systems, and Experiment 2 led to another unexpected discovery, as it was found that orienting does not play a large role in the capacity to predict audiovisual integration as previously expected. The alerting system, which they previously found to be closely linked to higher order cognitive processes (Fan et al., 2008), was more closely connected to predicting audiovisual integration. However, similar to Van der Stoep and colleagues (2015), we discovered that the predictability of the location of a stimulus and the appearance of invalid cues led to significant differences in the ability to integrate audiovisual information. Participants who were unable to ignore invalid cues show a significantly lower capacity for AVI than those who have less difficulty with flanker conflict and invalid cues. The third experiment found weak correlations between the level of vigilance, as assessed by the Mackworth Clock Test, and the capacity of audiovisual integration. The 1948 study found a significant amount of increases in missed cues as the duration of the task increased, seeming to signify a reduction in visual perception; our study showed weak correlations at best regarding the link between readiness and capacity. However, the Navon Task implemented in this experiment produced significant correlations between global precedence and capacity at the 600ms and 800ms time intervals, but the correlations were not statistically significant at the 200ms and 400ms intervals. This is similar to the findings of Wilbiks and Dyson (2016), as they had previously shown that the ability to perceive and integrate increases when slower intervals are used. Navon's (1977) finding that increased speed comes from observing the large details of a stimuli before the smaller, more complex characteristics indicates a wider field of focus and correlated significantly with the capacity for audiovisual integration, illustrating the importance of global precedence.

Having established perceptual and cognitive abilities that are related to audiovisual integration capacity, Experiment 4 combined the measures of unimodal perception, attention and focus and established a model that is predictive of an individual's capacity for audiovisual integration. Although multiple object tracking was found to be the most predictive of capacity at all speeds of presentation, susceptibility to location conflict, attentional alerting and local precedence all increased predictability, particularly at slower speeds of presentation. In fact, at the slowest speed of presentation, 34.3% of variability in capacity could be accounted for, compared to only 12.1% at 200 ms. The results of the current research show an increase in capacity at slower time intervals which contrasts with the position supported by Van der Burg and colleagues (2013) that audiovisual integration capacity is limited to one item but corresponds appropriately with Wilbiks and Dyson's (2016) research asserting the difficulty of successfully processing incoming visual stimuli at speeds as fast as 200ms. The fourth experiment in the series also shows that age is a negative factor in predicting capacity, as there is a general decline in visual perception as we age, although older individuals in general do combine more multisensory stimuli – but this often takes the form of errant integration.

Given the importance of the temporal binding window for audiovisual integration, the findings of Zmigrod and Zmigrod (2015) could be applied in future research. They found that applying cathodal transcranial direct stimulation over the right posterior parietal area of the brain, they were able to narrow the temporal binding window by up to 30 percent (Zmigrod & Zmigrod, 2015). If one were to apply these findings to individual differences in audiovisual integration capacity research, it seems likely that capacity may increase, provided the difficulties could be accounted for by factors that influence the width of the TBW such as age and autism. Recent research addressing the effects of anosmia, or the loss of olfaction, suggests that that the absence of the sense of smell heightens an individual's ability to integrate auditory and visual stimuli and detect multisensory temporal asynchronies, due in part to a narrower temporal binding window (Peter et al., 2019). Additionally, those who had been born with anosmia demonstrated a greater ability to distinguish asynchronies than those with acquired sensory loss (Peter et al., 2019). The current research focuses on the importance of individual differences in the capacity for audiovisual integration. When comparing our perceptual findings to these findings, it appears as though our theory surrounding how our individual differences influence our perceptual processes is accurate.

The current research supports the notion that to best predict individual capacity for audiovisual integration we must test the ability to track multiple objects, the tendency to focus on details before generality, and the ability to ignore invalid cues and shift attention rapidly in response to exogenous cues. The strongest predictor (and the only individual predictor that was statistically significant) within the model was the ability to track multiple objects, and considering the content of the two tasks, this is not a surprise. The multiple object tracking task involves keeping track of a number of objects as they move around the screen, while the audiovisual integration capacity task involves tracking the *state* of eight objects as they change repeatedly on the screen. So, while one involves tracking movement and the other involves tracking state, both of these tasks involve monitoring a number of visual objects. Similarly, the tendency to focus on specific details before overall elements of a visual display was non-significantly associated with higher audiovisual integration capacities. This suggests that focusing on individual items led to success, and this may well be the case in a task of this type, especially when associated further with shifting visual attention rapidly. These factors each play a role in creating the relationship between unimodal perception and audiovisual integration capacity. It is also possible that the causality we have inferred is reversed - that an individual’s audiovisual integration capacity predicts their ability to track multiple objects, etc. However, we believe that it is more likely the case that these unimodal factors are required before multisensory processing can be observed.

The maximum amount of variability we were able to account for in this study was 34.3%, and as such future research should seek to continue to examine similar research questions, in order to ascertain whether more of this variability can be accounted for. Given the success of audiovisual integration in neurotypical participants who are capable of tracking multiple objects, focusing on details before generality, and ignoring invalid cues and shifting attention rapidly in response to exogenous cues, there are several disorders that would be particularly interesting to investigate further with regard to their interaction with audiovisual integration capacity. Autism spectrum disorder (ASD) (Ashwin et al., 2009; Boer et al., 2013; Stevenson et al., 2016; Stevenson et al., 2017; Joseph et al., 2009), major depressive disorder (MDD) (Golomb et al., 2009; Marazziti et al., 2010; Richardson, & Adams, 2018; Rock et al., 2014; Serafini et al., (2017) and attention deficit disorder (ADD) (Bijlenga et al., 2017; Ghanizadeh, 2011) are but a few of a host of other disorders that alter our ability to normally process sensory information. Individuals with ASD typically have a wider temporal binding window, and often fare quite poorly when processing audiovisual information. ASD has many connections to unimodal and multimodal perception, but the continuous nature of capacity testing could ultimately lead to a more nuanced level of detection, allowing AVIC to be used as an early alerting system to perceptual abnormalities. This alerting system would allow for earlier interventions for children with ASD, which typically leads to better long-term outcomes.

MDD is also of significant interest as the effects of depression on visual and auditory perception have not been extensively studied. Golomb and colleagues (2009) found that individuals with MDD experienced a decline in spatial suppression which enhanced motion perception for typically suppressed stimuli. They also found that the degree of spatial suppression increased along with the duration of an individual’s depression and did not decrease upon resolution of depressive symptomatology. This suppression led to superior results in a high contrast motion discrimination task, which suggests that depressed individuals could be better at tracking moving objects, and potentially audiovisual integration, than their neurotypical peers.

The cognitive process of attention is also suitable for further research. As noted earlier, our prediction regarding the importance of the orienting system in AVI appeared to be inaccurate, however the alerting system played a more significant role. Given the importance of the alerting system and attention, there are many ways we could manipulate this process within the audiovisual integration task to alter the ability to integrate multiple audio and visual stimuli together. The results of these manipulations could encourage more understanding of the complex nature of human perceptual and cognitive systems that may allow us to further progress research into a more applied form, such as the construction of improved alerting systems for healthcare monitoring within a clinical setting. Additionally, more manipulations regarding the attentional processes may prove to be beneficial in the formation of the early alerting system proposed earlier to diagnose symptomology of autism spectrum disorder. Overall, this research shows that over a third of the wide individual variation observed in audiovisual integration capacity estimates is associated with an individual’s performance on unimodal perceptual and/or cognitive tasks.

References

Ashwin, E., Ashwin, C., Rhydderch, D., Howells, J., & Baron-Cohen, S. (2009). Eagle-eyed visual acuity: an experimental investigation of enhanced perception in autism. *Biological Psychiatry*, *65*(1), 17-21.

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision, 9*(7), 9-9.

Bijlenga, D., Tjon-Ka-Jie, J. Y. M., Schuijers, F., & Kooij, J. J. S. (2017). Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. *European Psychiatry*, *43*, 51-57.

Boer, L. D., Eussen, M., & Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience*, *7*, 8.

Brand-D'Abrescia, M., & Lavie, N. (2008). Task coordination between and within sensory modalities: Effects on distraction. *Perception & Psychophysics, 70*(3), 508-15.

Cavanagh, P., & Alvarez, G. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*(7), 349-54.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. Doctoral Thesis at McGill University (Canada).

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity [Target article and commentaries]. *Behavioral and Brain Sciences, 24*, 87-185.

de Dieuleveult, A. L., Siemonsma, P. C., van Erp, J. B., & Brouwer, A. M. (2017). Effects of aging in multisensory integration: a systematic review. *Frontiers in Aging Neuroscience*, *9*, 80.

Donohue, S. E., Woldorff, M. G., & Mitroff, S. R. (2010). Video game players show more precise multisensory temporal processing abilities. *Attention, Perception, & Psychophysics*, *72*(4), 1120-1129.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience, 28*(16), 4183-4191.

Eayrs, J., & Lavie, N. (2018). Establishing individual differences in perceptual capacity. *Journal of Experimental Psychology*. *Human Perception and Performance,44*(8), 1240-1257. doi:10.1037/xhp0000530

Fan, J., Gu, X., Guise, K., Liu, X., Fossella, J., Wang, H., & Posner, M. (2009). Testing the behavioral interaction and integration of attentional networks. *Brain and Cognition,70*(2), 209-220. doi:10.1016/j.bandc.2009.02.002

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research, 203,* 381–389. doi:10.1007/s00221-010-2240-4

Ghanizadeh, A. (2011). Sensory processing problems in children with ADHD, a systematic review. *Psychiatry Investigation*, *8*(2), 89.

Golomb, J. D., McDavitt, J. R., Ruf, B. M., Chen, J. I., Saricicek, A., Maloney, K. H., ... & Bhagwagar, Z. (2009). Enhanced visual motion perception in major depressive disorder. *Journal of Neuroscience*, *29*(28), 9072-9077.

Green, J. J., & Woldorff, M. G. (2012). Arrow-elicited cueing effects at short intervals: Rapid attentional orienting or cue-target stimulus conflict? *Cognition*, *122*(1), 96-101.

Hagmann, C. E., & Russo, N. (2016). Multisensory integration of redundant trisensory stimulation. *Attention, Perception, & Psychophysics*, *78*, 2558-2568.

Hancock, P. (1986). Sustained attention under thermal stress. *Psychological Bulletin, 99*(2), 263-281.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166-1186.

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, *49*, 461-467.

Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision*, *13*(1), 1-19.

Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 414.

Hulleman, J. (2005). The mathematics of multiple object tracking: From proportions correct to number of objects tracked. *Vision Research, 45*(17), 2298-2903.

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 307.

Joseph, R. M., Keehn, B., Connolly, C., Wolfe, J. M., & Horowitz, T. S. (2009). Why is visual search superior in autism spectrum disorder?. *Developmental Science*, *12*(6), 1083-1096.

Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 336-358.

Kawakami, S., Uono, S., Otsuka, S., Zhao, S., & Toichi, M. (2018). Everything has its time: Narrow temporal windows are associated with high levels of autistic traits via weaknesses in multisensory integration. *Journal of Autism and Developmental Disorders*, 1-11.

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: standardization and normative data. *Applied Neuropsychology, 7*(4), 252-258.

Koldewyn, K., Weigelt, S., Kanwisher, N., & Jiang, Y. (2013). Multiple object tracking in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(6), 1394-1405.

Koo, T. K. & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163.

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in Integrative Neuroscience, 4*, 129. doi:10.3389/fnint.2010.00129

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence, 14*(4), 389-433.

Lavie, N. (2005). Load theory of selective attention and cognitive control. *Trends in Cognitive Science, 9*, 75-82.

Lichstein, K., Riedel, B., & Richman, S. (2000). The Mackworth clock test: A computerized version. *The Journal of Psychology, 134*(2), 153-61.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance, 31*(2), 235.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual research. *Quarterly Journal of Experimental Psychology, 1*, 6-21.

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology*, *626*(1), 83-86.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences,9*(6), 296-305.

Matusz, P. J. & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review, 18*(5), 904-909.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746.

Meyerhoff, H. S., & Gehrer, N. A. (2017). Visuo-perceptual capabilities predict sensitivity for coinciding auditory and visual transients in multi-element displays. *PLoS One*, *12*(9), e0183723.

Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, *79*, 1255-1274.

Miller, L., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience,25*(25), 5884-93.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353-383.

Oksama, Lauri & Hyönä, Jukka (2004) Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach, *Visual Cognition, 11*(5), 631-671, DOI: 10.1080/13506280344000473

Olivers, C. N. L., Awh, E., & Van der Burg, E. (2016). The capacity to detect synchronous audiovisual events is severely limited: Evidence from mixture modeling. *Journal of Experimental Psychology: Human Perception and Performance, 42*(12), 2115-2124.

Peter, M. G., Porada, D. K., Regenbogen, C., Olsson, M. J., & Lundström, J. N. (2019). Sensory loss enhances multisensory integration performance. *Cortex*, *120*, 116-130.

Posner, M., Walker, J., Friedrich, F., & Rafal, R. (1984). Effects of parietal injury on covert orienting of attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience,4*(7), 1863-74.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, *3*(3), 179-197.

Richardson, L., & Adams, S. (2018). Cognitive deficits in patients with depression. *The Journal for Nurse Practitioners*, *14*(6), 437-443.

Robison, M. K., & Unsworth, N. (2017). Individual differences in working memory capacity predict learned control over attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(11), 1912.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine, 44*(10), 2029-2040.

Serafini, G., Gonda, X., Canepa, G., Pompili, M., Rihmer, Z., Amore, M., & Engel-Yeger, B. (2017). Extreme sensory processing patterns show a complex association with depression, and impulsivity, alexithymia, and hopelessness. *Journal of Affective Disorders*, *210,* 249-257.

Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology, 23*(2), 150-155.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72-101.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology, 13*(13), R519-R521.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1517-1529.

Stevenson, R. A., Toulmin, J. K., Youm, A., Besney, R. M., Schulz, S. E., Barense, M. D., & Ferber, S. (2017). Increases in the autistic trait of attention to detail are associated with decreased multisensory temporal adaptation. *Scientific Reports, 7*(1), 14354.

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., Camarata, S., & Wallace, M. T. (2016). Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Research, 9*(7), 720-738.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096-1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24-31.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., & Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London B. Biological Sciences, 268,* 31-38.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098-1114.

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience and Biobehavioral Reviews, 61,* 208-224. doi:10.1016/j.neubiorev.2015.11.002

Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, *108*(1), 1-25.

Van der Burg, E., Awh, E., & Olivers, C. N. L. (2013). The capacity of audiovisual integration is limited to one item. *Psychological Science, 24*(3), 345-351.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053.

Van der Stoep, N., Van der Stigchel, S., & Nijboer, T. C. W. (2015). Exogenous spatial attention decreases audiovisual integration. *Attention, Perception, & Psychophysics, 77*(2), 464-482.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748-751.

Vroomen, J., & Gelder, B. D. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of experimental psychology: Human perception and performance*, *26*(5), 1583.

Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans. *Learning and Motivation, 14*, 406-432.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638-667.

Wilbiks, J. M. P., & Dyson, B. J. (2016). The dynamics and neural correlates of audiovisual integration capacity as determined by temporal unpredictability, proactive interference, and SOA. *PLoS ONE 11*(12), e0168304.

Wilbiks, J. M. P., & Dyson, B. J. (2018). The contribution of perceptual factors and training on varying audiovisual integration capacity. *Journal of Experimental Psychology: Human Perception and Performance, 44*(6), 871-884.

Wilbiks, J. M. P., Pavilanis, A. D. S., & Rioux, D. M. (2020). Audiovisual integration capacity modulates as a function of illusory visual contours, visual display circumference, and sound type. *Attention, Perception, & Psychophysics,* Advanced Online Publication. https://doi.org/10.3758/s13414-019-01882-6

Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex*, *17*(suppl_1), i118-i124.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, *12*(3), 219-224.

Yantis S. (1992) Multi-element visual tracking: attention and perceptual organization. *Cognitive Psychology* 1992; 24: 295-340

Zmigrod, S., & Zmigrod, L. (2015). Zapping the gap: Reducing the multisensory temporal binding window by means of transcranial direct current stimulation (tDCS). *Consciousness and Cognition*, *35*, 143-149.

Zmigrod, L., & Zmigrod, S. (2016). On the temporal precision of thought: individual differences in the multisensory temporal binding window predict performance on verbal and nonverbal problem solving tasks. *Multisensory Research* DOI: 10.1163/22134808-00002532.

Author's Note

J. M. P. W. was supported by a Margaret and Wallace McCain Postdoctoral Fellowship from Mount Allison University for portions of this research.

Open Practices Statement

The data for all experiments are available at https://osf.io/69gt7/.

Figure Captions

Figure 1. Means and standard errors (dots with error bars), along with individual data points (crosses) for each participant in Wilbiks & Dyson (2016). Regardless of stimulus parameters, a large degree of individual variability of capacity is present in each sample.

Figure 2. Schematic for visual stimuli presented in the main audiovisual integration capacity task in all experiments.

Figure 3. Capacity estimates for Experiments 1 to 4 as a function of stimulus onset asynchrony. Error bars depict standard errors; asterisks show statistically significant differences.

Figure 4. Scatter plots of significant correlations between multiple object tracking (MOT) and capacity estimates in Experiment 1 along with line of best fit and 95% confidence interval around the line of best fit.

Figure 5. Scatter plots of significant correlations between attentional cue validity and capacity estimates in Experiment 2 along with line of best fit and 95% confidence interval around the line of best fit.

Figure 6. Scatter plots of level of global precedence and capacity estimates in Experiment 3 along with line of best fit and 95% confidence interval around the line of best fit.

Tables

Table 1

Pearson correlations for capacity at each display duration with each other, and with visual working memory span (VWM) and multiple object tracking span (MOT). Correlations significant at p < .05 are indicated by bold text

		200 ms	400 ms	600 ms	800 ms	VWM	MOT
200 ms	Pearson's r	—					
	p-value	—					
400 ms	Pearson's r	**0.723**	—				
	p-value	**< .001**	—				
600 ms	Pearson's r	**0.780**	**0.833**	—			
	p-value	**< .001**	**< .001**	—			
800 ms	Pearson's r	**0.569**	**0.664**	**0.755**	—		
	p-value	**< .001**	**< .001**	**< .001**	—		
VWM	Pearson's r	0.179	0.127	0.180	0.088	—	
	p-value	0.223	0.389	0.221	0.550	—	
MOT	Pearson's r	**0.404**	**0.377**	**0.401**	0.255	**0.316**	—
	p-value	**0.004**	**0.008**	**0.005**	0.080	**0.029**	—

Table 2

Pearson correlations for capacity at each display duration with each other, and with subscores on the ANT-R. Correlations significant at p < .05 are indicated by bold text

		200 ms	400 ms	600 ms	800 ms	ALERT	VALID	FCon	LCon	ORIEN
200 ms	Pearson's r	—								
	p-value	—								
400 ms	Pearson's r	**0.665**	—							
	p-value	**< .001**	—							
600 ms	Pearson's r	**0.535**	**0.687**	—						
	p-value	**< .001**	**< .001**	—						
800 ms	Pearson's r	**0.656**	**0.711**	**0.669**	—					
	p-value	**< .001**	**< .001**	**< .001**	—					
ALERT	Pearson's r	0.003	0.067	0.013	0.054	—				
	p-value	0.986	0.672	0.937	0.735	—				
VALID	Pearson's r	**-0.331**	-0.289	**-0.398**	**-.366**	-0.040	—			
	p-value	**0.032**	0.063	**0.009**	**0.017**	0.803	—			

F Con	Pearson's r	-0.222	**-0.318**	**-0.359**	**-.338**	-0.073	-0.038	—		
	p-value	0.157	**0.040**	**0.020**	**0.029**	0.647	0.814	—		
L Con	Pearson's r	0.034	-0.005	0.089	0.052	-0.263	-0.232	-.104	—	
	p-value	0.833	0.977	0.575	0.741	0.092	0.140	0.512	—	
ORIEN	Pearson's r	-0.119	-0.198	**-0.313**	-.256	-0.100	0.268	0.280	0.185	—
	p-value	0.452	0.209	**0.044**	0.102	0.529	0.086	0.073	0.240	—

Table 3

Pearson correlations for capacity at each display duration with each other, with oddball detection rates on Mackworth Clock Task, and global precedence on the Navon task. Correlations significant at p < .05 are indicated by bold text.

		200 ms	400 ms	600 ms	800 ms	MCT	Navon
200 ms	Pearson's r	—					
	p-value	—					
400 ms	Pearson's r	**0.559**	—				
	p-value	**< .001**	—				
600 ms	Pearson's r	**0.417**	**0.653**	—			
	p-value	**0.007**	**< .001**	—			
800 ms	Pearson's r	**0.363**	**0.597**	**0.818**	—		
	p-value	**0.020**	**< .001**	**< .001**	—		
MCT	Pearson's r	0.228	0.152	0.145	0.184	—	
	p-value	0.152	0.344	0.365	0.249	—	
Navon	Pearson's r	0.214	0.185	**0.311**	**0.324**	0.065	—
	p-value	0.179	0.246	**0.048**	**0.039**	0.686	—

Table 4

Descriptive statistics for the factors entered into the hierarchical linear regression in Experiment 4.

Measure	M	SD
MOT	3.970	0.693
Navon	-1.484	77.890
Alerting	39.921	51.371
Validity	109.125	59.443
Flanker Conflict	109.987	125.551
Location Conflict	0.882	44.410
Orienting	94.974	73.681

Table 5

Partial Correlation Table for Predictors of K at Each Display Duration

	200 ms			400 ms			600 ms			800 ms		
Variable	SPC	*p*	*	SPC	*p*	*	SPC	*p*	*	SPC	*p*	*
MOT	.099	.015	*	.071	.042	*	.142	.003	**	.139	.003	**
Navon	.034	.144		.017	.307		.012	.372		.031	.138	
Alert	.005	.558		.006	.543		.001	.769		.015	.301	
Validity	.001	.833		.025	.223		.013	.347		.004	.594	
F Conflict	.000	.985		.010	.427		.003	.646		.001	.805	
L Conflict	.008	.483		.012	.392		.030	.165		.037	.106	
Orienting	.000	.956		.001	.867		.003	.655		.008	.439	
Age	.013	.366		.008	.481		.003	.651		.004	.573	
Gender	.036	.134		.060	.061		.032	.148		.049	.064	
Handedness	.020	.267		.000	.967		.021	.247		.025	.187	

Note. SPC = Semi-partial correlation; this value is the equivalent of ΔR^2 for each predictor if it were the last variable added to a regression equation containing all of these predictors.

* p < .05; ** p < *.01;* *** p < .001.

Table 6

Intraclass correlation coefficients for each measure used in this study, as assessed post hoc

Task	ICC	95% CI	*p*	*
Capacity 200 ms	.586	[.390, .731]	< .001	***
Capacity 400 ms	.645	[.467, .773]	< .001	***
Capacity 600 ms	.774	[.647, .859]	< .001	***
Capacity 800 ms	.629	[.447, .762]	< .001	***
ANT Alerting	.369	[.123, .571]	.002	**
ANT Validity	.625	[.431, .760]	< .001	***
ANT Orienting	.295	[.041, .512]	.012	*
ANT Flanker Conflict	.782	[.657, .865]	< .001	***
ANT Location Conflict	.113	[-.148, .359]	.197	
ANT F x L Congruency	-.138	[-.381, .122]	.852	
Mackworth Clock Task	.274	[.017, .498]	.019	*
Corsi Task	.576	[.373, .726]	< .001	***
Navon Task	.343	[.092, .552]	.004	**
MOT Task	.852	[.759, .911]	< .001	***

Note. ICC = Intraclass correlation coefficient (2,1)

* *p* < .05; ** *p* < .01*;* *** *p* < .001.