Cognitive Processes in the Perception of Actions and Agents

By

Lewis John Baker

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
in
Psychology
August, 2016
Nashville, Tennessee

Approved:	Date:
Daniel T. Levin, PhD	5/18/2016
Brandon A. Ally, PhD	5/18/2016
John J. Rieser, PhD	5/18/2016
Megan M. Saylor, PhD	5/18/2016

To my friends, family and colleagues. Scientists stand on the shoulders of giants, but this dissertation was carried on the backs of a hundred human beings.

ACKNOWLEDGEMENTS

This dissertation was completed under the careful guidance of Dan Levin and Meg Saylor, to whom I am forever indebted for their open-minded and kind mentorship. Many of the ideas presented throughout this paper were generated through discussions with my committee members, John Rieser and Brandon Ally, nearly two years ago during my qualifying exam, and I can only hope to carry their insights forward through my career.

This work would not have been possible without the financial support of a Graduate Research Fellowship from the National Science Foundation, as well as equally generous funding from Peabody College.

It cannot be stated enough how much I have relied on the community of scientists, thinkers and otherwise good people at Vanderbilt University. Much of this document, but particularly Chapter 1, stems from preliminary research conducted by Stephen Killingsworth and Alicia Hymel. I owe the entirety of my statistical knowledge to Jim Steiger and Kris Preacher, the latter of whom I owe a great deal of thanks for corralling me into the quantitative methods doctoral minor. I am also indebted to the weekly meetings with Dan, Meg, Chris Jaeger, Sofia Jimenez, Jonathan Lane, Maria Osina and Claire Weaver, where I was permitted an embarrassingly large proportion of time to discuss my ideas. I am thankful for undergraduate assistants Jillian Allard, Anne Havard, Anna Isaacs, Chris Jaeger, Michael Schoenfield and Lionel Pomerantz for their assistance in data collection. Finally, I am forever appreciative of Danica Alexander, Kao-Wei Chua, Leandra Desjardins, Holly Gardner, Sergio Andres Lobo-Navia and Jennifer Yih for contributing in one way or another to the ideas presented below, but mostly for their steadfast friendship.

TABLE OF CONTENTS

Chapter

LIST OF TABLES

LIST OF FIGURES

Figure Page

INTRODUCTION

Human beings are bombarded with information from the sensory world. As an information processor with limited capacity, the human perceptual system must selectively allocate resources to relevant stimuli while ignoring the trivial. The spectacular functionality of the perceptual system leaves us with the impression of a vast and complete sensory experience, where our knowledge of the objects and actions about us is only limited by the physical constraints of our eyes and ears. However, experiments reveal the depth of our subjective experience is an illusion. We readily fail to detect enormous changes in our environment (Levin & Simons, 1997; Simons & Levin, 1998; Rensink, 2010), neglect objects at or near the center of fixation (Henderson & Hollingworth, 1999; Smith, Lamont & Henderson, 2012), and forget items in our own hands (Triesch, Ballard, Hayhoe and Sullivan, 2003). Although many cognitive scientists have studied the failures of attention and memory, relatively few have investigated the environmental and task factors that facilitate cognition in everyday situations: why we pay attention to select items in the real world and ignore others. Here I present three studies that investigate how the perception of actions and agents guides attention and memory. Chapter 1 investigates the limits of attention and memory when tracking multiple events, Chapter 2 explores constraints to the rapid perception of another agent’s visual access, and Chapter 3 tests how social agents guide visual search.

The relatively young field of event perception offers insight into the cognitive processes governing the moment-to-moment representations that give rise to conscious awareness. Research in event perception demonstrates that individuals parse continuous activity into discrete units determined by relatively few but informative dynamic properties in an event, such as global shifts in spatial relationships or more local changes in goals and intentions. This contrasts with inattention research suggesting that people represent relatively few details in a scene (Rensink, O’Regan & Clark, 1997; Simons & Rensink, 2005) and do not readily recall those details that are encoded (Simons, Chabris, Schnur & Levin, 2002; Mitroff, Simons & Levin, 2004; Varakin, Levin & Collins, 2007). The perception of new events at specific points of transition suggests that certain properties receive preferential encoding even as inattention research demonstrates that the majority of stimulation may fall beneath awareness. The prevailing theory of event perception, Event Segmentation Theory (EST; Zacks et al, 2007), proposes that individuals cyclically perceive, encode and anticipate events through an ongoing representation of the immediate context in working memory. These *event models* are selective representations of specific features relevant for comprehending ongoing behavior. Research indicates that viewers encode properties into event models at the boundaries between discrete events (Baker & Levin, 2015), and rapidly forget information from completed events (Morrow, Bower & Greenspan, 1989; Radvansky, Tamplin & Kraweitz, 2010; Swallow et al, 2011). Researchers have proposed that an event model is a representation of "what is going on now" in working memory (Zacks et al, 2007) that is abstracted at event boundaries to form the units of long-term episodic memory (Ezzyat & Davachi, 2011). Event perception as described by EST involves a system that selectively attends to statistically relevant properties, ignores irrelevant information, facilitates awareness to immediate contextual information and determines the formation of long-term memory, all while operating efficiently enough to permit planning and coordination during everyday activity. However, no experiment has empirically has demonstrated how individuals monitor multiple events or the effect of task load on awareness of ongoing events. Chapter 2 measures the effect of viewing multiple events on event perception using a novel measure adapted from Hymel, Baker & Levin (2015), indicating that event perception requires attentional focus and working memory.

Recent evidence suggests that social stimuli may guide early visual processing. Developmental psychologists have long considered theory of mind (the ability to think about the thoughts, feelings, beliefs and desires of someone else; Premack & Woodruff, 1978) as a relatively late developing (Flavell et al, 1981) and cognitively taxing process (Lin, Keysar & Epley, 2004). However, emerging research suggests that adults (Samson et al, 2011) and infants as young as 5 months (Kovacs, Teglas & Endress, 2010) can rapidly judge whether an agent has limited visual access to a scene. In a recent study, we confirmed that this "early system of perspective taking" is not due to low-level attentional cues, and likely stems from a rapid computation of an agent's line-of-sight access to a target of attention (Baker, Levin & Saylor, 2016). However, the exact circumstances that give rise to this default perspective taking are still largely unknown. I expand upon these findings in Chapter 2 to investigate whether the default calculation of another's perspective extends to complex situations featuring multiple agents or targets of attention. Results suggest a much more limited system than previously hypothesized that signals when the observer has privileged access to an attended set of objects.

The final chapter of this dissertation extends the study of actions and agents to investigate how social stimuli shape the deployment of visual attention. The results of Chapter 2 suggest some form of rapid but limited computation of another's visual perspective. Several researchers have proposed that this system operates by selectively guiding attention to social stimuli (Leslie, Freidman & German, 2005; Apperly & Butterfill, 2009), but there has been no empirical demonstration that social agents alter the deployment of visual attention. Chapter 3 tests how the presence of another agent alters visual search, leading to the somewhat surprising discovery that social agents bias visual search to regions an agent cannot access.

The three studies in this dissertation explore the cognitive processes that facilitate for the perception of actions, agents and the targets of joint visual access. Together, these studies explore the control processes involved in awareness of scenes, events and social interaction, and further our understanding of one of the most basic questions in the history of psychology: Why do we attend to some things and ignore others?

CHAPTER 1

DECREASED AWARENESS FOR PARALLEL EVENTS REVEALS LIMITED CAPACITY OF EVENT PERCEPTION

Abstract

Event segmentation theory (EST) proposes that individuals comprehend ongoing activity by selectively attending to relevant features and holding those features in working memory for the duration of the event. However, no research has explored how individuals monitor multiple events, or comprehend activity while under cognitive load. This study tests whether tracking multiple events taxes impairs the depth of event perception. Participants detected fewer causally impossible actions when viewing two events in parallel, suggesting a limited ability to encode and compare properties during event perception. Participants detected fewer target actions in parallel events when cued to attend to the target-containing sequence and ignore the distractor, although cuing partially mediated failures to detect target actions. Finally, working memory load from a secondary task also decreased target detection when viewing a single sequence of events. These results confirm that event perception relies on a limited working memory capacity that is further constrained by the focus of attention.

Introduction

Even a relatively demure event, like drinking your morning coffee, contains an overwhelming amount of sensory information. It is an astounding feat of the perceptual system that we not only make sense of dynamic activity, but that our comprehension of events generally agrees with other observers. Studies indicate that individuals generate models of a situation as it unfolds, preferentially attending to and encoding the spatio-temporal relationships, agents, goals, intentions and causes that best inform event comprehension (Zacks, Speer & Reynolds, 2009; Kives, Ware & Baker, 2015). Researchers have suggested that these *event models* (sometimes referred to as situation models; Zwaan & Radvansky, 1998) are maintained in working memory and refreshed at the start of each new event (Zacks, Speer, Swallow, Braver & Reynolds, 2007; Swallow, Zacks & Abrams, 2009; Swallow et al, 2011). However, theories of event perception have yet to demonstrate how a single event model could possibly process the spontaneity of everyday life, which often requires attention to two or more disparate events. This study therefore begins with a simple question: What happens when viewers must monitor two events in parallel? Here, I demonstrate that tracking two events greatly reduces the encoding of details within each event. In five studies, I demonstrate that failures of event perception arise through a combination of divided attention and working memory capacity. Although many researchers have suggested that event perception requires attention and working memory, this is the first study to reveal failures of event perception when these domain-general resources are taxed.

The Limited Capacity of Event Perception

Everyday events contain vast amounts of information. Researchers throughout the history of psychology have theorized how the mind selects relevant information from the stream of consciousness and ignores that which is irrelevant (James; 1904; Koffka, 1922; Wertheimer, 1922; Gibson, 1966; Johannsson, von Hofsten & Jansson, 1980). However, it was only relatively recently that Newtson (1973) proposed studying dynamic events as discrete units in time. When Newtson simply asked participants to press a key when they believed a new event occurred in a film, he found that participants not only segmented temporal information into discrete units, but that they did so reliably and in good agreement across raters (Newtson & Enquist, 1976; Newtson, Rindner, Miller & LaCross, 1978; Newtson, Hairfield, Bloomingdale & Cutino, 1987). The ability to reliably parse continuous information into coherent units suggests a system that periodically and selectively encodes only features necessary for comprehension. Research has demonstrated the segmentation of events hinges on attention to several specific changing properties, such as spatial relationships, agents and goals (Zwaan & Radvansky, 1998; Schwann & Garsoffsky, 2004; Swallow, Zacks & Abrams, 2009; Swallow et al, 2011; Huff, Papenmeier & Zacks, 2013). These specific transitions not only signal the onset of a new event, but also trigger increased comparison of features in working memory (Baker & Levin, 2015), determine the features encoded in long-term memory (Ezzyat & Davachi, 2011) and signal the rapid forgetting of irrelevant information from previous events (Radvansky, Tamplin & Kraweitz, 2010). Such results suggest that event perception requires selective attention to relevant properties and an ability to hold and compare properties in recent memory to perceive a complete unit of time.

The prevailing theory of event perception, Event Segmentation Theory (EST), has attempted to model the factors that influence event perception. In general, the theory states that visual events are perceived in a hierarchical fashion, nesting primitive actions (e.g. motion or spatial displacement) within overarching conceptual events (defined by goals, schemas, patterns or previous experience). The central component of EST is the event model, a working memory representation built from multimodal input, including visual, textual and auditory information (Baggett, 1979; Magliano, Miller & Zwaan, 2001; Zacks, Speer & Reynolds, 2009). Event models contain perceptual information of objects and spatial relationships, as well as goal state information of where an object has been and where it is most likely to be next (Spapé & Hommel, 2010). This latter information is used to form perceptual predictions that frame sensory input and assist in comprehension (Reynolds, Zacks & Braver 2007). EST states that perceptual predictions are generated in part through bottom-up inference of causality and trajectory (Reynolds, Zacks & Braver, 2007; Zacks, 2004; Zacks, Swallow, Speer, Vettel & McAvoy, 2006; Hard, Tversky & Lang, 2006; Zacks, Kurby, Eisenberg & Haroutunian, 2011). Meanwhile, observers also track information pertaining to agent goals, including their immediate physical course (Baird & Baldwin, 2001; Baldwin, Baird, Saylor & Clark, 2001; Baldwin, Andersson, Saffran & Meyer, 2008), but also relatively long term intentions (Robertson & Suci, 1980; Wegner & Guiliano, 1983; Morrow, Bower & Greenspan, 1989; Wilson, Rinck, McNamara, Bower & Morrow, 1993; Magliano, Taylor & Kim, 2005). Future events are further extrapolated using prior experience and deductive inference (Cohen & Ebbeson, 1978; Ebbeson, 1980; Zacks et al, 2007; Magliano & Zacks, 2011; Radvansky, Tamplin, Armendarez & Thompson, 2014). Thus, event models contain a select but still quite substantial subset of the total information available for the comprehension of ongoing activity.

Event models have three related functions. First, event models are representations in working memory containing information relevant to the immediate context. Event models thus facilitate increased awareness of relevant objects and actions for the duration of the event (Hanson & Hirst, 1989; Lassiter & Slaw, 1991; Zacks & Tversky, 2001; Schwan & Garsoffsky, 2004; Speer & Zacks, 2005; Kurby & Zacks, 2008; Radvansky, Tamplin & Krawietz, 2010; Zacks et al, 2011; Swallow et al, 2011; Kurby & Zacks, 2012; Sargent et al, 2013). Second, event models guide perception to predict future behavior (Reynolds, Zacks & Braver, 2007; Zacks, Kurby, Eisenberg & Haroutunian, 2011). Models act as the priors of each event, guiding perception to subsequent activity through low-level causal perception (Cutting, 1981; Strickland & Keil, 2011) or through experience-driven prediction (Cohn & Paczynski, 2013; Swallow, Zacks & Braver, 2009). Lastly, event models are basic units of long-term memory (Ezzyat & Davachi, 2010; Staresina & Davachi, 2009). Information present at the event boundary is the most likely to be stored in long-term memory (Newtson et al, 1987; Boltz, 1992; Davachi, 2006), indicating that episodic memory is less like continuous video playback, and more like discrete panels in a comic strip (illustrated – literally – in Cohn, 2013).

Event perception as explained by EST is an expansive control network that not only facilitates attention to changing properties, but generates near-future predictions based on calculations from immediate perception and previous experience, all while updating information in working memory. Event perception theoretically occurs automatically, or at least must operate efficiently enough that perceivers can comprehend ongoing events while engaging in any number of cognitively demanding tasks, such as planning actions or engaging in conversation. This is additionally impressive given the known limits of attention and working memory. Furthermore, EST assumes viewers are tracking, updating and forming predictions for one event at a time, using a single expansive event model. The complexity of EST makes the capacity limits of event perception unclear, and it is uncertain what the consequences of monitoring multiple events may be. The current study investigated the cognitive costs of event perception by testing participant awareness of two events viewed in parallel. Using a variety of experimental designs, I show that failures of event perception arise through the limits of attention and working memory.

Potential Constraints on Event Perception

This study tested whether tracking multiple events impairs event perception. This relatively simple question informs our understanding of the cognitive components necessary to comprehend ongoing behavior. Participants either viewed a sequence of actions (e.g., a man replacing flashlight batters) in isolation or in parallel with another sequence in an ABAB fashion (e.g., switching between the flashlight event and an entirely separate event with a different man using a screwdriver). They then responded whether a shot from either sequence appeared out of sequential order (e.g., using the screwdriver before picking it up).

Several hypotheses can be generated as to how the perceptual system responds to multiple ongoing streams of activity. First, it is not a foregone conclusion that viewing two events would hinder event perception at all. Evidence suggests that event perception relies on the detection of critical transitions that denote the initiation of new events. Crucially, event perception remains more or less unperturbed so long as these event boundary properties exist (Cohn, 2013). Filmmakers have long observed that shortening an event through editing does not alter viewer understanding of that event (Cutting, Delong & Nothelfer, 2010), and that editing an event from multiple viewpoints does not alter event perception (Schwan, Garsoffky & Hesse, 2000). In fact, displaying multiple events in parallel is common practice in most conventional films, and individuals with no prior experiment with film comprehend events presented in parallel (Schwan & Ilidrar, 2010). Perhaps most strikingly, the temporal order of events does not appear to alter event perception: participants viewing an unfamiliar event (e.g., constructing a saxophone) normally or in reverse segmented an event at the exact same locations, regardless of their prior knowledge (Hard, Tversky & Lang, 2006). Such findings reinforce that event perception is grounded in the perception of specific objects and actions. It is possible that the perceptual system detects the specific properties indicative of new events and uses top-down inference to fill in the gaps, and that the same properties will be encoded regardless of the number of events viewed. Experiment 1 addresses this possibility by comparing event perception when viewing a single ever versus two events in parallel.

However, events are not viewed as a series of static images, and likely rely on the dynamic temporal relationships linking objects and actions. Nonetheless, event perception might operate by detecting the relatively specific transitions indicative of new events, such as low-level motion (Scholl & Tremoulet, 2000; Scholl & Nakayama, 2001; Smith & Santacreu, 2016**)** or spatial cues (Meyerhoff et al, 2011; Huff & Schwann, 2012; Baker & Levin, 2015). This *perceptual continuity* hypothesis would predict that viewing events in parallel would disrupt event perception by severing the continuity from one moment to the next. Experiment 2 specifically addresses this low-level explanation for event perception by testing event perception for pairs of actions in parallel and single events. If event perception relies on perceptual continuity from one moment to the next, then event perception should be unimpeded so long as the perceptual flow is not interrupted between views of an event.

An alternative to the perceptual continuity hypothesis is that event perception requires unbroken attention to a stream of activity, and that dividing attention between two events would reduce the amount of detail encoded during event perception. Phenomena such as inattentional blindness (Mack & Rock, 1998), conceptual masking (Potter, 1976) and the attentional blink (Raymond, Shapiro & Arnell, 1992) suggest that splitting attention between items reduces perceptual awareness. Given that the perception of new events captures attention (Raisig, Wilke, Hagendorf & van der Meer, 2010; Mital, Smith, Hill & Henderson, 2011; Huff, Papenmeier & Zacks, 2012), an *attentional shift hypothesis* would predict that rapidly shifting attention between events would reduce the quality of information encoded. Experiments 3 and 4 test the perception of parallel events when participants are instructed to attend to a target sequence and ignore an irrelevant sequence. The attentional shift hypothesis would predict improved quality of event perception when attention is directed to the target sequence.

On final alternative would posit that event perception requires an active maintenance of recent information in working memory. Event perception is more than just the attention to features and transients, as comprehension often requires knowledge of long-term goals, mental states and spatial relationships (e.g., Magliano, Taylor & Kim, 2005). Researchers have consistently found relationships between the perception of new events and the contents of working memory (Speer & Zacks, 2005; Swallow, Zacks & Abrams, 2009; Radvansky, Tamplin & Krawietz, 2010; Swallow et al, 2011), and the reliability with which individuals segment and remember events is significantly correlated with their performance on working memory tasks (Sargent et al, 2013). Therefore, a *working model hypothesis* would posit that event perception requires the encoding and comparison properties in working memory. Individuals compensate for the limited capacity of working memory by only encoding a select subset of the information available during event perception. Individuals monitoring two events would therefore encode fewer features for each individual event as working memory reaches capacity. Experiment 5 directly tests the effect of working memory load on event perception.

Measuring Event Perception

Most tests of event perception have focused on the segmentation of continuous activity into discrete events (e.g., Newtson & Enquist, 1976; Zacks, Tversky & Iyer, 2001) or have measured memory of properties from recent events (e.g., Swallow, Zacks & Abrams, 2009; Ezzyat & Davachi, 2011; Swallow et al., 2011). However, these designs are insufficient to measure the quality of encoding during online event perception. For instance, segmentation behavior does not reveal the level of detail encoded when a new event begins, or participant's confidence in their judgments. Likewise, memory paradigms measure recall of properties within an event but do not measure participant's on-line perception of the entire sequence of actions surrounding those properties. For example, a participant could remember that an event contained a red car even though they failed to perceive that the red car ran a stop sign. Furthermore, differences in memory paradigms may be attributable to errors in recall rather than encoding. The current study therefore adapted a task from Hymel, Levin & Baker (2015) that measured on-line perception of the causal sequence of events.

Hymel, Levin & Baker asked participants to watch a rapid sequence of actions demonstrating a single common event, where each step in the event was shown in a new shot. Critical trials reversed the order of two actions, effectively showing an impossible event. For example, a misordered coffee event showed a woman picking up a coffee pot, pouring coffee, reaching for a sugar packet, pouring sugar, stirring the coffee *and then* reaching for a coffee stirrer. The ability to detect this causal misordering required participants to perceive each action as well as that action's temporal relationship with previous actions. Participants found the task challenging: participants rarely detected misordered actions in an incidental task, and performed worse during an explicit task when under cognitive load from a secondary verbal overshadowing task. The misordering-detection task reveals that event perception is capacity-limited. Consequently, testing the factors that strain this capacity can inform the cognitive components necessary for event perception.

Experiment 1

Experiment 1 measured event perception while monitoring two events in parallel. If event perception is capacity-limited then monitoring two events in parallel should lead to reduced encoding as capacity is filled.

Method

Participants

Two hundred participants (106 male, median age of 31) were recruited from Amazon's Mechanical Turk website for a $.35 compensation. Researchers have validated Mechanical Turk for use in cognition research (e.g., Paolacci, Chanlder & Ipeirotis, 2010; Germine et al, 2012). Participants were randomly assigned to the parallel-event or single-event condition.

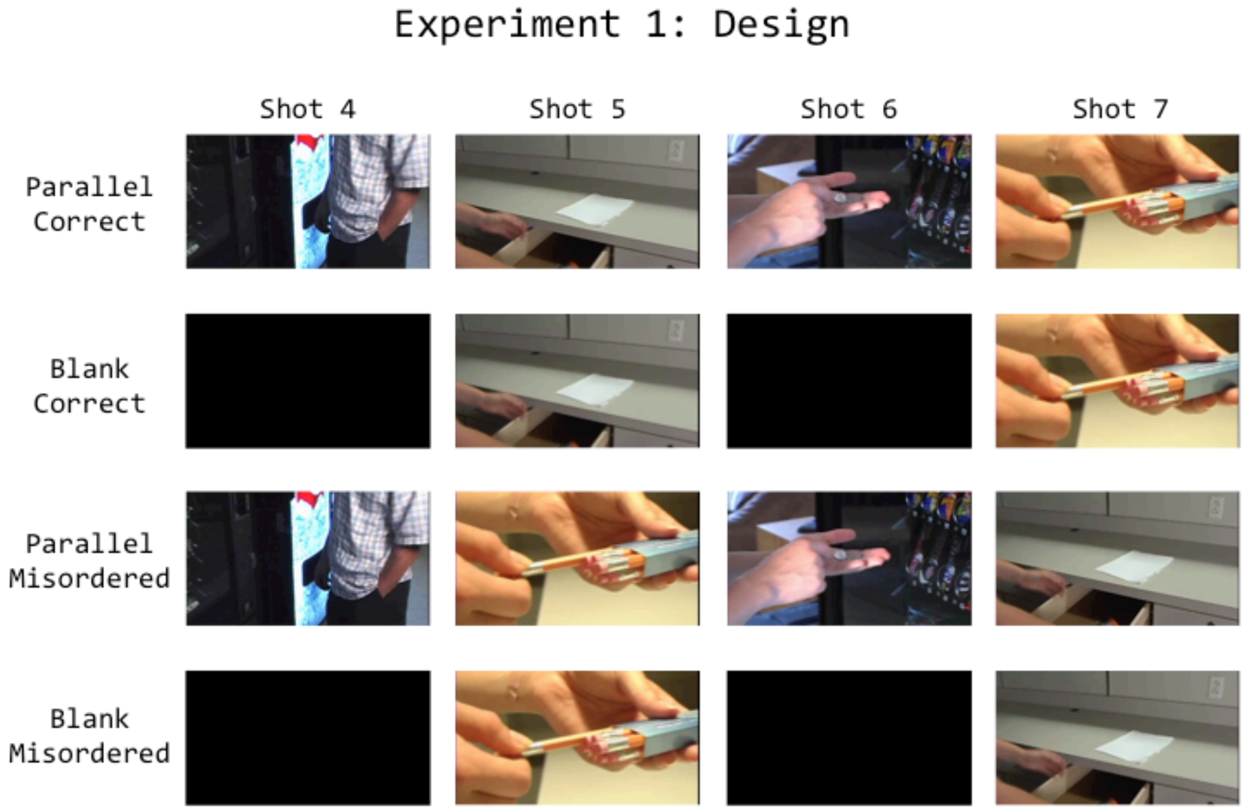


Figure 1.1 Representative stimuli from Experiment 1. Parallel movies alternated between two events (shown: using a vending machine and sharpening a pencil) or between one event and a temporally matched blank screen. Misordered sequences swapped the order of sequential actions within one event (shown: selecting a pencil in Shot 5 before grabbing a pencil pack in Shot 7)

Stimuli

Twelve video stimuli taken from Hymel, Levin and Baker (2015) were modified into two conditions (*Figure 1.1*). In the *parallel* condition, Adobe Premiere 6 was used to edit 12 events into 6 parallel sequences. Parallel sequences alternated events by each shot, in an ABAB fashion. Each paired sequence was edited into correct and misordered versions. Correct sequences presented each event in correct temporal order. Misordered sequences reversed the temporal order of one of the events (e.g., A1, B1, A2, B2, **A4**, B3, **A3**, B4…). Only one event was misordered at a time. Each edit was separated by 2 black frames (67 ms), to prevent potential illusory motion due to overlap of actions across shots (as did Hymel, Levin & Baker, 2015;

notably, performance was unaffected by removing these frames). Sequences in the single-event, *blank* condition were identical to the parallel sequences, except that observers saw a temporally matched black screen instead of a second event (e.g., A1, __, A2, __, A3…). Sequences were edited without sound. Correct and misordered sequences were made for each of the twelve events.

Procedure

Upon giving consent, participants were randomly assigned to either the parallel or the single-event conditions. Participants were told they would watch "videos of people performing everyday actions. For instance, a 'drinking soda' video might show a woman opening a refrigerator, grabbing a can, closing the fridge door, opening the can and sipping the drink." They were then told to "report whether you think an action occurs out of order. So if the woman in our soda movie drank from the can before opening it, you would say the action was out of order." Participants in the parallel condition were additionally told that they would watch two sequences at a time, and to respond if an action in either sequence occurred out of order.

Participants watched 6 sequences: 3 misordered and 3 correctly ordered. Participants were given a two-alternative forced choice response to the question, "Were any actions in that sequence out of order?" All combinations of condition, video and ordering were counterbalanced across the entire experiment.

	Parallel	Blank
N	98	102
Accuracy	52.0% (±3.2%)	73.5% (±3.9%)
Hit Rate	20.4% (±5.2%)	54.6% (±6.6%)
Correct Reject Rate	83.7% (±5.4%)	92.5% (±3.1%)
Discrimination (*d'*)	0.15	1.55

Table 1.1. Summary results from Experiment 1. Items in parentheses denote 95% confidence intervals. Signal discrimination was calculated from the entire sample.

Results

Summary results are shown in Table 1.1. Participants were significantly less accurate in the parallel condition than in the blank condition ($t_{198} = 8.359$, $p < .001$, $d = 1.19$). Signal detection analysis revealed greater discrimination in the blank condition (d' = 1.55) than in the parallel condition (d' = .15). Accordingly, participants in the blank condition were significantly better at correctly detecting misorderings ("hits", $t_{198} = 7.987$, $p < .001$, $d = 1.14$) and identifying correctly ordered actions ("correct rejects", $t_{198} = 2.798$, $p = .005$, $d = 0.40$).

Discussion

Participants in Experiment 1 were significantly worse at detecting misordered actions when viewing events in parallel. Importantly, participants in the blank condition still only detected roughly half of the misordered actions in the videos while generating relatively few false alarms, thus replicating performance in Hymel, Levin & Baker (2015). Furthermore, errors in detecting target action pairs cannot be attributed to an inability to see the target, as Hymel, Baker & Levin (2015) found that participants can accurately detect any action in these sequences if cued before viewing. These results support the claim that event perception is a capacity limited process, and that tracking two events in parallel reduces the detail of encoding during event perception.

The results of Experiment 1 may stem from either a lack of perceptual continuity between actions, attentional load as participants shift between sequences, or working memory load as participants attempt to represent ongoing activity from both sequences. According to the perceptual continuity hypothesis outlined in the introduction, individuals may use relatively simple perceptual transients to assess the continuity of objects, actions and spatial relationships between shots. This continuity is broken in the blank condition when the flow of action or relative space is interrupted in an otherwise seamless presentation between views, leading to detection of misordered actions. Meanwhile, alternation between sequences in the parallel condition eliminated direct comparison between shots of the same sequence, thus reducing the ability to detect misordered actions. According to the perceptual continuity hypothesis, the parallel condition did not reduce awareness because of attentional shifting or working memory load; rather, inserting perceptual information from another event disrupted comparison of the target misordered pair of shots. Experiment 2 therefore tested whether the deleterious effects of parallel events are due to disruption of perceptual continuity.

Experiment 2

Experiment 2 controlled for perceptual continuity when viewing single or parallel events by modifying the videos from Experiment 1 to always present critical misordered actions in pairs. If failures of event perception stem from the discontinuity of actions, objects or spaces between actions in a sequence, then participants in the blank condition should miss just as many errors as participants in the parallel condition. However, if failures of event perception stem from divided attention or working memory load, then participants should be significantly worse in the parallel condition.

Method

Participants

Participants (N = 100, 53 male, median age of 28) volunteered via Amazon's Mechanical Turk website for a $.35 compensation. Power analyses based on Experiment 1 (*d* = 1.19, post-hoc Power >.999) indicated that Power of .95 could be achieved with 50 participants. Given the minimal costs of Mechanical Turk participation, responses from 100 participants were collected. Participants were randomly assigned to either the parallel (n = 53) or single-event blank condition (n=47).

Stimuli & Procedure

The six videos pairs from Experiment 1 were modified so that actions were always presented in pairs from the same event. Videos in the parallel condition alternated events in an A1-A2-B1-B2 fashion. Reversals resulting in misordered actions were always presented within pairs, so that misordered actions were never separated by intervening events (e.g., …A3-A4-**B4-B3**-A5-A6…). Videos in the single-event, blank condition were identical to parallel stimuli, with the exception that actions from the second event were replaced by temporally matched black screens (e.g. A1-A2-__-__-A3-A4…). Edits were once again separated by 2 black frames (67 ms). Correct and misordered sequences were made for each of the twelve events. The stimuli and procedure were otherwise identical to Experiment 1.

Results

Results closely resembled those from Experiment 1 (*Table 1.2*). Participants were significantly less accurate in the parallel condition than in the blank condition ($t_{99} = 4.213$, $p < .001$, $d = .85$). Signal detection analysis also revealed greater discrimination in the blank condition ($d' = 1.29$) than in the parallel condition ($d' = .39$). Participants in the blank condition were again significantly better at correctly detecting misorderings (*hits*; $t_{99} = 4.893$, $p < .001$, $d = .99$), but were not significantly different in identifying correctly ordered actions (*correct rejects*; $t_{99} = 0.286$, $p = .775$).

	Parallel	Blank
N	53	47
Accuracy	55.0% (±14.8%)	71.3% (± 6.3%)
Hit Rate	23.3% (± 7.2%)	54.6% (±10.6%)
Correct Reject Rate	86.8% (± 5.7%)	87.9% (± 5.4%)
d'	0.39	1.29

Table 1.2. Summary results from Experiment 2. Items in parentheses denote 95% confidence intervals.

Discussion

Experiment 2 replicated the finding that viewing multiple events reduces awareness of misordered actions. These results rule out a perceptual continuity explanation, as misordered actions were preceded by the exact same action in both conditions. These results are consistent with a working model hypothesis, which predicts that viewing multiple events inducing working memory load, reducing awareness of actions within an event. However, these results are also consistent with the attentional shift hypothesis that switching attention between two events reduces the depth of encoding during event perception

The attentional shift hypothesis suggests that splitting attention between sequences reduces awareness of each sequence, similar to an attentional blink, where capture of attention by one stimulus promotes inattention to an immediately following stimulus (Raymond, Shapiro & Arnell, 1992), or to conceptual masking, where attention to a rapidly presented stimulus (~50-500 ms) interferes with the encoding of a previously seen stimulus (Potter, 1976; Loftus & Ginn, 1985; Breitmeyer & Öğmen, 2006). Experiments 3 and 4 therefore tested whether the deleterious effects of parallel events were due to working memory load or to lapses caused by switching attention between sequences of information.

Experiment 3

Experiment 3 replicated the design of Experiment 1 while directing attention to the target-containing sequnce. Participants viewed parallel or blank sequences as in Experiment 1. Before viewing each stimulus, participants read a prompt that cued them to attend to one sequence while ignoring the other. If parallel events induce failures of event perception because they split attention between target and distractor events, then cuing participants to the target event should reduce or eliminate the effect. However, if viewing two events taxes the working memory resources necessary to comprehend events, then awareness should decline in the parallel condition even when attention is cued to the target event.

The design of Experiment 3 additionally controlled for the asymmetry of participant responses in the designs of Experiments 1 and 2. Although the relatively high proportion of correct rejection trials in earlier experiments suggests that responses were not inflated by increased opportunity, it is still true that participants in the blank condition always knew which sequence would potentially contain a target misordered pair (i.e., the only sequence visible) whereas participants in the parallel condition had no cue indicating which sequence might possibly contain the target. Participants in both conditions of Experiment 3 always knew which sequence potentially contained the target misordered pair, thus eliminating this confound.

Method

Participants

Participants (N = 200, 101 female, median age of 34) volunteered via Amazon's Mechanical Turk website for a \$.35 compensation. Several additional filters were implemented to exclude participants who experienced problems with the videos or had previously participated in similar experiments (either this study or Hymel, Levin & Baker, 2015). Participants were excluded if they 1) reported that the video paused or stopped during the experiment (n = 3), if they paused or rewatched any videos during the experiment (n = 10) or if they had previously participated in a similar experiment (n = 3). After exclusion, this left 95 participants in the parallel condition and 91 in the blank condition.

Stimuli and Procedure

Experiment 3 used the exact same stimuli and procedure as Experiment 1, with the exception that each video was preceded by a short sentence cuing participants to attend to one of the sequences and to ignore the other. In the instructions, participants were told “for each video, you will be given the title of the sequence that might contain an error.” For example, one stimuli read, “Pay attention to the ‘starting a car’ sequence while ignoring the other sequence.” Participants were only cued to the sequence containing the misordered action in target-present trials. Participants were cued to a random sequence in target-absent trials. Combinations of condition, stimuli and cue were counterbalanced across all participants.

	Parallel	**Blank**
N	95	91
Accuracy	64.7% (+/-4.1%)	73.0% (+/-4.2%)
Hit Rate	38.6% (+/-7.0%)	55.3% (+/-7.2%)
Correct Reject Rate	88.8% (+/-4.6%)	92.3% (+/-3.4%)
d'	0.92	1.56

Table 1.3. Summary results from Experiment 3. Participants were cued to the content of the target sequence and asked to ignore the distractor. 95% confidence intervals are in parentheses.

Results

Results resembled Experiments 1 and 2 (*Table 1.3*). Participants were significantly less accurate in the parallel condition than in the blank condition (t_{1184} = 3.356, p < .001, d = .45). Signal detection analysis also revealed greater discrimination in the blank condition (d' = 1.560) than in the parallel condition (d' = 0.925). Participants in the blank condition were significantly better at correctly detecting misordered actions (*hits*; t_{184} = 3.354, p = .001, d = .48), but were no different in identifying correctly ordered actions (*correct rejects*; t_{184} = 1.213, p = .227).

Further analyses tested whether cuing participants to target events in Experiment 3 increased detection above the no-cue baseline of Epxeriment 1. Participants in the parallel condition of Experiment 3 detected significantly more misordered actions than participants in the parallel condition of Experiment 1 (*hits*; t_{191} = 4.035, p < .001, d = .60). However, participants in the parallel condition of Experiment 3 still detected significantly more misordered actions than participants in the blank condition of Experiment 1 (*hits*; $t_{195} = 3.318$, $p = .001$, $d = .47$). There were no differences between the blank conditions of Experiments 1 and 3 ($t_{195} = .148$, $p = .882$).

Discussion

Participants in Experiment 3 detected fewer misordered actions when viewing two events in parallel, even when they were explicitly told which sequence potentially contained a target misordered action. This supports the proposal that events are encoded automatically in working memory, and that increasing the number of events increases working memory load, resulting in reduced awareness. However, it is interesting that participants in the parallel condition of Experiment 3 performed better than those in the parallel condition of Experiment 1. It is possible that failures of event perception in the parallel condition of Experiment 1 were due in part to dividing attention between two sequences. However, it is possible that the cue in Experiment 3 was insufficient in directing attention to one sequence over the other, and that a more salient cue might eliminate differences between blank and parallel conditions altogether. Experiment 4 tested whether a salient perceptual cue further mediated event perception in parallel sequences.

Experiment 4

The results of Experiment 3 in part support an attentional shift hypothesis, which predicts that event perception requires sustained attention, and that dividing attention between parallel events leads to failures of event perception. However, it is possible that the conceptual cue to the event containing a misordered action was too underspecified to adequately draw attention to the target sequence. Participants may have had attend to each shot long enough to determine if it was part of the target event, generating sufficient attentional load to reduce perception of misordered pairs. To account for this, Experiment 4 used a salient perceptual cue to indicate the sequence potentially containing a target event. If parallel events constrain event perception because they divide attention between target and distractor events, then perceptually cuing participants to the target event should reduce or eliminate the effect. However, if viewing two events taxes the working memory resources necessary for awareness, then awareness should decline in the parallel condition even when attention is cued to the target event.

Method

Participants

Participants (N = 208, 108 male, median age of 32) volunteered via Amazon's Mechanical Turk website for a \$.35 compensation. Participants were excluded if they 1) reported that the video paused or stopped during the experiment (n = 4), if they paused or rewatched any videos during the experiment (n = 24) or if they had previously participated in a similar experiment (n = 12). After exclusion, this left 87 participants in the parallel condition and 83 in the blank condition.

Stimuli and Procedure

Experiment 4 followed the exact procedure of Experiment 3 with one exception. Instead of cuing participants to the target-containing sequence with a text prompt before the video, target-containing sequences in the parallel and blank conditions were bounded by a 100-pixel red stripe on the top and bottom of the screen. Target-absent sequences (with correctly ordered actions) in the parallel condition and temporally matched black screens in the blank condition were never bounded by red stripes. A random sequence was bounded in target-absent trials. Stimuli were counterbalanced for bounded target-present events, bounded target-absent events and condition across all participants.

	Parallel	**Blank**
N	87	83
Accuracy	60.5% (+/-4.4%)	71.7% (+/-4.1%)
Hit Rate	34.1% (+/-7.5%)	50.6% (+/-5.7%)
Correct Reject Rate	87.0% (+/-7.4%)	92.8% (+/-3.4%)
d'	0.71	1.47

Table 1.4. Summary results from Experiment 4. Target sequences were indicated by red stripes bounding the frame. 95% confidence intervals are in parentheses.

Results

Results were nearly identical to Experiment 3 (*Table 1.4*). Participants were significantly less accurate in the parallel condition than in the blank condition ($t_{168} = 3.613$, $p < .001$, $d = .56$). Signal detection analysis also revealed greater discrimination in the blank condition ($d' = 1.474$)

than in the parallel condition ($d' = .715$). Participants in the parallel condition were significantly worse at correctly detecting misordered actions (*hits*; $t_{168} = 3.045$, $p = .002$, $d = .47$), and were marginally worse at identifying correctly ordered actions (*correct rejects*; $t_{168} = 1.721$, $p = .087$, $d = .27$).

Differences between experiments were then analyzed. A 3 (Cue: No Cue [Exp. 1], Conceptual Cue [Exp. 2], or Perceptual Cue [Exp. 3]) by 2 (Condition: Parallel or Blank) between-subjects ANOVA revealed a significant effect of Cue ($F_{2,555} = 3.656$, $p = .027$, $\eta_p^2 = .013$) and Condition ($F_{2,555} = 66.127$, $p < .001$, $\eta_p^2 = .106$), with a significant interaction effect ($F_{2,555} = 4.444$, $p < .012$, $\eta_p^2 = .016$). Paired comparisons controlling for false discovery rate using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) revealed that participants detected significantly fewer misordered actions in the parallel conditions of all three cue types ($t_{559} = 8.006$, $p <.001$, $d = .67$). Paired comparisons between the parallel conditions of each cue type revealed that participants detected significantly fewer misordered actions with No Cue than with a Conceptual Cue ($t_{191} = 4.035$, $p <.001$, $d = .58$) or Perceptual Cue ($t_{185} = 4.038$, $p <.001$, $d = .59$), but there was no difference between Conceptual and Perceptual Cues ($t_{182} = .874$, $p = .383$). Detection in Blank conditions did not differ by cue type ($t < .34$, $p > .38$).

Discussion

Participants in Experiment 3 and 4 detected significantly fewer misordered actions when viewing two events in parallel, even when explicitly told to attend to one event and ignore the other. The results of Experiment 3 and 4 were virtually identical. However, it is interesting to note that both experiments yielded significantly greater detection of misordered actions in parallel trials than in Experiment 1. Cuing participants to the sequence containing the target misordered action appears to mediate some of the deficits incurred by viewing parallel events. Despite this, participants were less aware of sequence errors in parallel conditions, supporting the hypothesis that event perception automatically recruits working memory to comprehend and track events, and that viewing multiple events induces working memory load. These results suggest that event perception requires both working memory and focused attention.

Two modifications to Experiment 4 may shed more light on the role of focused attention in event perception. The perceptual cues in Experiment 4 appeared at the onset of each shot in the cued sequence, and may have insufficiently prepared participants to attend. Furthermore, the speed of stimulus presentation may have increased the difficulty of the task. Participant performance might improve if cue onset occurred 10 frames (~333 ms) before shot onset, giving participants adequate time to attend to or ignore the upcoming action. If event perception automatically recruits working memory, then parallel events should impair event perception event in this modified cuing procedure. Though intriguing, this experiment will be held to future investigation.

Experiment 5

Experiments 1-4 demonstrated that tracking multiple events in parallel impedes event perception. It is hypothesized that viewing parallel events increases working memory load, reducing the capacity to encode features in either event. Previous research has demonstrated that working memory load from a secondary task decreases awareness of visual properties (de Fockert, Rees, Frith & Lavie, 2014). Participants in this prior experiment detected fewer oddballs from a rapid serial visual presentation task when they encoded a difficult sequence of numbers in working memory (e.g., "0,3,1,2,4") rather than an easy sequence of numbers (e.g., "0,1,2,3,4"). Following this, Experiment 5 directly tested the effect of external working memory load on event perception. Participants watched each stimulus video while holding some number of objects in working memory. If event perception requires working memory, then individuals under high load should detect fewer misordered actions.

Method

Participants

Participants (N = 136, 77 male, median age of 31.5) volunteered via Mechanical Turk for a $.35 compensation. Participants were excluded if they 1) reported that video playback stopped during the experiment (n = 5), if they paused or re-watched any videos during the experiment (n = 6) or if they had previously participated in a similar experiment (n = 4). Four participants were excluded from analysis for missing all questions on the working memory task. After exclusion, this left 56 participants in the high load condition and 62 in the low load condition.

Stimuli and Procedure

Experiment 5 used the 12 edited sequences used in Hymel, Levin & Baker, 2015. Unlike Experiments 1-4, these videos were not edited into parallel or blank versions. Otherwise stimuli were identical to those used in the previous studies. Participants were instructed to identify misordered actions in the stimuli. Participants were additionally informed that they would memorize a display of some number of colored squares before each movie (*Figure 1.2*). Once the movie finished, they would identify which of the two squares swapped positions. Participants then completed two practice working memory trials with feedback before beginning the experiment.

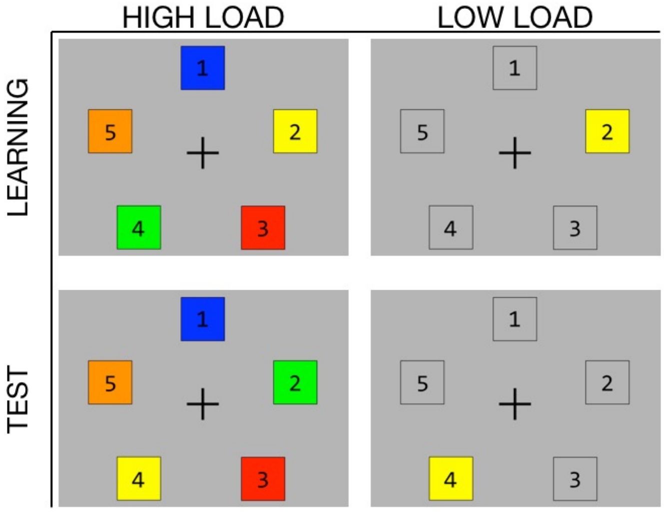


Figure 1.2. Representative high and low working memory load trials. Participants memorized the colors of squares in five locations prior to watching each video (Learning phase). Participants selected which two locations changed colors in the Test phase (squares 2 and 4 above).

Experimental trials began with a WM learning phase, followed by stimulus video presentation and forced-choice target misordering detection, and then ending with a WM test phase. Participants in the High WM condition saw five uniquely colored 100x100 pixel squares evenly displayed about a central 75x75 pixel fixation cross. Participants in the Low WM condition saw a display with the exact same properties, with the critical difference that only one square was colored. Participants memorized the learning display for five seconds before viewing a test video depicting a single event across several cuts, with two black frames (67 ms) separating each shot. Immediately following the video, participants responded whether they detected a misordered action in the clip. They then saw the WM test image and selected which two boxes swapped position using a drop-down menu. Participants were only scored as correct if they identified both of the squares that swapped colors. Participants repeated this procedure through all 12 videos, which were randomly presented and counterbalanced for target presence across participants.

	High Load	Low Load
N*	56	62
WM Accuracy	82.3% (± 6.7%)	93.1% (± 2.7%)
Accuracy	63.2% (± 4.2%)	65.7% (± 2.8%)
Hit Rate	37.4% (± 7.5%)	37.4% (± 5.7%)
Correct Reject Rate	92.8% (± 4.4%)	94.1% (± 2.6%)
d'	1.22	1.32

Table 1.5. Summary results from Experiment 4. Participants detected misordered actions from single event sequences under high or low working memory load. 95% confidence intervals in parentheses. Target detection analysis was limited to working memory accurate trials.

Results

Summary statistics are shown in Table 1.5. Accuracy in the working memory task validated the difficulty of the measure. Participants correctly identified significantly more color-locations changes in the low load condition than in the high load condition ($t_{116} = 3.663$, $p < .001$, $d = .67$). Detection of misordered actions was calculated for trials with correct working memory responses. There were no significant differences in overall accuracy in the misordered action detection task ($t_{116} = 1.020$, $p = .310$), nor were there significant differences in target detection ($t_{116} = 0.006$, $p = .995$) or correct rejection ($t_{116} = 0.479$, $p = .632$).

Results from Experiment 5 were then compared to Experiment 1. A one-way ANOVA coding condition as a between-subjects factor revealed a significant difference in hit rate ($F_{3,313} = 25.602$, $p < .001$, $\eta^2 = .186$). Paired comparisons corrected using the Benjamini-Hochberg procedure revealed that participants in the parallel viewing condition were significantly worse than participants in both the high load ($t_{152} = -3.257$, $p <.001$, $d = .53$) and low load ($t_{159} = -3.906$, $p <.001$, $d = .62$) conditions. Furthermore, participants in the blank viewing condition performed significantly better than participants in the high load ($t_{156} = 3.899$, $p <.001$, $d = .601$) and low load ($t_{163} = 3.252$, $p <.001$, $d = .501$) conditions.

Discussion

Experiment 5 tested whether increasing magnitudes of working memory load would impair event perception. Results suggest that there were no differences in detection of misordered actions whether participants sustained high or low working memory load. However, comparison with the results of Experiment 1 suggests that both high and low load conditions reduced awareness below a no-load baseline. Conversely, awareness of misordered actions was far lower when viewing parallel events in Experiment 1 than during either working memory load condition of Experiment 5. These results suggest that taxing working memory decreases awareness of events, but that switching between events exacerbates this effect. These results support both the working model and attentional shift hypotheses.

A further consideration should be the degree to which this design manipulated working memory rather than short term memory. As discussed above, the present task was adapted from a previous working memory task used to induce visual inattention (de Fockert et al, 2001). However, participants in Experiment 5 updated of object locations rather than item positions in a list, to better compare task load in this experiment with the visual load presumably generated by parallel visual events. Rather than generating working memory load, the visual nature of the task may have been within the span of visual short term memory (Todd & Maroi, 2004), which has been demonstrated to have a capacity of around 4-5 objects (Alvarez & Cavanagh, 2004). A task requiring participants to maintain and update information in working memory across trials might lead to differences between conditions. Such a manipulation would be more akin to the verbal overshadowing task used by Hymel, Levin & Baker (2015) to reduce detection of sequence errors.

General Discussion

I have demonstrated in five experiments that event perception requires working memory and focused attention. Participants in Experiments 1 and 2 were significantly worse at detecting impossibly misordered actions when viewing two events in parallel. Experiments 3 and 4 replicated this effect, even when participants were cued to the specific sequence containing the target misordered action. Interestingly, cuing attention to a specific sequence moderated some of the deficits incurred from viewing parallel events. Experiment 5 revealed that any degree of working memory load held throughout viewing decreased awareness below performance in Experiment 1, although there were no differences between high and low load.

These results, particularly from Experiments 2 and 5, rule out the plausibility of the perceptual continuity hypothesis of event awareness, where moment-to-moment awareness is determined by detection of discontinuous objects, actions and spaces within a stream of activity. Such a hypothesis would predict that participants would perceive events as continuous so long as objects, actions and their spatial relationships are held intact between views. Detection of misordered events would stem from between-view comparisons of object and spatial motion. As detection of misordered actions was unaffected by showing target actions in unbroken pairs within a parallel sequence in Experiment 1, it is unlikely that event perception is sustained by moment-to-moment evaluation of perceptual continuity.

The results of Experiments 3, 4 and 5 add credibility to both the attentional shift and working model hypotheses. According to the former, event awareness requires sustained attention to a single sequence of events; when attention is divided, the viewer is less likely to notice impossible events. Meanwhile, the working model hypothesis predicts that event comprehension is rooted in the encoding and comparison of relevant properties in a working memory model of the current situation, and that increasing the number of events held in working memory should reduce awareness of individual actions. Both of these hypotheses proved to have merit. Individuals appear to encode information relevant to comprehension of ongoing events in working memory. Monitoring multiple events decreases the amount of information available for each in event as working memory reaches capacity, while dividing attention may reduce the quality of information encoded in the first place.

Future Directions

Experiment 5 demonstrated that *any* form of working memory load reduced awareness below the single event control of Experiment 1. Another experiment should be run to directly compare awareness between a single set of observers using identical stimuli, as the blank condition of Experiment 1 had substantially different timing between shots than the stimuli used in Experiment 5. As these results stand, it is possible that simply spacing the presentation of information between shots increases awareness, although it is notable that participants in the blank condition of Experiment 1 demonstrated similar rates of detection as found in Hymel, Levin & Baker (2015), whose rates of presentation mirrored those of Experiment 5. In addition to this replication, several future experiments are required to fully test the role of working memory on awareness of parallel events.

Experiment 5 differed from earlier experiments in that the detection task was entirely enveloped by the load task, as opposed to alternating with a second event. The effect of diminished awareness under this load task should be directly compared with a similar overlap in the display of events. Viewers may maintain an unfinished event model, much like the classic Zeigarnik effect (Zeigarnik, 1938), wherein unfinished tasks linger in working memory. In this hypothetical experiment, participants would watch half of event A followed by all of event B and concluded by the remainder of event A. If participants detect fewer misordered actions in Event B if it is bounded by a distractor event than if event B is bounded by blank frames, then it is possible that the incomplete event is taxing working memory.

Experiments 3 and 4 further suggest that participants may automatically attend to and encode events as they are presented, even when explicitly instructed to ignore the distractor event. One outstanding question is the degree to which participants comprehend the ignored event. In a follow-up of the attentional cuing experiments, participants could be cued to one event and instructed to ignore the parallel event. Participants would be tested on their detection of misordered actions in the cued event for the first five trials, and on the last trial be tested for their detection of a misordered action in the distractor event. To prevent potential demand characteristics from this single-trial experiment (where participants may feel tempted to report a misordered event so as not to appear “fooled”), this experiment will require participants to respond in a few words exactly what occurred out of order. In addition, participants would briefly summarize *both* events to test whether any comprehension of the ignored event penetrated awareness. If some level of event perception occurs automatically, then at least some participants should be aware of the parallel ignored event.

One final way to test whether that ignored distractor sequences induce working memory load is to compare parallel events with a condition that alternates a single event with the same shot of an unrelated action (e.g., a reach or throw) repeated between each shot of the target event (e.g., A1, B1, A2, B1, A3, B1…). If it is truly *comprehension* of the distractor event that induces load and reduces awareness, then repeating a meaningless action between shots should not reduce detection of misordered actions. However, if failures of event perception stem from a shift of attention to salient motion, then meaningless repeated parallel actions should interrupt detection of misordered actions just as much as viewing an entirely different event.

Summary & Conclusions

In five experiments, I have demonstrated that event perception requires both focused attention and limited working memory resources using a novel measure of event comprehension. This study is the first to directly the cognitive limitations of event perception. These results leave open a great deal of future research inquiries which test questions of automaticity, domain-generality and capacity in our perception of the real world.

CHAPTER 2

THE LIMITS OF DEFAULT PERSPECTIVE TAKING: MULTIPLE AGENTS AND ATTENTIONAL SETS

Abstract

Several researchers have now proposed that humans have a rapid, default and early-developing system for encoding another's perspective. Proponents of this theoretical system claim that the processing costs incurred when another agent holds a different perspective are due to an automatic calculation of that agent's perspective. However, detractors note that such behaviors might stem from domain-general cognitive abilities. It is therefore imperative to understand the exact circumstances under which other's perspectives affect cognition. In five experiments I investigate perspective taking in scenes containing multiple agents and attentional sets. If processing costs associated with default perspective-taking stem from representation of each unique perspective in a scene, then participants should display increasing processing costs as the number of conflicting perspectives increases. However, Experiments 1a, 1b and 2 found equal processing costs when any number of agents held a different perspective. Experiments 3a and 3b then demonstrated that another's perspective only influences behavior after participant selection of an attentional set. These experiments demonstrate that default perspective-taking is not a calculation of every available mental state. Rather, this system may be a heuristic signal that one has privileged access to an attended set of objects.

Introduction

When we debate politics, play sports or plan for retirement, we are putting aside our current mental state and adopting the perspective of someone else, be it a colleague, a competitor or a future version of ourselves. Although researchers have historically demonstrated perspective taking to be cognitively taxing and error prone (Lin, Epley & Keysar, 2010), recent findings suggest that observers may use limited forms of another's visual perspective by default to efficiently guide action in social situations (Kovacs, Teglas & Endres, 2010). Several theories posit that this early system functions in parallel with the effortful processes associated with more complex perspective taking (Leslie, Friedman & German, 2005; Apperly & Butterfill, 2009), whereas others claim that existing experiments hinge on domain-general cognitive abilities (Heyes, 2014; Santiesteban, Catmur, Hopkins, Bird & Heyes, 2014; Catmur, Santiesteban, Conway, Heyes & Bird, 2016). The debate on the source of these behaviors can only be settled once we understand the exact circumstances in which another's perspective influences cognition. Here, I test whether the processing costs associated with default perspective-taking incorporate the perspectives of multiple agents and multiple perspectives of an individual agent. In doing so, I demonstrate that default perspective taking arises when any agent has a different line of sight access to an attended set of objects. These behaviors suggest a decision making heuristic that one's own visuospatial access is privileged. This heuristic likely arises *after* attentional selection.

A robust literature has documented the apparent limitations of mental state reasoning in children and adults. Seminal observations in cognitive development revealed that children younger than four or five demonstrate a remarkable inability to imagine a scene from a different perspective (Piaget & Inhelder, 1958) or suppress their own knowledge when thinking about other's mental states (Wimmer & Perner, 1983). Although children consistently improve on perspective-taking tasks starting around age four, research suggests that even adults suffer from persistent *egocentrism*, default inability to consider another's mental state as different from one's own (Epley, Morewedge & Keysar, 2004). Research suggests that theory of mind ability is highly correlated with working memory capacity, finding that participants show more egocentric errors when making rapid responses (Keysar, Lin & Barr, 2003), under cognitive load (Lin, Keysar & Epley, 2010), or even when simply unmotivated (Klein & Hodges, 2001). Many researchers thus concluded that theory of mind was late-developing, cognitively taxing and egocentric.

Despite the frequent errors seen in adults, infants as young as five months have demonstrated tacit understanding of goals, intentions and beliefs. Infants respond to visual changes in perceived goals (Onishi & Baillargeon, 2005), can detect animate movement in nonliving things (Luo & Baillargeon, 2005), and differentiate intentional from nonintentional actions (Tomasello et al., 2005). Perhaps most striking, however, is the success of many infants in false belief tasks often considered the gold standard of mental state reasoning. Kovacs, Teglass and Endress (2010) had five-month-old infants watch a short film of a cartoon agent entering a room with a ball. The cartoon agent then watched the ball stay in place or move behind an occluder. After the agent left, the ball would then move behind a second occluder or stay in the same location. The agent then returned and the occluders were removed. Infant looking times and adult response times both increased when the ball appeared in a location that was inconsistent with either the participant's or *the agent's* beliefs. A second paradigm geared for adults further suggested that individuals may generate basic representations of other's perspectives by default. In Samson et al's *dot-perspective task* (2011), participants were instructed to take their own perspective or the perspective of a cartoon agent. Participants then responded whether that perspective could see a given number of objects in the range of subitization (i.e., 0-3; Trick & Pylyshyn, 1994). Participants across numerous replications have demonstrated slower and more error prone responses when the agent's perspective is inconsistent with the participant's perspective (McCleery et al, 2011; Qureshi, Apperly & Samson, 2010; Samson, Apperly, Braithwaite, Andrews & Scott, 2010; Samson & Apperly, 2010; Surtees & Apperly, 2012; Mattan, Rosthein & Quinn, 2016; Surtees, Samson & Apperly, 2016; Surtees, Apperly & Samson, 2016; Baker, Levin & Saylor, 2016), even when participants were only asked to take their own perspective (Experiment 2 of Samson et al, 2011).

Two Systems for Perspective-Taking?

To reconcile the apparent success of infants in perspective-taking tasks previously thought difficult even for mature adults, researchers have proposed two-system theories of mental state representation (e.g., Fodor, 1992; Leslie, Friedman & German, 2004; Onishi & Baillargeon, 2005; Surian, Caldi & Sperber, 2007; Apperly & Butterfill, 2009, among many others). According to these theories, an early-developing default system selectively guides awareness to the presence of another's differing perspective. After this early system identifies another perspective, a later developing, cognitively effortful system can represent the differences between self and other perspectives[1]. Apperly and Butterfill have analogized these dual processes to number cognition, where an early-developing domain-specific module processes number-like properties such as relative magnitude and density while a later-developing domain-general process effortfully organizes information using mathematics knowledge. A two-system theory of perspective taking potentially reconciles conflicts between adult and infant theory of

---

[1] To prevent confusion, I follow the precedent of Apperly & Butterfill and refer to these systems as the early and late systems. The increases in response time delays and errors related to the early system are referred to collectively as processing costs.

mind research and opens new alternatives for the longstanding debate between theories of domain specificity or generality in theory of mind.

Many experiments explicitly suggest that these behaviors constitute a default representation of another's perspective (e.g., McCleery et al, 2011; Qureshi, Apperly & Samson, 2010; Samson, Apperly, Braithwaite, Andrews & Scott, 2010; Samson & Apperly, 2010; Surtees & Apperly, 2012). The initial thought was that such a modular system would be specific to living agents (Samson et al, 2010). However, several experiments revealed processing costs when displays used arrows instead of avatars (Heyes, 2014; Santiesteban, Shah, White, Bird & Heyes, 2014; Santiesteban, Catmur, Hopkins, Bird & Heyes, 2014), leading to the suggestion that the effect stemmed from low-level spatial cuing, of the type seen with gaze-following experiments (Friesen, Kingstone & Ristic, 2004). However, processing costs persist when an agent's visual access to targets is manipulated using occluders rather than avatar location and orientation (Baker, Levin, & Saylor, 2016). This suggests that the early system monitors line-of-sight visual access, and that effects are not artifacts stemming from spatial cues of attention. Even taken with the most conservative interpretation, participants appear to rapidly adapt their behavior when an alternative perspective arises, even when ignoring the other agent would be beneficial.

The Current Study

Given that research has repeatedly demonstrated rapid and default detection of inconsistent perspectives, the goal of future research must be to define the mechanisms underlying this behavior. At this point, researchers posit that the early system is present by at least 5 months of age (Kovacs, Teglas & Endress, 2010) and is characterized by preferential representation of objects (e.g., Samson et al, 2010) within line-of sight visuospatial access of an agent (Baker, Levin & Saylor, 2016). In the current experiment, I test two critical assumptions of the default system of perspective taking: that processing costs stem from automatic representation of all other perspectives, and that this default representation directs attention to inconsistent perspectives. This study tested the limits of the early system to respond to multiple perspectives or multiple possible attentional sets from a single perspective. If participants incur increasing processing costs with the inclusion of additional agents, then participants might be representing multiple distinct perspectives. However, if participants demonstrate the same costs to any number of differing perspectives, then the early system may be more of a signal of privileged access to a scene rather than representation of an entirely different perspective from a different viewpoint. This signal would be more akin to a spatial grouping of a scene by an agent's line-of sight than to a robust representation of the agent's mental state.

Furthermore, if the default system operates as an "efficient" subsystem of theory of mind that functions "to guide young children's attention to cases in which [participants'] epistemic perspective diverges from that of someone else" (Apperly & Butterfill, 2009), then the system should function as an exogenous cue of attention that signals when any object appears in contrast to another's perspective. I therefore tested whether participants incur processing costs when an agent holds a different perspective of objects that the participant is trying to ignore. If the early system selectively guides attention to another's restricted perspective, then participants should incur costs when an agent has a different perspective of any set of objects, even those outside of the current attentional set.

Experiment 1

Human social interaction regularly involves events containing dozens or even hundreds of other agents. There are two ways in which the default system may process multiple perspectives. If the default system represents every available perspective (as suggested by Leslie, Friedman & German, 2004), then every additional perspective should incur a processing cost. On the other hand, the default system may be indiscriminately sensitive to *any* deviation from one's own perspective. According to this hypothesis, the early system operates as a heuristic cue that any other agent in the scene has limited visuospatial access. This system would fall short of full perspective-taking, as participants would not need complete representation of the objects, angles and occlusions available to all other agents. Rather, participants would only calculate obstructions within one's own perspective that might limit another's perspective. Thus the early system would be an indicator of privileged knowledge, a heuristic signal of divided space.

Experiments 1 and 2 used a previously validated modification of the dot-perspective task (Experiment 2 of Baker, Levin & Saylor, 2016) to test the effect of multiple conflicting perspectives on processing costs. Participants responded whether they or another agent could see a number of targets within a scene (Figure 2.1). If early-system perspective taking involves complete representation of a relevant agent's perspective, then we should see increasing processing costs with the addition of more diverging perspectives in a scene. However, if the early system operates categorically by signaling whether objects may be unseen by *any* agent, then processing costs in inconsistent perspective situations should be relatively unaffected by the inclusion of additional agents.

Whereas previous versions of this design had two classes of perspective conflict (Self and Other perspectives which were Consistent or Inconsistent with the participant), this design affords conditions of perspective-taking leading to three principled hypotheses (*Figure 2.2)*. According to an e*gocentric deviation* hypothesis, seeing any perspective different from one's own causes interference. In this case, interference is not due to representation of another's perspective per se, but sensitivity to any perspective that differs from egocentric knowledge. Thus, one differing perspective would cause the same processing cost as two differing perspectives. We would conclude that the default system is largely categorical, calculating whether one's own perspective is inaccessible from any other perspective.

It is also possible that the early system represents multiple perspectives individually, which would consequently increase processing costs as a function of the number of discrete perspectives in the scene. Two hypotheses may be generated from this possibility. According to an *individuation-convergence* hypothesis, the early system detects multiple perspectives and represents their intersection. Thus, multiple perspectives that differ from one's own to the same degree (as with *Figure 2.2b*, when you see three targets but both avatars see 1) might increase interference beyond a single differing perspective (*Figure 2.2c*). To put another way, the early system weighs each perspective equally; observers demonstrate greater processing costs when their own perspectives are outweighed by the combination of two conflicting perspectives. In a similar vein, an *individuation-divergence* hypothesis would predict that the early system individuates perspectives and attention is drawn to regions of conflicting perspectives. In this case, interference should increase with contrasting perspectives. In the case of Figure 2.2d, the participant's perspective differs from the avatars' perspectives in two ways, the male avatar seeing one and the female avatar seeing two. This might trigger the early system twice, effectively generating load from representing three diverging perspectives at once. Either the individuation-convergence or the individuation-divergence hypothesis would suggest a default system that can not only rapidly individuate multiple perspectives, but also represent their contents additively.

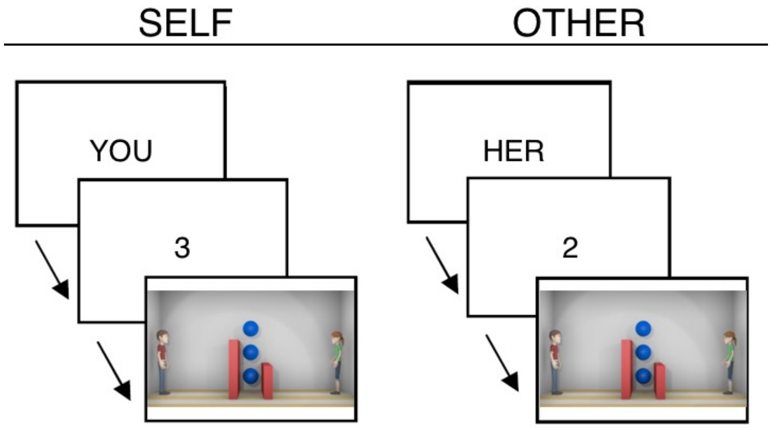


Figure 2.1 Design of all experiments. Participants responded whether the perspective cued could see the number of objects cued in the display ("match" trials shown).

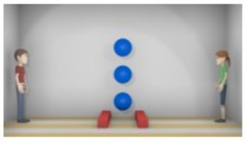
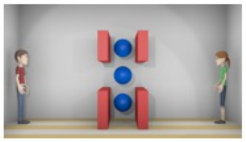
		Condition Name	Egocentric Deviation	Individuation Convergence	Individuation Divergence
a		Consistent	Baseline	Baseline	Baseline
b		Self-Other Inconsistent	XX	XXX	X
c		Partial Inconsistent	XX	XX	XX
d		All Inconsistent	XX	X	XXX

Figure 2.2. Schematic of conditions and hypotheses for Experiments 1 & 2. X's denote expected magnitude of increased response time and error rates above baseline in Self-perspective trials.

Method

Participants

Twenty-four Vanderbilt University undergraduates participated for class credit (mean age = 19.0, 18 Female, all right-handed). All participants throughout this study were treated in compliance with IRB approval and APA ethical standards.

Stimuli

Stimuli were created with the 3D editing software, Blender (v2.66; The Blender Foundation, www.blender.org, 2015), using cartoon avatars downloaded from an open source file-sharing site (VMComix, 2011). The stimuli were reproduced following the method described in Baker, Levin & Saylor (2015; Experiment 2), which replicated Samson et al (2010) using occlusion to vary avatars' visual access to targets rather than avatar direction. This experiment changed the design to incorporate a second cartoon agent. The agents, one male and one female, appeared on either side of the display. The respective location of the agents was fixed, with the male on the right and the female on the left, to prevent potential search costs or training effects.

Test images consisted of a 3-dimensional room with up to three blue target spheres flanked by two agents. Two walls appeared between the agents and the targets. Each wall was segmented into three subunits each, for a total of six possible occluders (*Figure 2.1*). In all cases, the participant could see all of the displayed dots, while the avatars could see varying numbers of dots depending on wall position.

Twenty stimulus images were created in four experimental categories, for a total of 80 test images. In *Consistent* images (*Figure 2.2a*), the participant and both avatars were able to see the same number of targets. Three sets of Inconsistent perspective types were created. In the *Self-Other Inconsistent* condition (*Figure 2.2b*), both avatars saw the same number of targets but were inconsistent with the participant's view (e.g., you see 3, male avatar sees 2, female sees 2). In the *Partial Inconsistent* condition (*Figure 2.2c*), one of the avatars saw the same number of targets as the participant while the other saw fewer targets (e.g., you see 2, male sees 1, female sees 2). The consistent perspective was counterbalanced across the 20 trials, so that each avatar was consistent with the participants' perspective on half of trials. In the *All Inconsistent* condition (*Figure 2.2d*), both avatars saw a different number of targets than the participant and from each other (e.g., you see 3, male sees 1, female sees 2). The number of targets seen by each avatar was counterbalanced, so that the female saw more targets than the male on half of trials.

Six unique cue sets were created for each image, again following the method of Samson et al. (2011). Participants first viewed a perspective cue that either primed the participant's perspective (*Self* trials: "YOU") or one of the avatars' perspectives (*Other* trials: "HIM" or "HER"). Participants next saw a number cue, which either aligned with the perspective cued (*Matched*) or were did not align with the perspective cued (*Mismatched*). Critically, Mismatched trials for Inconsistent images displayed the number of targets seen by an irrelevant perspective. Mismatched trials from one of the Other perspectives corresponded with the number of objects participant's perspective, while mismatched trials from the participant's perspective corresponded with the number of objects seen by one of the avatars. Take the example of an image where the male and female avatars see two targets but the participant sees three targets (as in Figure 2.2b). In a Self-Matching trial, the cues would be "YOU" and "3", but in Self-Mismatching trials, the cues would be "YOU" and "2".

It is worth noting that number cues for Mismatching trials in All-Consistent displays did not correspond with any perspective. In these situations, a random number generator was used to select a number between zero and three that did not correspond with any perspective. All participants viewed the same randomly generated number for mismatching trials on individual All-Consistent displays. Previous studies have viewed Mismatching-Consistent trials as fillers, as the number cues generate no conflict with any potential representation (Samson, Apperly, Braithwaite, Andrews & Scott, 2010; Qureshi, Apperly & Samson, 2010, McCleery et al, 2011; Surtees & Apperly, 2012). However, such exclusion criteria might bias results towards false acceptance of an underpowered effect. Here I follow the logic presented in Baker, Levin & Saylor (2016) and analyze results from all trials containing targets.

Procedure

Upon giving consent, participants completed the entire experiment using the psychophysics toolbox for MATLAB (Kleiner, Brainard & Pelli, 2007) on an Apple Mac Mini with a 19.5" LCD monitor (1600x900p, 60 Hz). Participants read instructions and then completed a tutorial before beginning experimental trials. The tutorial consisted of four practice trials, which followed the experimental procedure outlined below. Each practice trial demonstrated a different condition (consistent, self-other inconsistent, partial inconsistent and all inconsistent). Practice trials gave feedback and repeated until participants made the correct response. The experimenter verbally confirmed that participants understood the tutorial before the experiment began.

Each trial began with a fixation cross for 750 ms followed by a 500 ms blank screen. Participants then saw a perspective cue for 750 ms and a second 500 ms blank. They then saw a number cue for 750 ms immediately followed by the stimulus. Participants viewed each of the 80 displays six times, once for each combination of perspective cue (YOU vs. HIM vs. HER) and number cue (Matching vs. Mismatching), for a total 480 trials in all. For each display, participants responded whether the perspective cued matched the number cued. Trials were terminated if participants took longer than 2000ms to respond. Trials were completely randomized over six blocks. The entire experiment took a maximum of 40 minutes, but participants completed the experiment in 29.7 minutes on average.

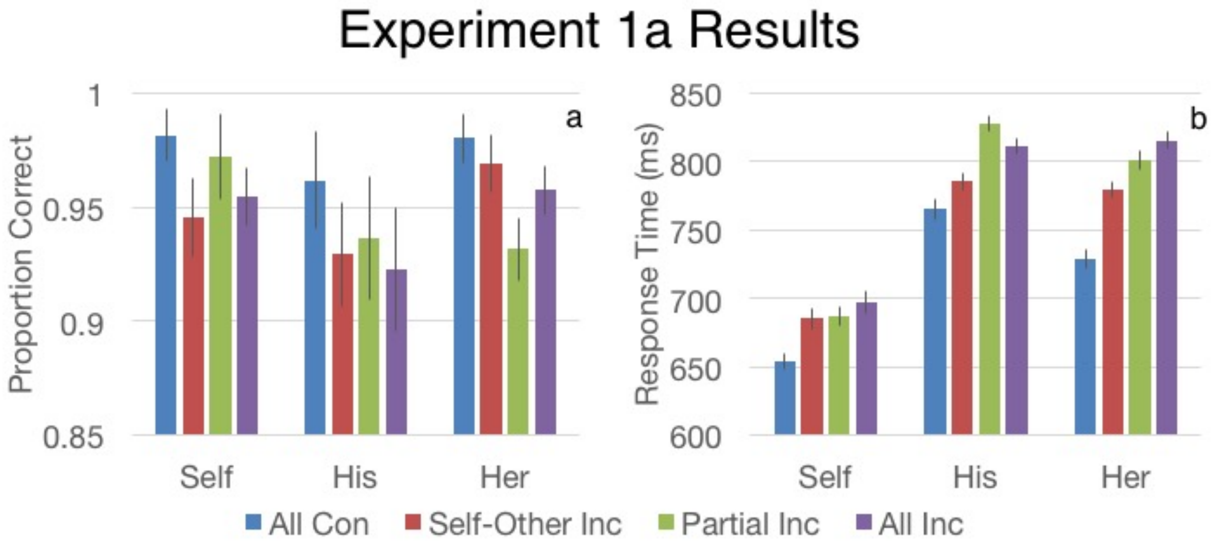


Figure 2.3. Processing costs in Experiment 1 by Perspective cued and Consistency type. Participants were slower to respond to their own perspectives (Self) when any combination of alternative perspectives differed from their own. Error bars here and throughout the remaining experiments represent within-subject standard error using the procedure outlined in O'Brian & Cousineau (2014).

Results

Both accuracy and response times were analyzed (Figure 2.3). Unless otherwise stated, all tests were two-tailed, performed within-subjects and reflect correction for false discovery rate using the Benjamini-Yekutieli-Hochberg method (Benjamini & Yekutieli, 2001).

Accuracy

The primary analysis was whether the avatar's perspective influenced accuracy for Self trials. On these critical Self trials, participants were significantly more accurate on All Consistent trials than S-O Inconsistent trials (mean difference = 3.6%, $t_{23} = 2.36$, $p = .014$, $d = .49$) and All Inconsistent trials (mean difference = 2.7%, $t_{23} = 2.20$, $p = .020$, $d = .46$). All Consistent trials were not significantly different from Partial Inconsistent trials (mean difference = 0.9%, $t_{23} = 0.80$, $p = .216$).

Overall results reveal effects of Perspective and Consistency. A 3 (Perspective) by 4 (Consistency) within-subjects ANOVA revealed a main effect of Perspective ($F_{2,46} = 6.654$, $\eta_p^2 = .219$, $p = .003$) and a main effect of Consistency ($F_{3,69} = 11.66$, $\eta_p^2 = .336$, $p < .001$), as well as a Perspective*Consistency interaction effect ($F_{6,138} = 4.638$, $\eta_p^2 = .168$, $p < .001$). Participants were significantly less accurate when taking the male avatar's ("His") perspective (93.6% accurate) than either the female avatar's (95.8%; $t_{23} = 1.684$, $p = .05$, $d = .35$) or their own (mean: 96.2%; $t_{23} = 2.308$, $p = .015$, $d = .48$) perspective (p's < .002). There was no significant difference between Self and Her perspectives ($t_{23} = .309$, $p = .380$). Participants were significantly more accurate in the Consistent condition (97.5%) than in the Self-Other Inconsistent (94.8; $t_{23} = 4.450$, $p < .001$, $d = .928$), Partial Inconsistent (94.7; $t_{23} = 4.451$, $p < .001$, $d = .929$) or All Inconsistent trials (94.5%; $t_{23} = 4.311$, $p < .001$, $d = .899$). There were no significant differences between Inconsistent conditions (t_{23}'s < 1.169, p's >.873)

Response Times

Again, the critical measure was whether perspective type influenced response times in Self trials. Following accuracy results, participants were significantly delayed in Self trials when any perspective was inconsistent with their own (all t's > 1.83, p's < .05). Inconsistent conditions did not differ significantly (all p's > .45).

A 3 (Perspective: YOU, HIM, HER) by 4 (Consistency Type) within-subjects ANOVA revealed significant main effects of Perspective ($F_{2,23} = 94.12$, $\eta_p^2 = .804$, $p < .001$) and

Consistency ($F_{3,23} = 31.53$, $\eta_p^2 = .578$, $p < .001$), as well as a Perspective*Consistency interaction ($F_{6,23} = 2.675$, $\eta_p^2 = .104$, $p = .017$). Consistent with egocentric bias, pairwise comparisons revealed that participants were faster when taking their own perspective (mean: 675 ms) than when taking the perspective of either the male (794 ms; $t_{23} = 9.911$, $p <.001$, $d = 2.06$) or the female avatar (778 ms; $t_{23} = 11.430$, $p <.001$, $d = 2.38$). Furthermore, responses for each Consistency type were significantly different from one another (lowest $t_{23} = 2.353$, $p < .018$) with the exception of All Consistent and Partial Inconsistent trials ($t_{23} = .367$, $p = .358$).

There was a slight tendency for participants to respond faster to the female avatar's perspective than the male's perspective (mean difference = 16.40 ms, $t_{23} = 2.57$ $p = .008$, $d = .54$). Importantly, there was no difference in Self-perspective, Partial-Inconsistent trials depending on which avatar saw the same number of targets as the participant ($t_{23} = .363$, *ns*). There were no interactions of participant sex with any experimental conditions ($F_{1,23} = .152$, $p = .440$), as might be predicted if individuals preferred the avatar that matched their gender.

The differences between Male and Female avatars is noteworthy, as it is possible that differences in perspective type may be masked by a Simon effect. Briefly, the Simon effect is the robust finding that response times are quicker when a stimulus appears in the same spatial region as the response (Simon, 1969). In all our experiments, participants responded "match" with their right index finger and "mismatch" with their left index finger. However, as the female avatar always appeared on the right side of the screen, it is possible that participants were faster to respond when the female avatar saw the correct number of targets.

Controlled Replication

Twenty Vanderbilt undergraduates (mean age = 18.7, 9 female, 2 left-handed) participated for class credit. I ran a replication that counterbalanced avatar location across participants to control for potential Simon effects caused by responding with a hand in a different spatial location than perspective primed. Participants responded by pressing vertically aligned keys centered below the screen with their dominant hand ("correct" with the middle finger and "incorrect" with the index finger). This control was implemented for the remainder of this study. The procedure and design were otherwise identical to Experiment 1a. Results are illustrated in Figure 2.4

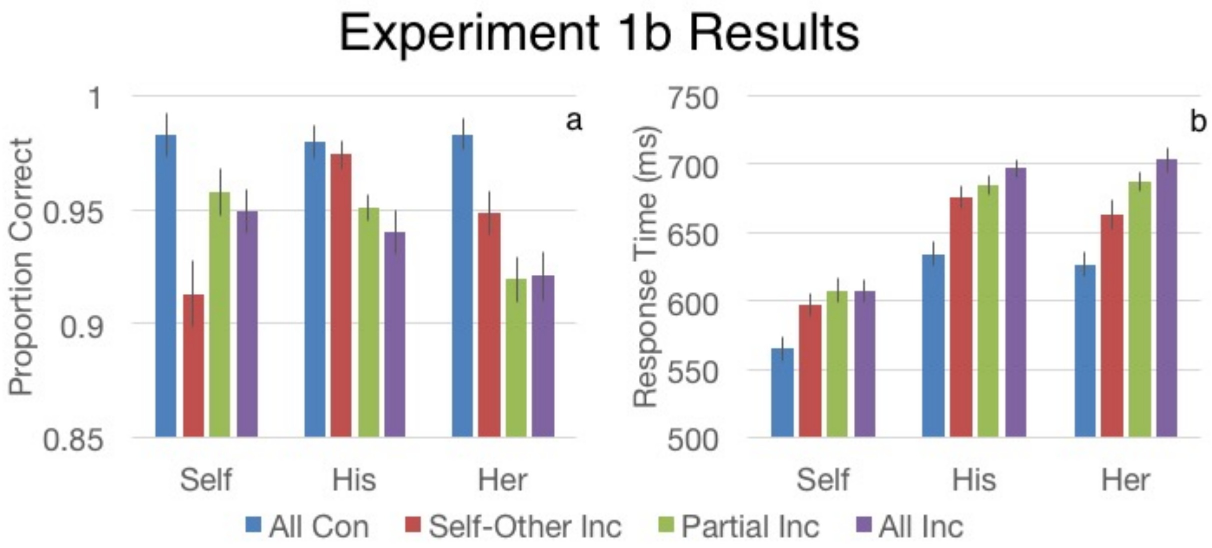


Figure 2.4. Processing cost in the controlled replication of Experiment 1. Participants were significantly delayed in all Inconsistent trial types when responding from their own perspective. There were no differences between the avatars' perspectives (HIS and HERS) within consistency type. Error bars denote standard error.

Accuracy

Critical planned contrasts for Self trials revealed significantly greater accuracy in Consistent trials (98.3%) than Self-Other Inconsistent trials (91.3%, $t_{19} = 3.474$, *p* = .001, *d* = .80) or All-Inconsistent Trials (94.9%; $t_{19} = 1.884$, *p* = .038, *d* = .43), but not greater than Partial-Inconsistent trials (95.7%, $t_{19} = 1.260$, *p* = .112). Additionally, participants were significantly more accurate on Partial Inconsistent trials than Self-Other inconsistent trials ($t_{19} = 1.731$, *p* = .049, *d* = .40).

A 3 (Perspective) by 4 (Consistency Type) within-subjects ANCOVA revealed no effects of handedness ($F_{1,16} = 0.033$, *p* = .858) or avatar location ($F_{1,16} = 1.145$, *p* = .300), but did show a marginal effect of participant gender ($\text{mean}_{\text{female}} = 96.2\%$, $\text{mean}_{\text{male}} = 93.8\%$; $F_{1,16} = 4.018$, *p* = .063) and a significant three-way interaction between Perspective, Consistency and Gender ($F_{6,108} = 2.926$, *p* = .011). Specifically, males were less accurate on Self perspective S-O Inc trials (mean difference = 8.18%, $t_{19} = 2.00$, *p* = .03, *d* = .46) and Her perspective Partial Inc trials (mean difference = 8.26%; $t_{19} = 3.67$, *p* = .002, *d* = .84). As previous studies have identified no differences by gender, and as there are no principled reasons why differences should occur in these instances, all subsequent analyses were performed without gender as a covariate. A within-subjects ANOVA on participant accuracy revealed significant effects of Perspective ($F_{2,38}$ = 4.287, p = .021) and Consistency ($F_{3,57}$ = 11.781, p < .001), as well as a Perspective*Consistency interaction ($F_{6,114}$ = 3.826, p = .002). Paired comparisons revealed only marginal differences between Perspectives across all conditions (highest t_{19} = 1.530, *p*'s > .071). Paired comparisons by Consistency found significantly greater accuracy in All Consistent trials (98.2%) than Self-Other Inconsistent (94.5%; t_{19} = 4.047, p < .001, d = .93), Partial Inconsistent (94.2%; t_{19} = 4.448, p < .001, d = 1.02) or All Inconsistent trials (93.7%; t_{19} = 4.040, p < .001, d = .92). Inconsistent trials did not differ from one another (highest t_{19} = .566, *p*'s > .711).

Response Times

Paired comparisons again revealed that participants were slower on Self trials when in the presence of any form of inconsistent perspective (lowest t_{19} = 2.135, all *p*'s < .023, *d*'s > .49). There were no differences in response times between any of the Self-Inconsistent trials (highest t_{19} = 1.206, p = .88).

A 3 (Perspective) by 4 (Consistency) mixed effects ANCOVA found no effects of participant sex ($F_{1,16}$ = 2.349, p = .145), handedness ($F_{1,16}$ = .055, p = .818), or avatar location ($F_{1,16}$ = 1.249, p = .280) on participant response times. A within-subjects ANOVA on participant response times revealed significant main effects of Perspective ($F_{3,38}$ = 42.800, η_p^2 = .693, p < .001) and Consistency ($F_{3,57}$ = 18.810, η_p^2 = .498, p < .001), but no significant Perspective*Consistency interaction ($F_{6,114}$ = 0.859, p = .547).

Experiment 1b successfully controlled for differences in response times by avatar. Pairwise comparisons revealed no differences in response times to the male or female avatar's perspective (t_{19} = .843, *ns*). Across all conditions, participants were significantly faster in Self trials (594 ms) than when taking His (672 ms; t_{19} = 5.951, p < .001, d = 1.37) or Her perspective (671 ms; t_{19} = 7.702, p < .001, d = 1.77). There was no difference between avatar perspectives (t_{19} = 1.231, p =.88, d = 1.77). Participants were significantly faster across all Perspective types in Consistent trials (610 ms) than Self-Other Inconsistent (647 ms; t_{19} = 3.053, p = .003, d = .70), Partial Inconsistent (659 ms; t_{19} = 4.313, p < .001, d = .99) or All Inconsistent trials (670 ms; t_{19} = 5.053, p < .001, d = 1.16). Overall response times for Self-Other Inconsistent trials were faster than All Inconsistent trials (t_{19} = 3.310, p = .003, d = .69). There were no other differences between Inconsistent trial types (highest t_{19} = 1.177, p > .126).

The critical findings of Experiment 1a were replicated. Participants in Self Perspective All Consistent trials were significantly faster and less error prone than in Self-Other Inconsistent or All Inconsistent trials, and significantly faster than Partial Inconsistent trials.

Discussion

Experiments 1a and 1b demonstrated that individuals show a categorical increase in processing costs when in the presence of any inconsistent perspective, and aside from a small difference in accuracy between Partial-Inconsistent and Self-Other Inconsistent trials in Experiment 1b, revealed no differences between types of inconsistency. These processing costs appeared even when controlling for avatar placement, gender, handedness and Simon effect confounds.

This evidence suggests that the default perspective-taking system does not necessarily represent all perspectives in a scene, but may rather signal that any other perspective differs from one's own. In other words, the default system may be a heuristic signal of privileged knowledge. This tentatively supports an egocentric deviation hypothesis, in which the default perspective-taking system responds whether a perspective deviates from one's own perspective, but is not sophisticated enough to differentiate between multiple diverging perspectives.

However, it is also possible that the early system of perspective taking *does* respond to increasing magnitudes of differing perspectives but that the design of Experiment 1 was insufficient in capturing this nuance. It may be the case that the early system responds to social convergence or divergence only when it is necessary to attend to differences in groups of perspectives. For this reason, Experiment 2 tested whether multiple differing perspectives influenced perspective taking when participants were required to attend to groups, not individuals.

Experiment 2

Experiment 1 found that the presence of *any* perspective different from one's own elicited errors and response delays. Additional conflicting perspectives, which at times aligned with or against one's personal perspective, appeared to have no impact on response times, and largely did not alter error rates. Experiment 2 tested the additional possibility that the early system of perspective taking may be more sensitive to agreement or conflict between multiple other perspectives if the task demanded attention to whole groups of agents.

Experiment 2 replicated Experiment 1 with one critical difference. Instead of responding to the individual perspectives of Self, Him and Her, participants reported their own perspective or the perspective of one or both avatars. In Either trials, participants responded whether either one of the avatars could see the cued number of targets in a display. In Both trials, participants responded whether both of the avatars saw the exact number of cued targets in the display. If the early system is sensitive to the magnitude of conflicting perspectives when all perspectives are task relevant, then we expect response delays in Self trials that were not seen in Experiment 1. If delays do not differ between inconsistent trials, then it is likely that the early system of perspective taking does not individualize alternative perspectives, regardless of task demands.

Method

Participants

Thirty-five Vanderbilt undergraduate students participated for class credit. Twelve participants were excluded for chance performance on all Self trials, indicating a misunderstanding of instructions in this more complicated design. These participants responded in a manner suggesting they interpreted the number cue to mean “at least X objects” rather than “exactly X objects.” This left 23 subjects for analysis (mean age = 18.9, 17 female, 18 right-handed).

Stimuli & Procedure

The stimuli and procedure were identical to those used in Experiment 1b, save for a critical change in the Perspective cues. The Perspective cues in Experiment 1 prompted participants towards their own or either the male or female avatar’s individual perspectives. Perspective cues in Experiment 2 instead directed participants to report what Both avatars could see in a display or what Either avatar could see, in addition to their own (Self) perspective. The timing of trials remained identical to Experiment 1: participants viewed a Perspective cue, followed by a number cue and a display, and participants responded whether exactly the number of objects cued could be seen by the perspective cued. For example, participants given a “Both” cue followed by a “1” and then saw the display in Figure 2.2b, they would give a “yes” response, but they would not if they saw Figure 2.2c, as both avatars do not see *exactly* one target. However, participants given an “Either” cue would respond affirmatively to Figure 2.2c if they saw either a “1” or a “3” as at least one of the avatars can see one or three objects. Trial conditions were counterbalanced and completely randomized over six blocks. Avatar location was randomly assigned to each participant, with 12 seeing the female avatar on the left and male avatar on the right, and 11 seeing the reverse.

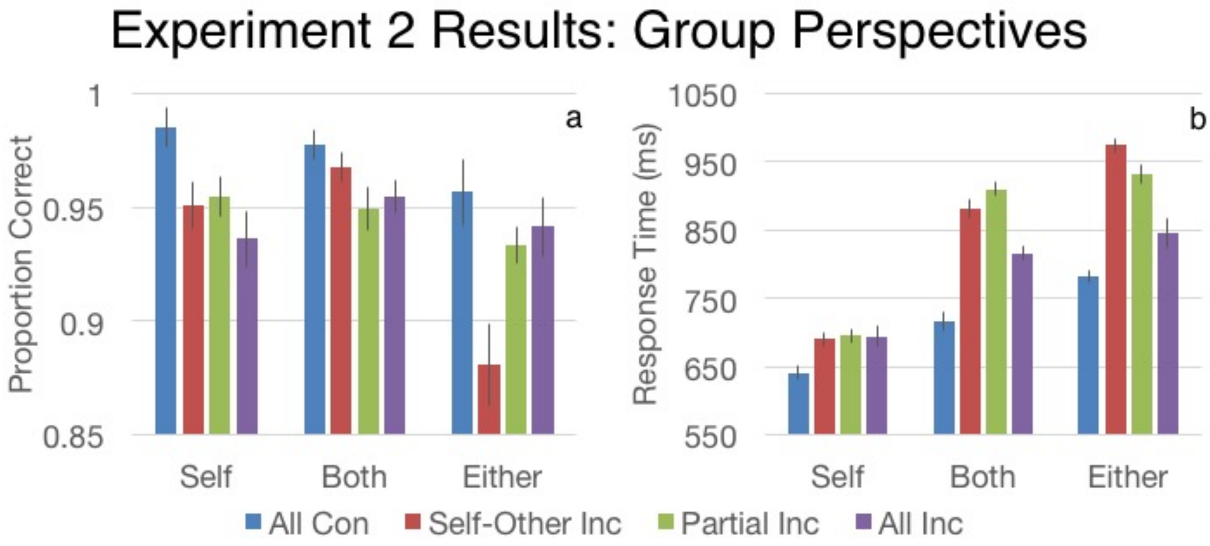


Figure 2.5. Processing costs in Experiment 2. Participants were significantly delayed on Self trials when any combination of perspectives differed from their own. Error bars denote standard error of the mean.

Results

Accuracy

Looking specifically at Self-perspective trials, paired comparisons revealed that Consistent trials (98.2%) were significantly more accurate than Self-Other Inconsistent (93.2%; $t_{22} = 3.263$, $p = .002$, $d > .681$), All Inconsistent (94.6%) and Partial Inconsistent (95.3%) trials (all $t_{22} > 1.95$, $p < .034$, $d > .473$). There were no differences between types of inconsistency (all $t_{22} < .525$, $p < .303$).

Overall accuracy data is illustrated in Figure 2.5a. A 3 (Perspective) by 4 (Consistency) mixed effects ANCOVA found no effects of avatar location ($F_{1,19} = 1.461$, $p = .242$), participant gender ($F_{1,19} = 0.064$, $p = .802$) or handedness ($F_{1,19} = 1.863$, $p = .188$). A within-subjects ANOVA revealed significant effects of Perspective ($F_{2,44} = 6.696$, $p = .003$, $\eta_p^2 = .233$) and Consistency ($F_{3,66} = 12.190$, $p <.001$, $\eta_p^2 = .356$), and a Perspective*Consistency interaction ($F_{6,132} = 3.961$, $p = .001$, $\eta_p^2 = .153$). Paired comparisons by Perspective indicated that participants were significantly less accurate on Either trials (93.1%) than Self (95.2%; $t_{22} = 2.131$, $p = .044$, $d = .45$) or Both trials (96.1%; $t_{22} =4.174$, $p < .001$, $d = .89$), but there was no difference between Self and Either trials ($t_{22} =1.264$, $p = .220$). Similar analyses by Consistency revealed that participants were significantly more accurate in Consistent trials (97.2%) than Self-Other Inconsistent (94.3%; $t_{22} = 4.813$, $p < .001$, $d = 1.03$), Partial Inconsistent (94.6 %, $t_{22} = 3.859$, $p < .001$, $d = .82$) and All Inconsistent trials (93.4%; $t_{22} = 5.477$, $p < .001$, $d = 1.17$). There were no differences in accuracy between Inconsistent types (lowest $t_{22} = 1.712$, $p > .101$)

*Response Time*s

Analysis of critical Self trials revealed that Consistent trials (651 ms) were significantly faster than Self-Other Inconsistent (708 ms; $t_{22} = 4.384$, $p = .001$, $d =1.06$), Partial Inconsistent (708 ms; $t_{22} = 2.756$, $p = .007$, $d =.668$) and All Inconsistent trials (706 ms; $t_{22} = 2.478$, $p = .012$, $d =.601$). There were no differences between Inconsistent trial types (highest $t_{22} = 1.539$, $p > .929$).

Overall response time data is visualized in Figure 2.5b. A 3 (Perspective) by 4 (Consistency) mixed effects ANCOVA found no effects of avatar location ($F_{1,19} = 2.822$, $p = .109$), participant gender ($F_{1,19} = 0.242$, $p = .629$) or handedness ($F_{1,19} = 1.210$, $p = .285$). A within-subjects ANOVA revealed main effects of Perspective ($F_{2,44} = 185.713$, $\eta_p^2 = .894$, $p < .001$) and Consistency ($F_{3,66} = 82.794$, $\eta_p^2 = .790$, $p < .001$), as well as a significant Perspective*Consistency interaction ($F_{6,132} = 14.272$, $\eta_p^2 = .393$, $p < .001$). Paired comparisons by Perspective revealed that Self trials (689 ms) were significantly faster than Either trials (887 ms; $t_{22} = 15.018$, $p < .001$, $d = 3.13$) and Both trials (834 ms; $t_{22} = 15.191$, $p < .001$, $d = 3.17$). Both trials were also significantly faster than Either trials ($t_{22} = 6.169$, $p < .001$, $d = 1.29$). Paired comparisons between Consistency types revealed that Consistent trials (717 ms) were significantly faster than Self-Other Inconsistent (784 ms; $t_{22} = 8.186$, $p < .001$, $d = 1.75$), Parial Inconsistent (845 ms; $t_{22} = 12.365$, $p < .001$, $d = 2.64$), or All Inconsistent trials (847 ms; $t_{22} = 9.805$, $p < .001$, $d = 2.09$). Every Inconsistent type differed from every other Inconsistent type (lowest $t_{22} = 6.771$, $p < .001$, $d > 1.44$), with the exception of All Inconsistent and Partial Inconsistent trials (mean difference = 1.6 ms; $t_{22} = 0.207$, $p = .838$).

Discussion

Experiment 2 tested whether attending to groups of agents altered processing costs associated with multiple inconsistent perspectives. Results replicated those found in Experiment 1: participants incurred processing costs when any perspective differed from their own, regardless of the number or overlap of differing perspectives. These results support the egocentric deviation hypothesis. The early system appears to calculate the *presence* of differing perspectives, but is not sensitive to the magnitude or contents of those perspectives. This suggests that the early system is not sensitive to magnitudes of differing perspectives. These experiments therefore provide evidence against the assertion that individuals automatically represent other's perspectives in their entirety (as with Leslie, Friedman & German, 2004) or selectively guides attention to the content's of another's perspective (Apperly & Butterfill, 2009). Rather, individuals may merely be sensitive to scenes where their own knowledge may be privileged.

One possible mechanism for the early system of perspective taking is that it exogenously cues attention to the contents of another's visuospatial perspective. Experiments 1-2 may demonstrate the limits of such an attentional mechanism, in that it stops once a single differing perspective is detected. However, if attention is captured by differences in perspectives, then the presence of an inconsistent perspective should induce cognitive costs even if the difference in perspectives is irrelevant to one's current task. Experiment 3 tests whether the cognitive costs associated with default perspective-taking appear only when perspectives differ to a relevant attentional set.

Experiment 3

One potential mechanism for the early system of perspective-taking might be to selectively guide attention to objects or regions where another agent cannot see. Perspective taking would thus constitute a rapid, broad gist similar to the perception of scene categories (Greene & Oliva, 2009). However, this seems highly unlikely given the sheer number of objects that another agent may or may not have access to in a real world scene. Take the example of a party. You see a friend go towards a table with several cans of soda, some hidden behind a punch bowl. Does the early perspective-taking system account for this incongruence if your friend has expressly asked for a slice of pizza, and he is to your knowledge indifferent towards a beverage? Moreover, what if the table is laden with miscellaneous snacks, with multiple objects partially or fully occluded from your friend's gaze? We can see that even a relatively mundane occurrence asks a great deal of an early-attentive perspective-taking system.

Whereas Experiments 1 and 2 tested the sensitivity of default system perspective-taking to multiple agents, Experiment 3 tested the default system's selection among multiple competing perspectives of a single agent. This experiment thus addressed two potential limitations of default perspective-taking. First, it tested whether the early system differentiates between two perspectives from the same agent. Second, it tested whether default perspective calculation is limited to the immediate attended set of features.

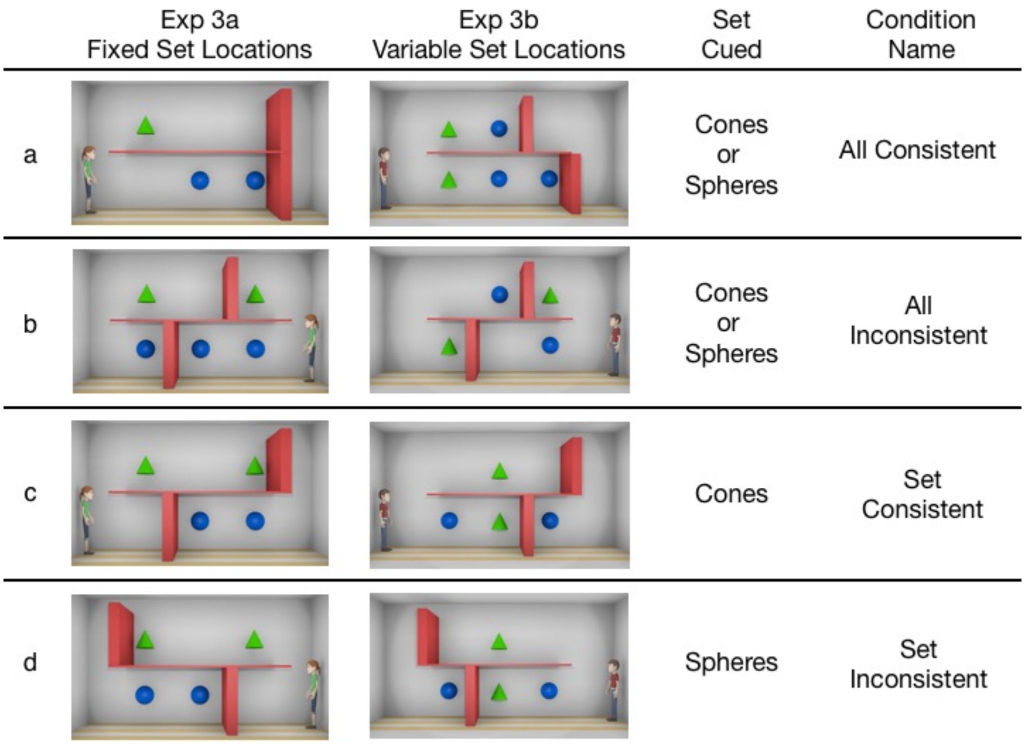


Figure 2.6. Multiple-set stimuli in Experiments 3a and 3b. Sets ranged from 0-3 items present in two sets (blue spheres or green cones), and wall segments moved independently to occlude 0-3 items from each set from the avatar's perspective. Note that the only difference between Set Consistent and Set Inconsistent trials is the set cued. Stimuli in Experiment 3b were identical save that blue or green sets were mixed between levels.

Method

Participants

Twenty-five Vanderbilt University undergraduates participated for class credit. One participant was excluded for performing more than three standard deviations below the mean in accuracy, leaving a total of 24 participants (mean age = 19.0, 17 Female, all right-handed).

Stimuli & Procedure

The design of displays was similar to Experiment 1 of Baker, Levin & Saylor (2015). Each display contained two sets of targets: a set of zero to three blue spheres and a set of zero to three green cones aligned horizontally (Figure 2.6). A thin wall centered at the avatar's eye height divided the objects into two rows. Each set of objects was potentially divided by a vertical wall in one of four locations on a horizontal axis. Cones appeared on the top row and spheres appeared on the bottom row. As with previous experiments, participants saw a perspective cue (YOU or HIM/HER) followed by a number cue (0-3). However, in this experiment the number cue was colored either blue or green. Participants responded whether the perspective primed could see the exact number of cued objects in the matching color set.

This design yielded four conditions (*Figure 2.6*). In All Consistent trials, participants and avatars saw the same number of both sets of objects. In All Inconsistent trials, participants and avatars saw different numbers of both blue spheres and green cones. In Set Consistent trials, participants and avatars saw the same numbers of targets but different numbers of distracters. Conversely, participants and avatars saw the same numbers of distracters but different numbers of targets in Set Inconsistent trials. If the early system activates after attentional selection, then processing costs should be reduced or absent in Set Inconsistent trials. However, if the default system guides attention objects outside of one's attentional set, then participants should demonstrate processing costs for both Set Consistent and Distracter Inconsistent trials.

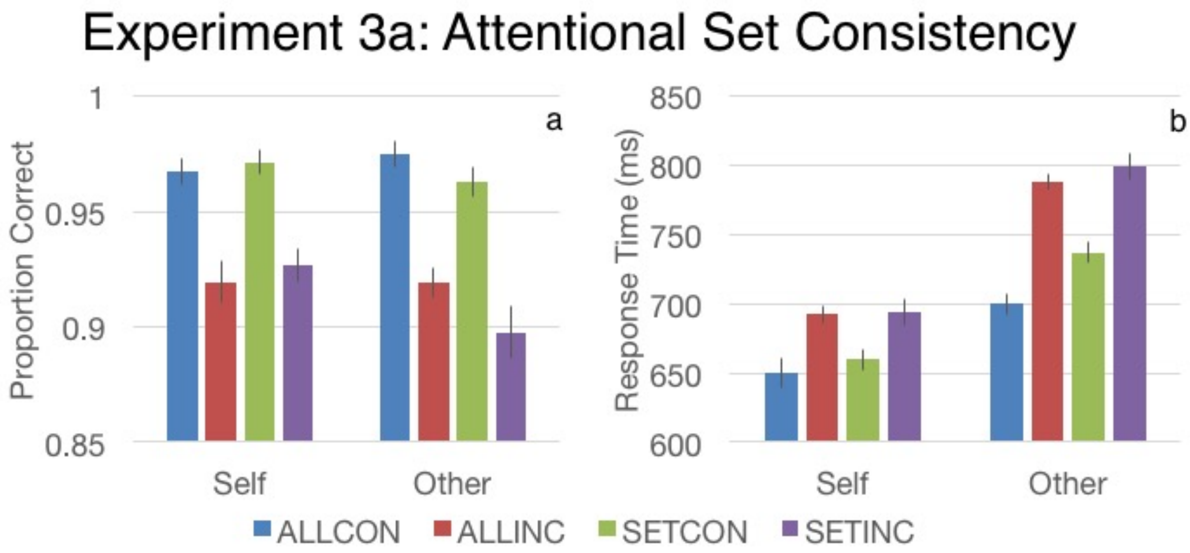


Figure 2.7. Processing costs in Experiment 3. Participants incurred processing costs on Self trials when another's perspective of an attended set differed from their own perspective. However, participants were unaffected by inconsistent perspectives of the unattended sets..

Results

Accuracy

Looking at critical Self-perspective trials, paired comparisons revealed that All Consistent trials (96.7%) were significantly more accurate than All Inconsistent trials (91.9%; $t_{23} = 4.237$, $p < .001$, $d = .88$) and Set Inconsistent trials (92.6%; $t_{23} = 3.893$, $p < .001$, $d = .81$). In agreement with a top-down control hypothesis, participants did not differ in their accuracy between All Consistent and Set Consistent trials (97.1%; $t_{23} = 0.029$, $p = .511$). Likewise, participants were significantly more accurate in Set Consistent trials than All Inconsistent ($t_{23} = 4.695$, $p < .001$, $d = .98$) and Set Inconsistent trials ($t_{23} = 4.278$, $p < .001$, $d = .89$). All Inconsistent trials did not differ from Set Inconsistent trials ($t_{23} = 0.028$, $p = .511$).

Overall accuracy data is illustrated in Figure 2.7a. A 2 (Perspective) by 4 (Consistency) within-subjects ANOVA revealed a significant effect of Consistency ($F_{3,69} = 32.311$, $p <.001$, $\eta_p^2 = .584$) and a Perspective*Consistency interaction ($F_{3,69} = 3.015$, $p = .035$, $\eta_p^2 = .116$), but no main effect of Perspective ($F_{1,23} = 1.461$, $p = .239$). Paired comparisons by Consistency indicate that participants were significantly more accurate on All Consistent trials (97.5%) than All

Inconsistent (91.9%, $t_{24} = 6.679$, $p < .001$, $d = 1.39$) and Set Inconsistent trials (91.2%, $t_{24} = 5.385$, $p < .001$, $d = 1.12$), but were not significantly different from Set Consistent trials (96.7%; $t_{24} = 0.345$, $p = .37$). All Inconsistent trials also did not differ from Set Inconsistent trials ($t_{24} = .259$, $p = .400$).

*Response Time*s

Analysis of critical Self trials revealed that Consistent trials (650 ms) were significantly faster than All Inconsistent (692 ms; $t_{23} = 2.368$, $p = .013$, $d =.49$) and Set Inconsistent trials (694 ms; $t_{23} = 2.355$, $p = .014$, $d =.491$), but did not differ from Set Consistent trials (660 ms; $t_{23} = .286$, $p = .611$). Set Consistent trials significantly differed from both All Inconsistent ($t_{23} = 3.973$, $p = .002$, $d = .83$) and Set Inconsistent Trials ($t_{23} = 3.244$, $p < .002$, $d = .68$). Inconsistent trials were not significantly different ($t_{23} = 1.192$, $p = .123$).

Overall response time data is visualized in Figure 2.7b. A 2 (Perspective) by 4 (Consistency) within-subjects ANOVA revealed main effects of Perspective ($F_{1,23} = 89.322$, $\eta_p^2 = .989$, $p < .001$) and Consistency ($F_{3,48} = 38.400$, $\eta_p^2 = .975$, $p < .001$), as well as a significant Perspective*Consistency interaction ($F_{6,96} = 4.813$, $\eta_p^2 = .173$, $p = .004$). Paired comparisons by Perspective revealed that Self trials (673 ms) were significantly faster than Other trials (756 ms; $t_{23} = 9.503$, $p < .001$, $d = 1.98$). Across both perspectives, paired comparisons between consistency types revealed that All Consistent trials (675 ms) were significantly faster than All Inconsistent (741 ms; $t_{23} = 6.868$, $p < .001$, $d = 1.43$) and Set Inconsistent trials (747 ms; $t_{23} = 6.658$, $p < .001$, $d = 1.39$). All Consistent trials were also significantly faster than Set Consistent trials (698 ms; $t_{23} = 2.285$, $p = .016$, $d = 0.48$). Set Consistent trials were faster than both All Inconsistent ($t_{23} = 6.658$, $p < .001$, $d = 1.39$) and Set Inconsistent trials ($t_{23} = 6.224$, $p < .001$, $d = 1.30$). Inconsistent trials did not significantly differ ($t_{23} = 0.367$, $p = .358$).

Discussion

Experiment 3 demonstrated that behaviors associated with default perspective-taking only emerge after attentional selection. Participants only incurred processing costs when an actively attended subset of objects were seen differently by another agent. This runs counter to a mechanism that guides attention to regions of conflicting perspectives. Rather, it appears that the perspective-taking module only activates *after* selection of a set of objects. This is consistent with a heuristic decision-making mechanism, by which the early system signals that one's perspective of an attended set is privileged. This scenario points to early-system perspective-taking as more of a decision making heuristic than a perceptual module.

However, Experiment 3 contained a potential confound that may have facilitated subsetting. Sets of green cones always appeared in the upper row, while sets of blue spheres always appeared in the lower row. The predictable and non-overlapping locations of each set may have reduced effort required for search and minimized potential perspective-taking effects in the unattended channel. I therefore ran a control study that replicated the exact stimuli and procedure used in Experiment 3, with the critical difference that green and blue sets were randomly presented in either the top or bottom rows (Figure 2.6). If default perspective taking behavior only emerges after selection of a set, and therefore ignores perspective differences in the unattended set, then participants should similarly show no processing costs in the Set Consistent trials with these new stimuli.

Controlled Replication

Twenty-five Vanderbilt undergraduates participated for class credit. One participant was dropped for performing over three standard deviations below the mean in accuracy, leaving 24 participants for analysis (mean age = 19.5, 18 female, 21 left-handed). The stimuli and procedure of this replication experiment were identical to the previous experiment, with the notable exception that target cones and spheres could potentially appear in both the top and bottom rows of the display. Positions of targets and distractors were randomized, though importantly these new stimuli otherwise mirrored the stimuli from Experiment 3a. For instance, if an All Inconsistent display from Experiment 3a showed three cones and two spheres, and the avatar could only three one cone and one sphere, the same was true of the complementary display in Experiment 3b with the exception that targets were mixed between rows. Overall accuracy and response time data are illustrated in Figure 2.8.

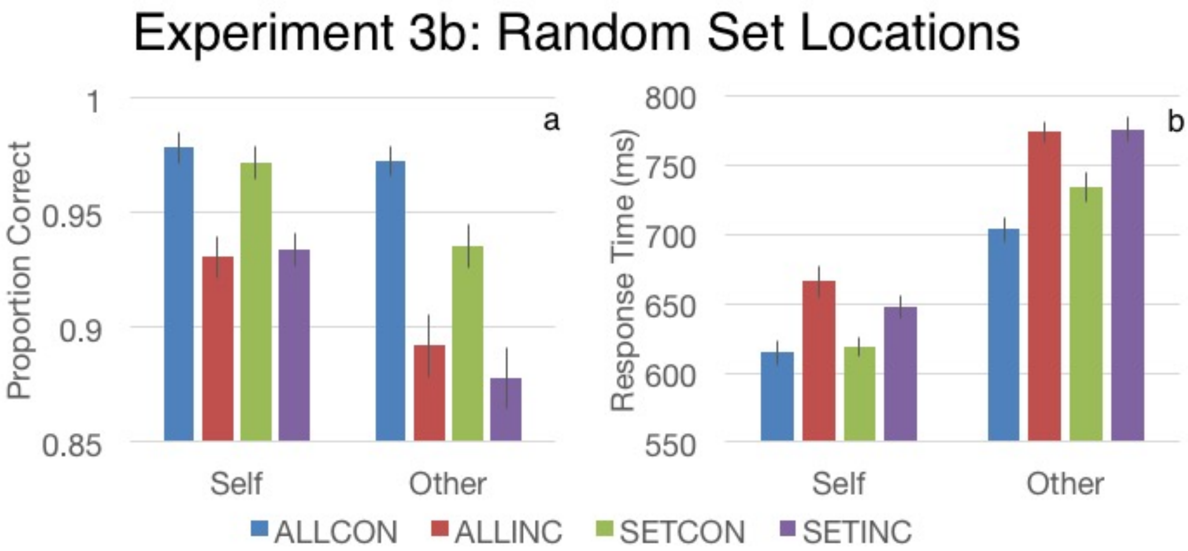


Figure 2.8. Processing cost in the controlled replication of Experiment 3. Participants were significantly delayed in all Inconsistent trial types when responding from their own perspective. There were no differences between the avatars' perspectives (HIS and HERS) within consistency type. Error bars denote standard error of the mean.

Accuracy

Critical planned contrasts for Self trials revealed significantly greater accuracy in Consistent trials (97.8%) than All Inconsistent (93.1%, $t_{23} = 4.125$, $p = .001$, $d = .86$) and Set Inconsistent trials (93.4%, $t_{23} = 4.230$, $p = .001$, $d = .86$), but did not differ from Set Consistent trials (97.2%, $t_{23} = 0.782$, $p = .221$). Likewise, Set Consistent trials significantly differed from All Inconsistent ($t_{23} = 3.820$, $p < .001$, $d = .80$) and Set Inconsistent Trials ($t_{23} = 3.819$, $p < .001$, $d = .79$). All Inconsistent trials did not differ from Set Inconsistent trials ($t_{23} = 0.417$, $p = .660$).

A 2 (Perspective) by 4 (Consistency) within-subjects ANOVA revealed a significant effect of Perspective ($F_{3,69} = 22.824$, $p <.001$, $\eta_p^2 = .498$), Consistency ($F_{3,69} = 28.723$, $p <.001$, $\eta_p^2 = .555$) and a Perspective*Consistency interaction ($F_{3,69} = 4.482$, $p =.006$, $\eta_p^2 = .163$). Participants were significantly more accurate across consistency trial types in Self trials (95.3%) than Other trials (91.7%; $t_{23} = 4.786$, $p <.001$, $d = 1.00$). Paired comparisons by Consistency indicate that participants were significantly more accurate on All Consistent trials (97.6%) than All Inconsistent (91.1%, $t_{23} = 6.305$, $p < .001$, $d = 1.31$), Set Inconsistent (90.6%, $t_{23} = 6.312$, $p < .001$, $d = 1.32$), and also Set Consistent trials (95.4%; $t_{23} = 3.851$, $p <.001$, $d = .80$). However, Set Consistent trials were significantly more accurate than both All Inconsistent ($t_{23} = 3.719$, $p <$

.001, $d = .78$) and Set Inconsistent trials ($t_{23} = 3.851$, $p < .001$, $d = 0.80$). All Inconsistent trials did not differ from Set Inconsistent trials ($t_{24} = .155$, $p = .439$).

Response Times

Analysis of critical Self trials revealed that Consistent trials (615 ms) were significantly faster than All Inconsistent (666 ms; $t_{23} = 3.854$, $p < .001$, $d =.80$) and Set Inconsistent trials (648 ms; $t_{23} = 3.13$, $p = .002$, $d =.65$), but did not differ from Set Consistent trials (618 ms; $t_{23} = .425$, $p = .663$). Set Consistent trials significantly differed from both All Inconsistent ($t_{23} = 3.973$, $p = .002$, $d = .83$) and Set Inconsistent Trials ($t_{23} = 4.162$, $p < .001$, $d = .87$). Inconsistent trials were significantly different ($t_{23} = 1.822$, $p = .041$, $d = .37$).

A 2 (Perspective) by 4 (Consistency) within-subjects ANOVA revealed a significant effect of Perspective ($F_{3,69} = 131.640$, $p <.001$, $\eta_p^2 = .851$), Consistency ($F_{3,69} = 44.953$, $p <.001$, $\eta_p^2 = .662$) and a Perspective*Consistency interaction ($F_{3,69} = 3.209$, $p =.028$, $\eta_p^2 = .122$). Paired comparisons by Perspective across consistency trial types found that participants were significantly faster in Self trials (638 ms) than Other trials (749 ms; $t_{23} = 11.503$; $p < .001$; $d = 2.40$). Paired comparisons by Consistency indicate that participants were significantly faster on All Consistent trials (659 ms) than All Inconsistent (720ms; $t_{23} = 8.629$, $p < .001$, $d = 1.80$), Set Inconsistent (712 ms; $t_{23} = 7.953$, $p < .001$, $d = 1.66$), and Set Consistent trials (676 $t_{23} = 2.051$, $p =.026$, $d = .43$). However, Set Consistent trials were significantly faster than both All Inconsistent ($t_{23} = 6.100$, $p < .001$, $d = 1.27$) and Set Inconsistent trials ($t_{23} = 5.625$, $p < .001$, $d = 1.17$). All Inconsistent trials did not differ from Set Inconsistent trials ($t_{24} = 1.163$, $p = .128$).

The critical findings of Experiment 3a were replicated. Participants were slower and more error prone when identifying targets of an attended set that could not be seen by another agent, even when the locations of these objects were unpredictable and mixed with distractor objects. If, as some suggest, the early system of perspective taking directs attention to the contents of another's visuospatial access, then participants should have incurred processing costs when an agent held an inconsistent perspective of any set of objects. This was not the case across two experiments. These results suggest that the hypothesized default system of perspective-taking does not direct attention to the contents of another's perspective. Rather, it appears that participants incur costs only after attentional selection. These results are consistent with a heuristic at the level of decision-making, where individuals rapidly calculate whether their own perspective is privileged *after* attentional selection.

General Discussion

A robust body of work since Piaget and Inhelder (1956) has suggested that perspective taking is cognitively effortful. Participants demonstrate *egocentric interference* as they set aside their own view of a scene to interpret what another person may see (i.e., Keysar, Lin, Barr, 2003; Epley, Morewedge & Keysar, 2004). Samson and colleagues challenged this notion with the introduction of the dot-perspective task, where participants viewing scenes where another agent held a different perspective incurred “altercentric interference”, presumably demonstrating that viewers represent another’s perspective by default. Recently, we demonstrated that the dot-perspective task was robust to low-level confounds such as spatial attentional cuing or perceptual grouping, validating the finding that perceiving a conflicting perspective leads to processing costs (Baker, Levin & Saylor, 2016). However, it remained uncertain exactly what these costs meant for scene perception, attention and representation.

This study found evidence that viewer’s calculations of other’s perspectives are quite limited. Viewers did not incur increasing processing costs as the number of alternative perspectives increased, as would be the case if they actively represented each divergent perspective. Viewers also did not incur processing costs for objects outside of one’s top-down attentional set, indicating that the early system is relegated to items within an attentional set. The absence of processing costs when another’s perspective of ignored sets differs from one’s own further suggests that if the early system guides attention, it does so after initial selection of items. The culmination of these effects leads to a much more limited conceptualization of the early system than previously discussed. Rather than drawing attention to perspectives and scaffolding representation to those perspectives, the early system appears to constitute a heuristic signal that one has privileged access to a scene. This *egocentric deviation* signal only occurs after attentional selection and is indifferent to the magnitude of the difference between perspectives. These results are consistent with an early-developing decision making heuristic that identifies whether attended objects are separated in space from any other agent.

Investigation of the early system to this point has largely concerned its existence: whether the effects ascribed to the phenomenon could be due to confounding variables, and whether the effects constitute domain-specific processing. This study takes a different tactic by tentatively accepting the existence of the early system and then testing its limitations. Many experiments demonstrating this phenomenon are ingeniously designed precisely to test the existence of this phenomenon, but lack specificity to state *how* the proposed system functions. Debates concerning the nature of default perspective-taking will be better informed by investigation on the specific circumstances by which other perspectives automatically influence behavior. Given the current findings, the early system appears to influence decision-making relatively late in the perceptual cycle. Future research should further investigate how the presence of other perspectives affects perception, attention, representation and decision-making before making conclusions on the nature and function of the proposed early system of perspective taking.

Chapter 3

SOCIAL AGENTS BIAS VISUAL SEARCH TO REGIONS OF PRIVLEDGED ACCESS

Abstract

Several recent theories of mental state reasoning suggest that human beings have an early-developing mechanism that directs attention to another agent's beliefs and visual-spatial perspectives. Supporting experiments found that individuals were slower and more error prone in decision making tasks when an agent in a scene sees or knows something differently than the viewer. However, it is unknown how the presence of social agents directs visual attention, if at all. Here, I test the effect of social agents on visual search. Participants responded whether they or a cartoon agent with unrestricted or limited visual access to a scene could see a target among a number of distracters. Consistent with theories of rapid perspective-taking, results suggest that participants altered their behavior in response to the agent's perspective. Somewhat surprisingly, search slopes and intercepts were unaffected by conflict between perspectives, suggesting that other perspectives do not tax visual attention. However, search slopes increased as a function of target distance from the agent, suggesting that individuals preferentially searched locations the avatar *could not see*. These results raise the provocative possibility that a rapid system of perspective-taking biases search to regions of privileged access.

Introduction

Perspective-taking is ubiquitous in human interaction. Navigating social situations requires prediction of the likely behaviors of other agents, their common and privileged knowledge, and one's ability to share this knowledge. Driving through a four-way stop, for instance, requires a number of perspective judgments. Does the other driver see the stop sign? Do they see me? Will they see the sign but nonetheless break the law? In the previous chapter, I explored claims about representation and attention in behaviors associated with taking another's perspective by default. That study supported the claim that individuals are sensitive to the presence of a differing perspective within limits: participants do not discriminate multiple perspectives, nor perspectives outside of the participant's attentional set. These results add to a burgeoning literature suggesting a degree of automaticity to perspective taking. However, it is still unknown how the underlying perspective-taking system functions. One frequent proposal is that social agents guide attention to regions of shared perceptual access. However, it could also be the case that delays associated with automatic perspective-taking are not due to attentional capture, but rather a heuristic decision-making process that delays responding after attentional selection. This study used a visual search paradigm to test how the presence of another social agent guides attention.

Arguments for a Domain-Specific Mechanism of Perspective-Taking

Despite its importance in everyday human life, perspective-taking is fraught with errors. Individuals often assume others share their knowledge (known as *egocentrism*, Piaget & Inhelder, 1956; Flavell, Everett, Croft & Flavell, 1981), and fail at perspective-taking tasks when under time constraints (Epley, Morewedge & Keysar, 2004), cognitive load (Lin, Keysar & Epley, 2010) or when they simply lack motivation (Klein & Hodges, 2001). And yet the fact remains that a tacit level of perspective understanding permeates the most basic of social interactions. People engaged in conversation adopt pronouns that implicitly acknowledge shared knowledge (Brennan & Clark, 1996; Garrod & Pickering, 2009); adults and children anticipate the goals of an action (Baldwin, Anderson, Saffran & Meyer, 2008; Zacks, Kumar, Abrams & Mehta, 2009), so much so that they often infer seeing goals in their absence (Strickland & Keil, 2011); and infants as young as five months look longer at displays that are contrary to an agent’s beliefs, even when the displays are consistent with the infant’s knowledge (Kovacs, Teglas & Endress, 2010). Most recently, adult research has demonstrated that the presence of another’s inconsistent perspective causes participants to be slower and more error prone when giving their *own* perspective (Samson, Apperly, Braithwaite, Andrews & Bodely Scott, 2011), even when controlling for spatial cuing and perceptual grouping effects (Baker, Levin & Saylor, 2016). Why can individuals sometimes calculate other’s knowledge with relative ease, only to show remarkable deficits in mental state reasoning in other tasks?

Several psychologists have proposed theories of perspective taking that merge cognitively effortful, error prone perspective-taking with the seemingly automatic perspective-taking suggested by developmental research (Butterworth & Jarrett, 1991; Gergely & Csibra, 2003; Ferguson & Bargh, 2004; Leslie, Friedman & German, 2005; Onishi & Baillargeon, 2005; Tomasello et al, 2005; Aperly & Butterfill, 2009; Samson et al, 2010; among others). To some degree, all models suggest that an early-developing and potentially modular system facilitates the automatic, limited computation of another's mental state, while a later-developing system elaborates upon those computations. Models diverge in precisely how this early system functions. Some argue that early system guides selective attention to locations of shared perceptual access (Butterworth & Jarrett, 1991; Apperly & Butterfill, 2009; Samson, Apperly, Braithwait, Andrews & Bodely-Scott, 2011). Others suggest that the early system facilitates limited representation of other perspectives (Leslie, Friedman & German, 2004) or otherwise motivates implicit reasoning about another's knowledge (Bargh & Fergusson, 2004; Onishi & Baillargeon, 2005), and that this knowledge influences judgments and attitudes. Still others contend that the default system attributes mental states to agents and actions, presumably leading to interference when mental states differ (Castelli, Happe, Frith & Frith, 2000; Fergusson & Bargh, 2004; Luo & Baillargeon, 2005; Tomasello et al, 2005; Kovacs, Teglas & Endress, 2010). Importantly, none of these theories deny the role of executive control in overt perspective-taking judgments. However, despite that many if not all of these theories allude to attentional guidance by social agents, no study to date has measured whether the presence of social agents influences deployment of attention. By incorporating the designs of perspective-taking studies with a classic test of visual attention, we can begin to investigate claims of rapid guidance of attention by social stimuli.

Visual Search and Perspective Taking

Visual search is one of the most prevalent human behaviors. Our eyes constantly scan for objects and features in our environment, either through top-down control (e.g., looking for one's car keys) or through bottom-up salience (e.g., stark contrasts in motion, color or luminance; Treisman & Gelade, 1980; Duncan & Humphreys, 1989; Wolfe, 1994). Due to the simplicity of the behavior and the high level of laboratory control permitted, the psychophysical study of visual search has become one of the most prevalent and reliable paradigms for studying visual attention. In general, participants in visual search paradigms respond whether a target is present among different numbers of distracters. Researchers derive two measures from participant response times over large numbers of trials. The *slope,* the increasing mean response time as a function of set size, indicates the time required to attend to each item in a display before terminating search. If slope approaches zero, the search may be considered *efficient*, meaning that detection of the stimuli is likely bottom-up and potentially pre-attentive (e.g., searching for a red square in a field of blue squares). Search slopes in excess of 15 ms are generally referred to as *inefficient* (Triesman & Gelade, 1980; Triesman & Gormican, 1988; Wolfe, 1994), meaning that participants must serially attend to individual items, features or groups of features. A second common measure is the response time *intercept*, the projected response time if set size were to equal zero. Differences in intercept are often attributable to task components outside of sequential deployments of attention, including cognitive, motor and decision-making components that accompany the search (e.g., searching for an object and then responding in the direction of its orientation; Wolfe, 1994).

Search slopes and intercepts reveal how bottom-up and top-down attention interact in scene perception. Evidence supports that a variety of efficient processes operate over large portions of the visual field to rapidly guide attention to unique shapes, colors, orientations or movements regardless of the number of items in the scene (Nakayama & Silverman, 1985; Theeuwes & Kooi, 1994). Attentional capture of efficient processes (e.g., onset of a bright or moving stimulus; Theeuwes, 2004; Fockert, Rees, Frith & Lavie, 2004) often increase search intercepts without altering slopes, suggesting that attention is briefly captured by the surprising stimulus before returning to the scene as a whole. Conversely, inefficient attentional processing requires individuals to dwell on individual items or groups of items, devote limited capacity resources to accept or reject each item as a target and then move to another location (Egeth & Yantis, 1997; Zelisnky & Sheinberg, 1997). Inefficient processing is typically identified by increased search slopes, as the identification of targets or processing of items in clusters is disrupted (Logan, 1996). This might occur as stimulus salience depreciates (Duncan & Humphreys, 1989) or as the number of conjunctive diagnostic criteria for target acceptance increases (Nakayama & Silverman, 1985). Notably, most theorists agree that efficient and inefficient processes cooperatively guide attention in visual search (Wolfe, 1994). An early system of perspective taking could therefore reasonably influence attention at multiple levels.

Researchers have demonstrated that higher-order cognition can impact inefficient processes as demonstrated by search slopes. Early work revealed that low-level differences in object category could facilitate efficient visual search (e.g., a left-leaning target among mixed right-leaning distracters; Wolfe et al, 1992). At a higher level, visually complex and abstract semantic object categories can also guide visual search, such as efficient search for animal targets among inanimate distractors (Levin, Takarae, Miner & Keil, 2001). Furthermore, individuals demonstrate more efficient slopes for items in categorically consistent locations (e.g., a blimp in the sky vs. on the ground (Neider & Zelinsky, 2006). Although knowledge-driven search is less efficient than perceptually-driven search (Yang & Zelinsky, 2009), knowledge influences deployment of visual attention to categorical targets far faster than chance alone would predict. Knowledge about mental states may direct visual search in the same manner.

Although no research has explicitly explored visual search during perspective taking, several studies have touched upon the influence of social cognition on search behaviors. Brennan et al. (2008) had pairs of participants search a display by coordinating verbally or by seeing each other's gaze overlaid on the display. Availability of another's gaze overlaid on a display not only improved search above solo trials, but was more efficient than search with speech alone, or gaze and speech combined (Neider et al, 2010). These cooperative search paradigms illustrated that overt knowledge of another perspective can alter visual search strategies. It is therefore possible that implicit mental state knowledge could influence visual search as well.

The Current Study

Various researchers have claimed that a default system of perspective-taking facilitates preferential attention to the targets of shared visual access. However, no experiment has explicitly tested the effect of social agents on attentional selection. In two experiments, I compared visual search behavior during perspective taking, measuring whether visuospatial access of another agent in a scene altered search behavior. It is important to point out that the design and implications of these experiments fundamentally differ from previous perspective-taking designs, including those described in Chapter 2. Nearly every demonstration of implicit perspective-taking behavior in adults and children has limited the number of perspectives and potential targets to a very small number within the range of subitization (McCleery et al, 2011; Qureshi, Apperly & Samson, 2010; Samson, Apperly, Braithwaite, Andrews & Scott, 2010; Samson & Apperly, 2010; Surtees & Apperly, 2012; Baker, Levin & Saylor, 2016). These designs tested whether the presence of a differing perspective led to cognitive load, presumably through conflicting representations of the agent's perspective and one's own perspective. In contrast, the design of the current study tested whether the presence of a differing perspective altered the deployment of attention throughout a scene.

Social agents could theoretically guide attention in a number of ways. Theories suggesting an automatic calculation of shared visual access with another agent (e.g., Samson et al, 2011) might predict an efficient process that calculates an agent's visual access over large portions of the visual field at one time. Such *global perspective-taking* would elicit an overall processing cost in search intercepts when searching a visual scene containing a difference in perspectives. Second, viewers may represent each searched item from the other agent's perspective (as with Leslie, Friedman & German, 2004), eliciting a small but steady cost for every item searched. This *serial perspective-taking* would increase the time needed to attend to each item in a scene containing divergent perspectives, increasing search slopes. Third, participants may change their search strategies in response to the agent, preferentially initiating search to regions accessible by the agent. Participants with *search initiation bias* would scan items near to the avatar first, leading to faster search slopes when the target appears near to the avatar and slower search slopes when targets are far from the avatar. Lastly, participants might demonstrate *response inhibition* if they detect a target and then calculate another's perspective before making a response about the target. This would involve slower search intercepts only for those targets that cannot be scene by the avatar.

Experiment 1

Experiment 1 tested whether the presence of another perspective influences visual search. Experiment 1 combined the perspective-taking task used by Samson, Apperly Braithwaite, Andrews and Bodley Scott (2011) with a visual search paradigm. Participants searched for a target among similar distractors. On half of trials, participants additionally reported whether an avatar in the scene could see the target, which was occluded from the avatar's line of sight on half of trials. This study is principally interested in trials where the participant took their own perspective, as these trials would reveal behavioral differences even when it was beneficial to ignore the avatar. The perspective-taking system might guide visual search in four independent ways. If perspective-taking occurs by capturing attention before search we would expect global perspective taking, as measured by increased intercepts when scenes contain differing perspectives. If perspective-taking requires increased attention to serially process each item in the display we would expect serial perspective-taking costs, illustrated by steeper slopes during when perspectives differ. If perspective taking occurs by capturing attention at the initiation of search we would expect search bias, demonstrated by faster slopes when targets are near the avatar. Finally, if perspective-taking occurs after search termination during response formulation we would expect response inhibition, characterized by increased intercepts only on those trials were the target cannot be seen by the avatar. Any combination of these hypotheses would indicate a change in search behavior evoked by the presence of another agent.

Method

Participants

Sample sizes were based upon data from a previous experiment integrating visual search and a secondary task (Woodman, Vogel & Luck 2001). In that study, an ANOVA testing search intercepts by increasing working memory load revealed an estimated effect size η_p^2 of .466 for a sample size of 10. With the assumption of an effect roughly half that size ($\eta_p^2 = .233$), power analyses suggested that twelve participants were necessary to find significant differences in search slope with 95% power. A total of 14 Vanderbilt University students (11 female, 12 right-handed, median age = 19) participated for class credit. All participants were treated within APA ethical guidelines approved by internal review board.

Stimuli & Apparatus

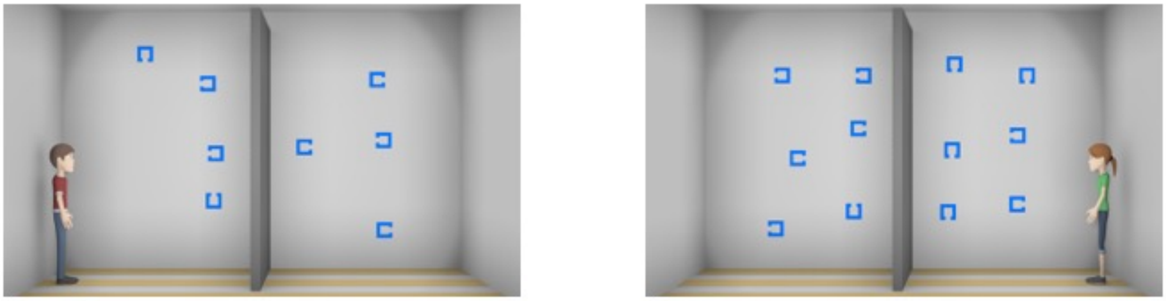
Stimuli were designed using the open source 3D design software, Blender (v2.66; The Blender Foundation, www.blender.org) with open-source images (VMComix, 2011). All measurements below were calculated relative to a 19.5 LCD monitor (actual display: 42cm x 24cm, 1600x900 px resolution, 60 Hz) at a viewing distance of approximately 70 cm. For reproducibility, units are reported in both pixels and approximate visual angle.

All stimuli consisted of a number of search items, a wall and a male or female avatar (matched to participant gender, to avoid potential confounds of cross-gender perspective-taking discussed in Klein & Hodges, 2001). Avatars (occupying an approximate volume of 80 px X 345 px; 1.729° X 7.458°) were located on the left or right sides of the screen, with 185 px (3.999°) from the edge of the display to the front of the agent. A dark grey wall (70 px X 700 px; 1.513° X 15.131°) appeared in the direct center of the screen (centered at 640 pixels from either display edge) in half of trails and on the far side of the screen opposing the avatar (with 250 pixels from the edge of the display to the front of the wall) on the other half of trials. The centered wall evenly divided all items shown, while the distal wall did not divide any targets.

Search arrays consisted of 4, 8 or 12 blue items, similar to Landolt C’s. The nontargets were 40 px^2 (.865°) outlined squares (10 px, .216°line thickness) with a 12 px (0.259°) gap on the left, right or bottom side. Target items were identical to nontarget items, except that the gap appeared on the top of the square. Items were presented in an evenly spaced 4 x 3 grid of possible locations with a minimum of 48 px (1.038°) between each item.

Visible + Far

Visible + Near



Occluded + Near

Occluded + Far

Figure 3.1. Representative search displays used in Experiment 1. Participants responded whether they or the avatar could see an upward facing C shape. Targets either appeared in the cluster Near or Far from the avatar. Placement of a wall determined whether the avatar’s perspective of the scene was Visible or Occluded. Note that avatars did not have visual access to targets in Occluded + Far trials.

Stimuli were completely counterbalanced across five dimensions. (1) Displays had set sizes of 4, 8, or 12 items. (2) An avatar either appeared on the left or right side of the display. (3) A wall either appeared in the center of the screen (*Occluded* displays) or was to the distal side of the avatar (*Visible* displays). (4) Targets appeared in one of the six item locations *Near* the avatar or *Far* from the avatar. Far targets in Occluded trials were blocked by the wall and thus not visible to the avatar, facilitating direct comparison to Far targets in Visible trials, which were located in the same regions of space but critically inaccessible to avatars. Eight displays were made for each unique combination of these first four dimensions, generating 192 displays. (5) A corresponding target absent display was made from each of the target present displays by rotating the target item. Thus all target absent trials were directly comparable to all target present trails. This led to a total of 384 displays.

Procedure

Upon giving consent, participants completed the entire experiment using a graphic interface designed in Python using the PsychoPy toolkit (Pierce, 2007). Participants read instructions before completing a tutorial. The tutorial consisted of five practice trials that demonstrated different trial types outlined below. Participants repeated the tutorial until they reached 100% accuracy. They then verbally confirmed their understanding of the instructions with the experimenter before proceeding.

Each trial consisted of a fixation cross in the center of the screen for 500 ms, followed by a perspective cue for 750 ms and the stimulus display. Perspective cues either primed the participant's perspective (*Self* trials: "YOU") or the avatar's perspective (*Other* trials: "HIM" or "HER", contingent on participant gender). Participants responded whether the avatar could see the target in Other trials, and responded whether targets were present anywhere in the display during Self trials. Stimuli were displayed until participants responded. Participants viewed each stimulus twice, once with a Self cue and once with an Other cue, for a total of 768 trials. Trials were randomly presented over six blocks. Participants were given a self-paced break between blocks.

The fixation cross provided feedback after every trial, displaying green following an accurate response and red following an inaccurate response. Participants were instructed to respond as fast as possible while still maintaining accuracy. Participant accuracy was displayed at the end of each block. Participants were instructed to aim for 95% accuracy, and were encouraged to speed up if accuracy was above 97% and slow down if accuracy fell below 93%. The entire experiment took less than 45 minutes.

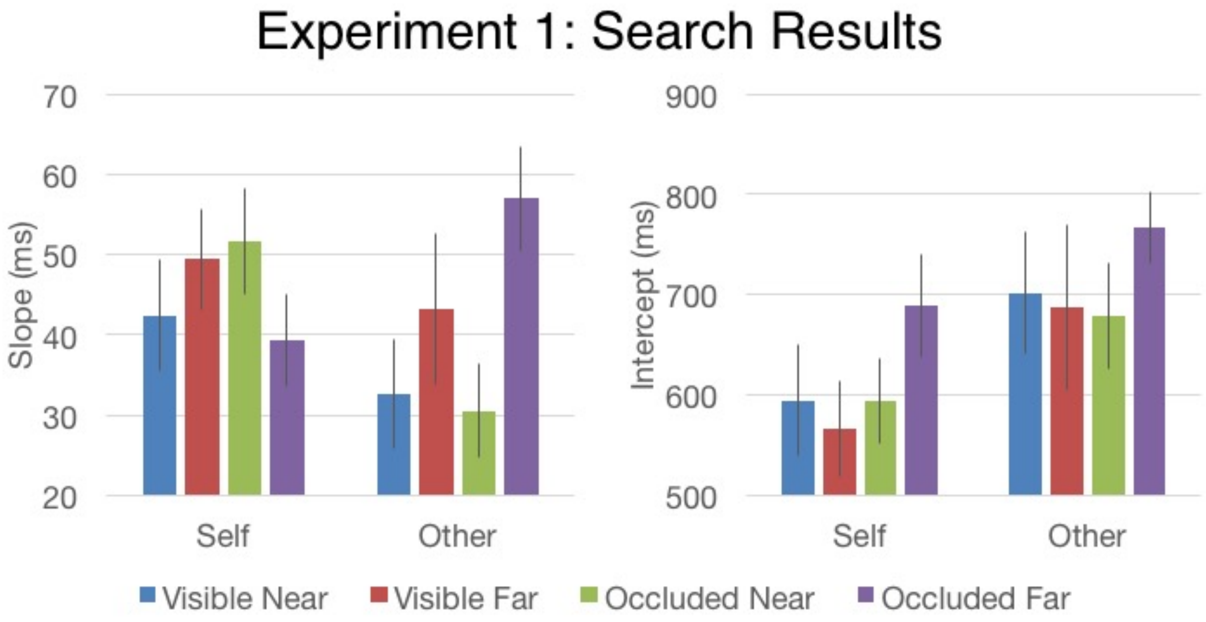


Figure 3.2. Average search slope and intercept in Experiment 1. Results calculated from target-present, accurate responses. Error bars here and throughout the paper denote standard error of the mean for repeated measures designs (O'Brian & Cousineau, 2014).

		Visible		Occluded	
		Near	Far	Near	Far
Self	Intercept (ms)	594 (52.9)	567 (45.5)	594 (40.7)	689 (48.3)
	Slope (ms/item)	42.4 (6.6)	49.4 (5.9)	51.6 (6.3)	39.3 (5.5)
	Error (%)	10.6 (4.3)	14.9 (5.5)	9.7 (4.5)	11.3 (4.9)
Other	Intercept (ms)	702 (57.8)	687 (79.3)	680 (50.5)	767 (33.8)
	Slope (ms/item)	32.6 (6.5)	43.2 (8.9)	30.5 (5.5)	57.0 (6.2)
	Error (%)	10.9 (4.7)	12.8 (5.9)	8.2 (4.6)	9.8 (5.4)

Table 3.1. Summary search performance in Experiment 1. Values are for target-present, accurate trials. Values in parentheses indicate one standard error of the mean.

Results

Table 3.1 summarizes search performance. The present experiment was centrally concerned with Self trials, as these reflected differences generated even when attempting to ignore the avatar, and also because Other trials required fundamentally different responses. For instance, the optimal strategy for Other-Occluded trials only required participants to scan half the screen shared by the avatar before making a decision. However, participants could also search for all targets before making a response. In both cases, participants in Other-Occluded-Far trials would have to detect a target and then form a negative response, effectively detecting a target but saying the avatar could not see it. These trials differed from Self trials, where participants simply had to search for a target and respond. If participants are sensitive to other perspectives, then these Self trials should be affected. Nonetheless, analyses for Other trials are included for their interest to future research. Outside of planned contrasts, all comparisons for the remainder of the paper were corrected for false discovery rate using the Benjamini-Hochberg-Yekutieli procedure for repeated measures (Benjamini & Yekutieli, 2001).

Accuracy

Participants demonstrated relatively few false-alarms to target-absent trials (M_{self} = 98.0% correct reject, M_{other} = 98.7% correct reject). Participants were also relatively accurate for target present trials in both conditions (M_{self} = 88.4% correct, M_{other} = 89.6% correct). Although participants demonstrated more errors than typical visual search experiments, errors were expected given the additional complexity of the perspective-taking task. A 2 (Perspective: Self vs Other) x 2 (Visibility: Visible vs Occluded) x 2 (Proximity: Near vs Far) ANOVA revealed a significant main effect of Visibility ($F_{1,13} = 5.712$, $p = .0327$), a marginal effect of Proximity ($F_{1,13} = 3.404$, $p = .088$) and no effect of Perspective ($F_{1,13} = .867$, $p = .329$). However, post-hoc comparisons revealed no significant differences between conditions (*p*'s > .44). Only accurate trials were analyzed further.

Response Times

A 3 (Set Size: 4,8, or 12) x 2 (Target: Present vs. Absent) ANCOVA revealed a significant linear effect of set size on response time ($F_{1,13} = 103.9$, $p < .001$). Target Absent trials were significantly slower than Target Present trials at a ratio of 1.45:1 ($F_{1,13} = 52.32$, $p < .001$).

There was also a significant Set Size*Target interaction, revealing a significant increase in search slope between Target Present and Target Absent trials at a 2.42:1 ratio ($F_{1,13} = 52.32$, $p < .001$). Given the presence of a significant slope effect, the remainder of analyses only used response times from target-present trials.

Search Slopes. A 2 (Perspective: Self vs. Other) x 2 (Visibility: Visible vs. Occluded) x 2 (Target Proximity: Near avatar vs. Far from avatar) ANOVA revealed no significant main effects of Perspective, Visibility or Perspective (all $F_{1,13} < 1.613$, $p > .226$). There was a marginal Perspective*Proximity interaction ($F_{1,13} = 3.980$, $p = .067$) and a significant Perspective*Visibility*Proximity interaction ($F_{1,13} = 7.819$, $p = .015$). However, paired comparisons identified no significant differences (*p*'s > .29).

Intercepts. A 2 (Perspective: Self vs. Other) x 2 (Visibility: Visible vs. Occluded) x 2 (Target Proximity: Near avatar vs. Far from avatar) ANOVA revealed no main effects of Perspective ($F_{1,13} = 2.662$, $p = .127$), Proximity ($F_{1,13} = 0.591$, $p = .456$) or Visibility ($F_{1,13} = 2.066$, $p = .174$). There was a marginal Visibility*Proximity interaction ($F_{1,13} = 3.792$, $p = .0734$). However, post-hoc tests revealed no significant differences (*p*'s > .20)

Planned Contrast 1: Global Perspective-Taking

The first planned contrast tested for the possibility that participants efficiently calculated the avatar's visual access over large portions of the visual field at one time. Global perspective-taking would result in increased search intercepts during Occluded trials. However, participants demonstrated no difference in intercepts between Self-Occluded and Self-Visible trials (mean difference: 58.8 ms; $t_{13} = 1.143$, $p = .274$). These results suggest that participants do not automatically process the avatar's perspective prior to search.

Planned Contrast 2: Serial Perspective-Taking

The second planned contrast tested for differences in slope as a function of the agent's visibility of a scene. If viewers represent each searched item from the other agent's perspective, we would predict greater search slopes for Occluded trials. However, participants demonstrated no difference in search slopes between Self-Occluded and Self-Visible trials (mean difference: 0.19 ms/item; $t_{13} = 0.028$, $p = .978$). The results suggest that participants do no process each item in a display from multiple perspectives.

Planned Contrast 3: Search Initiation Bias

The presence of an agent may shape participant behavior regardless of that agent's perspective. In practice, a search initiation bias would require participants to search through fewer items on average if the target is near the avatar than if it were far from the avatar, resulting in increased slopes as a function of target distance from the avatar. The third planned comparison thus tested whether search slopes increased when targets were Far versus Near the avatar. Participants demonstrated no overall differences between Near and Far slopes in Self trials (mean difference: 2.88 ms; $t_{13} = .659$, $p = .52$). There was no difference between Self-Visible-Near and Self-Visible-Far slopes (mean difference: 7.05 ms; $t_{13} = .856$, $p = .407$). Surprisingly, participant slopes were marginally *slower* in Self-Occluded-Near trials than in Self-Occluded-Far trials (mean difference: 12.26 ms; $t_{13} = 2.110$, $p = .055$). These results are consistent with participants preferentially searching the region inaccessible to the avatar.

Planned Contrast 4: Response Inhibition

Another's perspective may not influence search but nonetheless generate processing costs if individuals calculate perspective after search termination, as they formulate a response. This would demonstrate a cost of decision-making rather than visual attention. The final planned comparison therefore tested whether search intercepts increased in Self-Occluded-Far trials versus Self-Visible-Far trials. As displays were identical in every way between displays save for the location of the obstructing wall, differences between these trials constitute interference incurred after selection of an occluded target. Although participants trended towards faster intercepts in Self-Visible-Far trials than in Self-Occluded-Far trials, this difference was not significant (mean difference: 122.07 ms; $t_{13} = 1.339$, $p = .102$).

Discussion

Participants demonstrated no differences in accuracy, search slopes or search intercepts when searching a scene with or without a conflicting perspective. These results run counter to theories of perspective taking that would predict either rapid global attentional capture of another's perspective or a serial representation of each item's visibility by the agent. However, when targets were both far from the avatar and occluded, participants trended toward shallower search slopes. These results are consistent with a surprising inversion of search initiation bias: participants initiated search where the avatar *could not see*.

Search slopes, and therefore serial deployment of visual attention, did not increase for Self trials when the avatar had an Occluded perspective. If anything, participants showed trends of searching *faster* on a per-item basis when the avatar could not see the target. It is possible that participants registered the presence of a differing perspective and began their search in regions that the avatar could not access. This complements research on cooperative visual search (Brennan et al., 2008), though such behavior has never been recorded in spontaneous search behavior.

Despite the large number of trials, the number of factors involved in the analysis greatly reduced statistical power. Furthermore, it was possible that any differences between stimuli were diminished by the relative ease of the search task, and that a more difficult search task would reveal stronger effects of cognitive load from perspective-taking. I therefore ran a second experiment with more participants to explore whether trends for search preference in regions unseen by a social agent increased with more difficult stimuli.

Experiment 2

Experiment 2 replicated the design Experiment 1 using difficult search stimuli. If participants automatically calculate the presence of a differing perspective before or after search, then we should again see increased search intercepts in Self-Occluded-Far trials. Furthermore, participants in these trials should show reduced search slopes if they initiate search where the agent cannot visually access.

Method

Participants

Power analyses conducted using effects sizes for the critical paired comparison of Self-Occluded-Near vs. Self-Occluded-Far slopes (d_z = .58) suggested that twenty-five participants were necessary to replicate significant differences in search slope with 80% power. A total of 25

Vanderbilt University students (20 female, 22 right-handed, median age = 19) participated for class credit.

Stimuli & Procedure

Stimuli were identical in design and counterbalancing to Experiment 1 with one difference. To increase difficulty of the search, the diagnostic gap that identified items as targets or distractors was reduced from 12 px (0.259°) to 5 px (.108°). The size and spacing of items were otherwise unchanged.

Results

Accuracy

Summary statistics are documented in Table 3.2, and visualized in Figure 3.3. Participants demonstrated relatively fewer false alarms (M_{self} = 99.5% correct reject, M_{other} = 99.5% correct reject and greater target detection than in Experiment 1 (M_{self} = 92.9% correct, M_{other} = 93.2% correct). A 2 (Perspective: Self vs Other) x 2 (Visibility: Visible vs Occluded) x 2 (Proximity: Near vs Far) ANOVA revealed significant main effects of Visibility ($F_{1,24} = 19.36$, $p < .001$) and Proximity ($F_{1,24} = 4.971$, $p = .035$), but no effect of Perspective ($F_{1,24} = .427$, $p = .520$). There was a Visibility*Proximity interaction effect ($F_{1,24} = 8.216$, $p = .008$), but no interaction of Perspective*Proximity ($F_{1,24} = 2.914$, $p = .101$) or Perspective*Visibility ($F_{1,24} = 0.614$, $p = .441$), nor a three-way interaction ($F_{1,24} = 2.891$, $p = .102$). Post-hoc comparisons revealed that participants were significantly worse at Other-Occluded-Far trials than any Other trial (lowest $t_{24} = 2.825$, p's < .024), though again these trials required fundamentally different responses. All Self perspective pairs were non-significant (highest $t_{24} = 2.547$, $p > .13$).

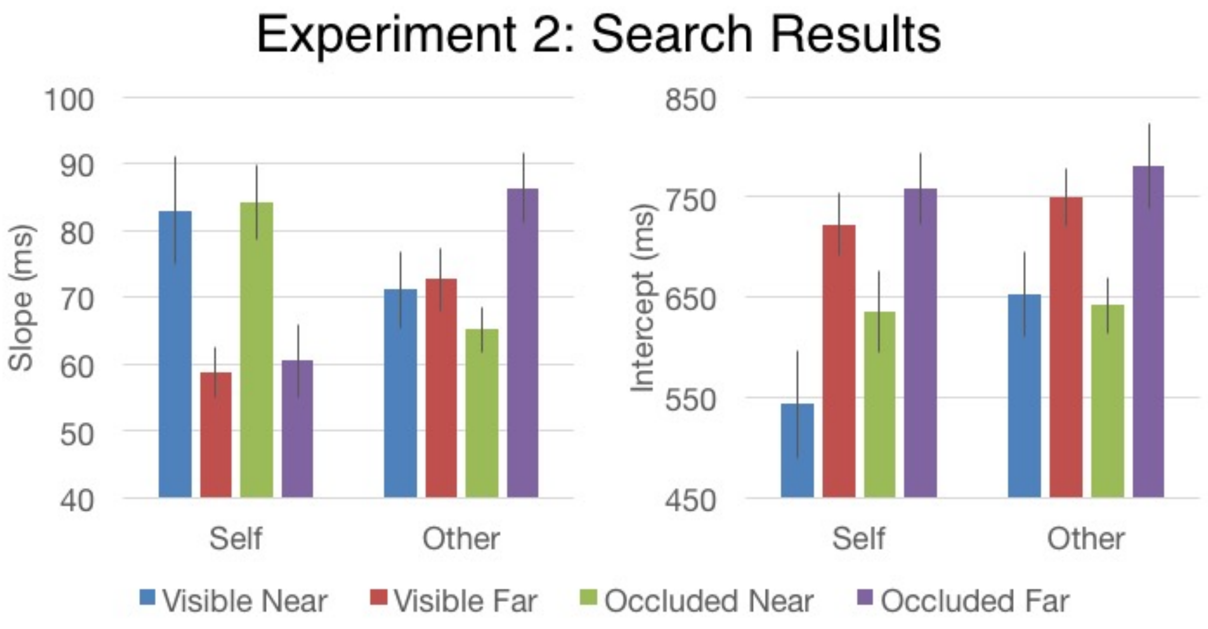


Figure 3.3. Average search slope and intercept in Experiment 2. Results calculated from accurate, target-present trials. Error bars denote standard error of the mean.

		Visible		Occluded	
		Near	Far	Near	Far
Self	Intercept (ms)	543 (51.9)	723 (29.9)	635 (38.7)	759 (33.4)
	Slope (ms/item)	82.9 (7.7)	58.8 (3.5)	84.2 (5.3)	60.5 (5.2)
	Error (%)	5.8 (0.7)	6.0 (0.8)	8.0 (0.9)	8.6 (0.7)
Other	Intercept (ms)	653 (40.1)	750 (27.6)	642 (26.4)	781 (40.2)
	Slope (ms/item)	71.1 (5.4)	72.8 (4.5)	65.2 (3.3)	86.4 (5.0)
	Error (%)	6.1 (0.7)	5.9 (0.4)	5.2 (0.5)	9.7 (0.9)

Table 3.2. Summary search performance in Experiment 2. Values illustrate accurate, target-present trials. Values in parentheses indicate standard error of the mean.

Response Times

A 3 (Set Size: 4, 8, or 12) x 2 (Target: Present vs. Absent) ANCOVA across all conditions revealed a significant linear effect of set size on response time ($F_{1,24} = 334.4$, $p < .001$). Target Absent trials were significantly slower than Target Present trials at a ratio of 1.45:1 ($F_{1,24} = 157.8$, $p < .001$), identical to Experiment 1. There was also a significant Set Size*Target interaction, revealing a significant increase in search slope between Target Present and Target Absent trials at a 2.07:1 ratio ($F_{1,24} = 158.4$ $p < .001$), again quite similar to Experiment 1. Given the presence of a significant slope effect, the remainder of analyses only used response times from target-present trials.

Search Slopes. A 2 (Perspective: Self vs. Other) x 2 (Visibility: Visible vs. Occluded) x 2 (Target Proximity: Near avatar vs. Far from avatar) ANOVA revealed no significant main effects of Perspective ($F_{1,24}$ =.435, $p = .516$), Visibility ($F_{1,24}$ =.363, $p = .553$) or Proximity ($F_{1,24}$ = 2.786, $p = .108$). There was no Perspective*Visibility interaction ($F_{1,24}$ = .086, $p = .772$) or Visibility*Proximity interaction ($F_{1,24}$ =1.951 $p = .175$). There was a significant Perspective*Proximity interaction ($F_{1,24}$ =28.572, $p < .001$). Post-hoc comparisons revealed that Self-Far slopes were significantly faster than Self-Near slopes (t_{24} = 4.642, $p < .001$) and that Other-Far slopes were significantly slower than Other-Near slopes (t_{24} = 2.682, $p = .046$). There was no three-way interaction ($F_{1,24}$ = 1.066, $p = .312$).

Intercepts. A 2 (Perspective: Self vs. Other) x 2 (Visibility: Visible vs. Occluded) x 2 (Target Proximity: Near avatar vs. Far from avatar) ANOVA revealed no effect of Perspective ($F_{1,24}$ = 2.979, $p = .097$) or Visibility ($F_{1,24}$ = 1.236, $p = .277$), but there was significant effect of Proximity ($F_{1,24}$ = 20.150, $p < .001$) with Near trials having overall faster intercepts. There were no interactions (all $F_{1,24}$ <1, $p > .322$).

Planned Contrast 1: Global Perspective-Taking

The first planned contrast tested for global processing of the agent's perspective, which would result in increased search intercepts when perspectives conflicted in Occluded trials. Participants demonstrated no difference between Self-Occluded intercepts and Self-Visible intercepts (mean difference: 43.4 ms; t_{24} = 0.757, $p = .456$).

Planned Contrast 2: Serial Perspective-Taking

The second planned contrast tested an effect of an agent's visual access on serial search, as identified by increasing slope as a function of agent perceptual access. Participants demonstrated no difference in search slopes between Self-Occluded and Self-Visible trials (mean difference: 10.53 ms/item; t_{24} =1.427, $p = .167$).

Planned Contrast 3: Search Initiation Bias

The third planned contrast tested for an effect of agent location on search initiation bias, as identified by increasing slope as a function of target distance from the avatar. Participants had significantly faster search slopes in Far trials (mean: 59.63 ms/item) than Near trials (83.56 ms/item; $t_{24} = 5.153$, $p < .001$). This was true in both Visible ($t_{24} = 2.603$, $p = 008$) and Occluded trials ($t_{24} = 3.074$, $p = 003$).

Planned Contrast 4: Response Inhibition

The final planed contrast tested whether participants incurred processing costs after selection of a target that was occluded from the avatar. However, intercepts did not differ between Self-Visible Far and Self-Occluded-Far trials (mean difference: 36.19 ms; $t_{24} = 0.029$, $p = .488$).

Exploratory Analysis

Previous designs measuring rapid perspective-taking were carefully controlled to present items within the range of subitization, between 0-4 targets, suggested to be the general capacity of efficient visual attentional processing (Trick & Pylyshyn, 1994). As discussed in the introduction, these previous experiments (including Chapter 2 of this dissertation) found significant processing delays when these small sets of objects were obstructed from another agent. Although Experiment 2 found no evidence in support of a global perspective-taking hypothesis, wherein participants rapidly calculate an agent's view of the entire scene as indicated by search intercepts, it is possible that global perspective-taking might arise in displays within the range of efficient attentional processing. Furthermore, the slope effects of both experiments suggest that individuals initiate search in regions far from the avatar. However, this slope increase might indicate a strategy implemented by participants when set sizes are too large to permit global perspective-taking. In order words, participants may search from the avatar's perspective when the number of objects in the display can be captured by efficient attentional processing, and adopt a strategy to look where the avatar cannot access when the number of objects exceeds attentional capacity.

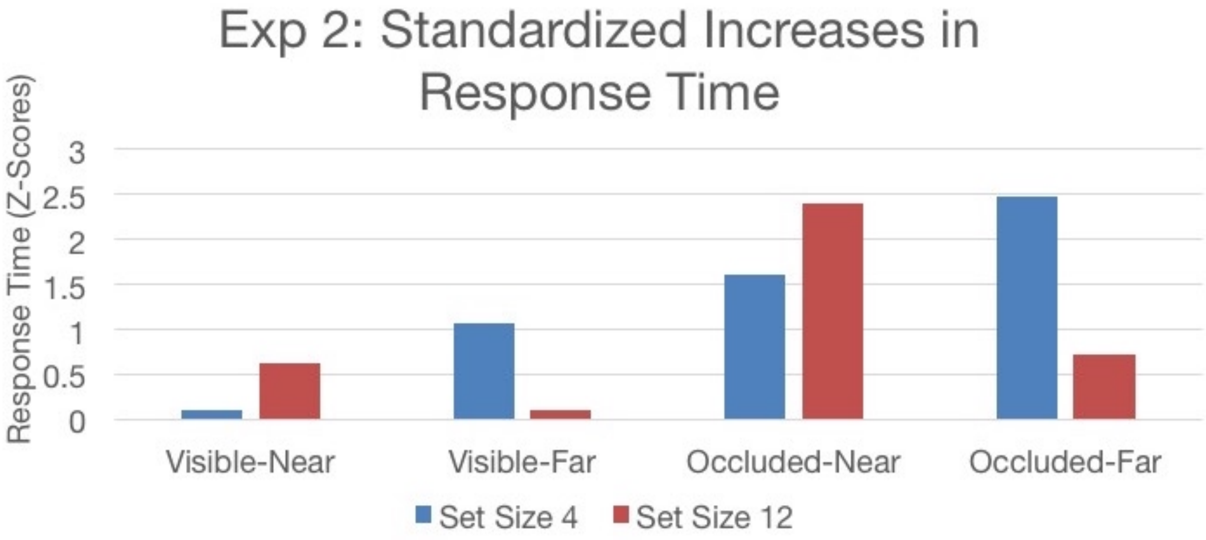


Figure 3.4. Standard scores of response times in Self trials for Experiment 2. Individuals responded faster to targets near the avatar in small set sizes and faster to targets far from the avatar in larger set sizes. Z-scores were calculated from average scores from each conditions within each set divided by standard deviation. The minimum score was then added to each z-score for visualization purposes. Statistical analyses were conducted on raw data.

Figure 3.4 visualizes standardized increases in response time by condition and set size. A 2 (Visibility: Visible vs. Occluded) x 2 (Target Proximity: Near avatar vs. Far from avatar) ANOVA for Self trials with a set size of 4 revealed that participants were significantly slower to respond to Occluded trials (mean: 990 ms) than Visible trials (mean: 908 ms; $F_{1,24} = 10.740$, $p = .003$), and were significantly slower in Far trials (mean: 975 ms) than Near trials (mean: 923 ms; $F_{1,24} = 5.569$, $p = .027$), with no interaction ($F_{1,24} = 0.037$, $p = .848$). Meanwhile, the same analysis for trials with a set size of 12 also revealed significantly slower response times for Occluded trials (mean: 1593 ms) than for Visible trials (mean: 1424 ms; $F_{1,24} = 12.322$, $p = .002$), but found significantly *faster* response times for Far trials (mean: 1431 ms) than for Near trials (mean: 1585 ms; $F_{1,24} = 17.919$, $p < .001$), and a significant Visibility*Proximity interaction ($F_{1,24} = 4.236$, $p = .05$). Paired comparisons controlling for false discovery rate using the Benjamini-Hochberg-Yekutieli procedure revealed significantly slower responding in Occluded-Near trials (mean: 1710 ms) than in Visible-Near (mean:1461 ms; $t_{24} = 3.122$, $p = .011$), Visible-Far (mean: 1387 ms; $t_{24} = 5.888$, $p < .001$) or Occluded-Far trials (mean: 1474 ms; $t_{24} = 4.259$, $p < .001$). No other comparisons were significant ($p > .312$).

These results, although exploratory, support an automatic detection of another's conflicting perspective revealed through increases in overall response time in occluded trials in both set sizes. It is methodologically interesting to find overall increases in response times in both Set Size 4 and Set Size 12 in Occluded scenes despite the lack of a significant intercept effect in the general results. The intercept effect was potentially offset by marginal variations in slope between trials. In this case, the increased slope in Occluded-Far trials greatly influences projection of search intercept, masking differences between conditions. Analysis of average RT across trials is unusual for visual search experiments, but nonetheless reveals processing costs when the avatar's perspective is occluded.

This exploratory analysis indicates that Experiments 1 and 2 may be capturing two distinct processes. The first is a potentially automatic guidance of attention to the agent, leading to slower response times in Occluded trials and faster response times when the target is near the avatar in small display sizes. However, this initial bias gives way to a second, potentially learned strategy to search areas far from the avatar when the number of objects exceeds the capacity of efficient processing.

Discussion

Experiment 2 replicated the main findings from Experiment 1. Participants demonstrated a significant search initiation bias, with faster search slopes in Self-Far trials than in Self -Near trials. This expanded replication of Experiment 1 found this effect for both Occluded and Visible trials. These results suggest that participants registered the location of the agent and initiated their search where the agent could not see.

Exploratory analyses examining response times for small and large set sizes further inform a potential source of search initiation bias. Increased average response times during Occluded trials suggest an overall processing cost associated with viewing a scene with an inconsistent perspective. It is possible that individuals deploy attention to the locations of joint perceptual access, but have a heuristic to search areas of privileged access when search is difficult. Whether this secondary search is a strategy learned through the design of this experiment or a more general heuristic applicable to other social scenes is a matter for future experiments specifically tailored to such questions.

General Discussion

Previous research suggests humans possess a system for rapidly guiding attention to the perspectives of others, but no research has demonstrated whether such a system actually alters deployment of attention. In two experiments, I tested how the presence of a social agent influenced visual search of a simple scene. These experiments indicated that individuals do in fact rapidly incorporate other people's perspectives into their search strategies, however they do so in a surprising manner: participants demonstrate a search initiation bias, preferentially looking at regions of space that are distant and occluded from the agent. These results support an emerging literature on automatic processes in social cognition, and are the first to illustrate a role of early visual attention in such processing.

The tendency of individuals to first search regions inaccessible to another agent aligns with previous research on collaboration in visual search tasks. Researchers found that pairs of individuals were most efficient when they could see each other's eye movements (Brennan, Chen, Dickinson, Neider & Zelinski, 2008). It is possible that the rapid calculation of other's perspectives as in Chapter 2 of this dissertation alerts participants to the regions of space accessible by another agent, and that visual search is biased for collaboration or competition.

There are several theories of automaticity in mental state reasoning, as detailed in the introduction. These theories vary in their assertion of domain-specificity and the core mechanisms involved, however all hypothesize a mechanism by which individuals detect the mental states of another agent. The current experiment sheds some light onto future models. First, the core finding that individuals alter their behavior in response to detecting another's perspective gives credit to the assumption that some form of calculation occurs rapidly in the perceptual stream. Second, differences in search slopes suggest that participants used this information to guide search, but does not suggest an overall attentional cost. Rather, these results, though admittedly preliminary, suggest that individuals detect the presence of an agent and heuristically search the space opposite the agent. This is consistent with the results of Chapter 2, which found that increasing number and magnitude of differing perspectives does not alter responding in the dot-perspective task. Rather than directing attention to those objects that another person can see, it may be that a rapid system detects the presence of other agents and guides search to privileged information.

Notably, these results neither confirm nor deny the capture of attention by social stimuli. In the present design, we would expect participants to demonstrate overall intercept delays if recognition of the avatar's visual access delayed responding overall but did not affect deployment of attention to each individual item in the scene (global perspective-taking). No such differences were found. Conversely, we would predict increased slopes if identification of items required evaluation of the agent's perspective for each item (serial perspective-taking). This difference was also not found. Crucially, exploratory analysis indicates that search initiation bias need not imply capture of attention by social agents. Rather, participants may have adopted a strategy to search the area opposing the avatar, either through training throughout the study or through a more general heuristic.

Several additional experiments are necessary before strong conclusions can be made. First, the exploratory findings of Experiment 2 suggest that the current design may capture two processes: an early-acting evaluation of another agent's visual access and a late-acting strategy to search items far from the avatar. One way of diverging these behaviors would be a stimulus onset asynchrony (SOA) testing the participant's ability to detect a target in a brief window of time (e.g., 120ms, 250ms, 500 ms or 1000 ms presentations). If attention is captured by social stimuli, detection of targets should increase with proximity to the avatar, even at low SOA's.

Although the current design necessitated cues to take a perspective before stimulus onset, it is possible that these cues also gave participants a 750 ms window to form a search strategy. Furthermore, it is possible that the act of preferentially searching the agent's space when taking the Other's perspective led participants to take the opposite strategy when taking an egocentric perspective. It would therefore be worthwhile to replicate these effects without taking the Other's perspective. Such an experiment could have periodic attention checks to ensure that participants continued to note the location of the avatar even without directly taking its perspective. Given that previous accounts of rapid perspective-taking have revealed effects during purely egocentric report (Samson et al, 2011), a replication of Experiment 2 should reveal identical effects if participants did not adopt a unique Self search strategy that opposed Other search strategies.

Visual search experiments were designed to assess how attention guides eye movements in a time when the computational power required to monitor and analyze eye movements was not readily available. In the current era, however, more direct methods can test where and when individuals look at regions of space. In a pilot study, I ran 16 participants through a modified version of Experiment 2 while monitoring their eye movements. Participants overall performed quite poorly on this task, perhaps due to accidental decreases in visual angle of the target stimuli as a consequence of fixed head positions in the eye tracking apparatus. Search slopes were twice of those in the first two experiments. However, participants in these experiments demonstrated the *opposite* finding of Experiment 2: Participants robustly searched the region near the avatar first. This trend appeared both in participant search slopes and in their first saccades and fixations. This data complicates interpretation of the current study, and necessitates replication to completely understand these behaviors.

REFERENCES

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106-111.

Apperly, I. A, & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–70.

Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 333–356.

Baird, J. A., & Baldwin, D. A. (2001). Making sense of human behavior: Action parsing and intentional inference. In Malle, B. F., Moses, L. J. & Baldwin, D. A. (Eds), *Intentions and Intentionality: Foundations of Social Cognition*, 193-206.

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14–29.

Baker, L.J., Saylor, M.M., & Levin, D.T. (2016). Default visuospatial perspective taking in complex layouts. *Journal of Experimental Psychology: Human Performance and Perception*, *42*(4), 508-516

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708-717.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–407.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 1165-1188.

Blender Foundation (2013). Blender. GNU General Public License v3.0. Retrieved from www.blender.org.

Boltz, M. (1992). Temporal accent structure and the remembering of filmed narratives. Journal of Experimental Psychology: *Human Perception and Performance*, *18*(1), 90–105.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, *62*(8), 1572–1595.

Brennan, S. E., & Clark, H. H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465-1477.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*(1), 1-16.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*(1), 55-72.

Cohen, C. E., & Ebbesen, E. B. (1979). Observational goals and schema activation: A theoretical framework for behavior perception. *Journal of Experimental Social Psychology*, *15*(4), 305-329.

Cohn, N. (2013). Visual narrative structure. *Cognitive Science*, *37*(3), 413–52.

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of agents: the processing of semantic roles in visual narrative. *Cognitive Psychology*, *67*(3), 73–97.

Cutting, J. E. (1981). Six tenets for event perception. *Cognition*, *10*, 71–78.

Cutting, J. E., DeLong, J. E., & Nothelfer, C. E. (2010). Attention and the evolution of Hollywood film. *Psychological Science*, *21*(3), 432-439.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*(6), 693–700.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433.

Ebbesen, E. B. (1980). Cognitive processes in understanding ongoing behavior. In R. Hastie (Ed.), *Person memory: The cognitive basis of social perception*, 179-225. Hillsdale, NJ: Erlbaum.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*(1), 269-297.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults : Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*, 760–768.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–86.

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, *22*(2), 243–52.

Ferguson, M. J., & Bargh, J. A. (2004). How social perception can automatically influence behavior. *Trends in Cognitive Sciences*, *8*(1), 33-39.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99.

Fodor, J. A. (1992). A theory of the child's theory of mind. Cognition, 44(3), 283-296.

Forster, S., & Lavie, N. (2009). *Harnessing the wandering mind: The role of perceptual load. Cognition*, *111*(3), 345–355.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319–29.

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, *1*, 292–304.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in Cognitive Science*, *7*(7), 287-292.

Gibson, J.J. (1966). The senses considered as perceptual systems. Houghton Mifflin Co.

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137-176.

Hanson, C., & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General*, *118*(2), 136–47.

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: building event schemas. *Memory & Cognition*, *34*(6), 1221–35.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243–259.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, *10*(5), 438–443.

Huff, M., & Schwan, S. (2012). Do not cross the line: Heuristic spatial updating in dynamic scenes. *Psychonomic Bulletin & Review*, *19*(6), 1065-1072.

Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, *20*(7), 848–864.

Hymel, A. M. Levin, D. T., & Baker, L. J. (2015). Default processing of event sequences. *Journal of Experimental Psychology: Human Perception & Performance*, *42*(2), 235-246.

James, W. (1904). *A World of Pure Experience. Journal of Philosophy, Psychology, and Scientific Methods*, *1*, 533-543, 561-570.

Johansson, G., von Hofsten, C., & Jansson, G. (1980). Event perception. *Annual Review of Psychology*, *31*, 27–63.

Kives, C.R., Ware, S. G., & Baker, L.J. (2015). Evaluating the pairwise event salience hypothesis in Indexter. *Proceedings of the Eleventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-15).*

Klein, K. J. K., & Hodges, S. D. (1993). Gender differences, motivation and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27*(6), 720–730.

Kleiner M, Brainard D, Pelli D, 2007, “What’s new in Psychtoolbox-3?” *Perception 36 ECVP Abstract Supplement.*

Koffka, K. (1922). Perception: and introduction to the Gestalt-theorie. *Psychological Bulletin*, *19*, 531-585.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–4.

Kurby, C. A. & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–9.

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition*, *40*(5), 812–26.

Lassiter, G. D., & Slaw, R. D. (1991). The unitization and memory of events. *Journal of Experimental Psychology: General*, *120*(1), 80-82.

Lassiter, G. D., Geers, a. L., & Apple, K. J. (2002). Communication set and the perception of ongoing behavior. *Personality and Social Psychology Bulletin*, *28*(2), 158–171.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*(3), 143–148.

Lavie, N., Hirst, A., De Fockert, J.W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*, 339–354

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience*, *31*(8), 2906–15.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences*, *8*(12), 528–33.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, *4*(4), 501–506.

Levin, D. T., Takarae, Y., Miner, a G., & Keil, F. C. (2001). Efficient visual search by category: specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics*, *63*(4), 676–97.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556.

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 435–441.

Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychological Review*, *103*(4), 603.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*(8), 601-608.

Mack, A., & Rock, I. (1998). Inattentional Blindness. Cambridge, MA: MIT press.

Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, *35*(8), 1489–517.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, *15*(5), 533–545.

Magliano, J. P., Taylor, H. A, & Kim, H.-J. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory & Cognition*, *33*(8), 1357–67.

Mattan, B. D., Rotshtein, P., & Quinn, K. A. (2016). Empathy and visual perspective-taking performance. *Cognitive Neuroscience*, 1-12.

McCleery, J. P., Surtees, A. D. R., Graham, K. a, Richards, J. E., & Apperly, I. a. (2011). The neural and cognitive time course of theory of mind. *The Journal of Neuroscience*, *31*(36), 12849–54.

Meyerhoff, H. S., Huff, M., Papenmeier, F., Jahn, G., & Schwan, S. (2011). Continuous visual cues trigger automatic spatial target updating in dynamic scenes. *Cognition*, *121*(1), 73-82.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics*, *66*(8), 1268–81.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56-60.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, *28*(3), 292–312.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*(6059), 264-265.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2010). Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, *17*(5), 718-724.

Neisser U., & Becklen R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, *7*, 480-494.

Neisser, U. (1976). Cognition and reality: Principles and implications of cognitive psychology. WH Freeman/Times Books/Henry Holt & Co.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436–450.

Newtson, D., Hairfield, J., Bloomingdale, J., & Cutino, S. (1987). The structure of action and interaction. *Social Cognition*, *5*(3), 191-237.

Newtson, D., Rindner, R., Miller, R., & LaCross, K. (1978). Effects of availability of feature changes on behavior segmentation. *Journal of Experimental Social Psychology*, *14*(4), 379–388.

O'Brien, F., & Cousineau, D. (2014). Representing error bars in within-subject designs in typical software packages. *The Quantitative Methods for Psychology*, *10*, 56-67.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255-258.

Peirce, J.W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8-13.

Piaget, J., & Inhelder, B. (1956) The child's conception of space. London Routledge & Kegan Paul.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509–522.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515-526.

Qureshi, A. W., Apperly, I. A, & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, 117(2), 230–6.

Radvansky, G. A, Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: environmental integration. *Psychonomic Bulletin & Review*, *17*(6), 900–4.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849-860.

Rensink, R. A. (2010). The dynamic representation of scenes. *Visual Cognition*, *7*(1-3), 17–42.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To See or not to See: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–43.

Robertson, S. S., & Suci, G. J. (1980). Event perception by children in the early stages of language production. *Child Development*, *51*, 89-96.

Samson, D., & Apperly, I. A. (2010). There is more to mind reading than having theory of mind concepts: New directions in theory of mind research. *Infant and Child Development*, *19*(5), 443–454.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Scott, S. E. B. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929–37.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*, 241–255.

Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science*, *13*(6), 493-498.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299-309.

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology*, *18*(1), 37–55.

Schwan, S., & Ildirar, S. (2010). Watching film for the first time: How adult viewers interpret perceptual discontinuities in film. *Psychological Science*, *21*(7), 970-976.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes during a real-world interaction. *Psychonomic Bulletin & Review*, *5*(4), 644–649.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20.

Simons, D. J., Chabris, C. F., Schnur, T. T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition*, *11*, 78-97

Smith, T. J., & Santacreu, J.Y.M.P. (2016). Match-action: the role of motion and audio in creating global change blindness in film. *Media Psychology*, 1-32.

Smith, T. J., Lamont, P., & Henderson, J. M. (2012). The penny drops: Change blindness at fixation. *Perception*, *41*(4), 489–492.

Spapé, M. M., & Hommel, B. (2010). Actions travel with their objects: Evidence for dynamic event files. *Psychological Research*, *74*(1), 50–8.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*(1), 125–140.

Staresina, B. P., & Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. *Neuron*, *63*(2), 267–76.

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*(3), 409–15.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580-586.

Surtees, A., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, *83*(2), 452–60.

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43-52.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97-105.

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, *23*(5), 1052–64.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257.

Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review*, *11*(1), 65-70.

Theeuwes, J., & Kooi, F. L. (1994). Parallel search for a conjunction of contrast polarity and shape. *Vision Research*, *34*(22), 3013-3016.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751-754.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(05), 675-691.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80-102.

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*, 86–94.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, *28*, 127–154

Varakin, D. A., Levin, D. T., & Collins, K. M. (2007). Comparison and representation failures both cause real-world change blindness. *Perception*, *36*(5), 737–749.

VMComix. (2011, July 04). Cartoon Character Pack 1. Public domain dedication under Creative Commons license 1.0 Universal. Downloaded August, 2012. Retrieved from http://www.blendswap.com/blends/view/14445.

Wegner, D. M., & Giuliano, T. (1983). Social awareness in story comprehension. *Social Cognition*, *2*(1), 1–17.

Wertheimer M. Untersuchungen zur Lehre von der Gestalt, I: Prinzipielle Bemerkungen.Psychologische Forschung. Translated extract reprinted as “The general theoretical situation”. In W. D. Ellis (Ed.), (1938). *A Source Book of Gestalt Psychology*. London, U. K.: Routledge & Kegan Paul Ltd., 12–16.

Wilson, S. G., Rinck, M., McNamara, T. P., Bower, G. H., & Morrow, D. G. (1993). Mental models and narrative comprehension: Some qualifications. *Journal of Memory and Language*, *32*, 141–154.

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–38.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34.

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.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*(16), 2095-2103.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*(6), 979–1008.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*(2), 201-216.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307–27.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–93.

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, *1076*(1), 150–62.

Zacks, J.M., Kurby, C.A., Eisenberg, M.L., Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*, 4057-4066.

Zeigarnik, B. (1938). On finished and unfinished tasks. *A Source Book of Gestalt Psycholog*y, 1, 1-15.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel–serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(1), 244.

Zwaan, R. A, & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–85.