

**MULTIPLE OBJECT TRACKING SUPPORTS  
THEMATIC ROLE FEATURES FOR LANGUAGE**

Thesis submitted in accordance with the requirements  
of the University of Liverpool for the degree of  
Doctor in Philosophy

**Andrew Allen Jessop**

September 2018



*To Mum and Dad.*



## **ACKNOWLEDGEMENTS**

I would like to express my deepest and sincerest gratitude to everyone that has made this incredible journey possible. I want to offer a special thank you to my mentor and advisor Franklin Chang. At the start of this PhD, you enabled me to fulfil a lifelong dream of visiting Japan as part of a research visit, for which I will be eternally grateful. But most importantly, your endless support and encouragement has taught me to think creatively and courageously, which has helped me to evolve as a scientist. I hope that I can continue to uphold these values in my future work. I would like to express my unreserved thanks to Marco Bertamini and the Visual Perception Lab at the University of Liverpool, for supporting me in the trenches throughout this research and providing invaluable feedback. Thanks also to the Liverpool Language Group who have inspired me to dive deeper into psycholinguistic research, not to mention making my PhD experience very enjoyable. Finally, I would never have made it to the finish line without the support of my family: Maureen, Allen, Laura, Paula, Jonathan, and India. I know you have all been waiting patiently to read this thesis for the last three years. So here it is.

This research was made possible by a grant awarded to Andrew Allen Jessop by the Economic and Social Research Council (ES/J500094/1), as part of a studentship from the Northwest Doctoral Training Centre.



## TABLE OF CONTENTS

<u>RATIONALE FOR USING AN ALTERNATIVE THESIS FORMAT</u>	<u>3</u>
<u>ABSTRACT</u>	<u>5</u>
<u>CHAPTER ONE: LITERATURE REVIEW</u>	<u>7</u>
Language-based Theories of Thematic Roles	7
The Visual Perception of Thematic Roles	12
Thematic Roles in Connectionist Models of Language	16
Multiple Object Tracking	25
The Present Work	31
<u>CHAPTER TWO: TRACKING THEMATIC ROLES IN THE VISUAL SYSTEM</u>	<u>36</u>
Fit Within the Thesis	36
Introduction	37
Experiment One: Tracking Roles Without Eye Movements	43
Method	46
Results	50
Experiment Two: Distraction Task	52
Method	54
Results	55
Experiment Three: Temporarily Invisible Objects Task	57
Method	58
Results	59
Combined Analysis	61
General Discussion	62
Additional Information	66

<u>CHAPTER THREE: THEMATIC ROLE TRACKING IN RELATIVE CLAUSES</u>	<u>67</u>
Fit Within the Thesis	67
Introduction	69
Experiment Four: Subject and Object Relative Clauses	78
Methods	82
Results	87
Experiment Five: Subject and Passive Relative Clauses	89
Method	92
Results	92
Experiment Six: Relative Clause Events in Active Transitives	94
Method	96
Results	98
Experiment Seven: Agent Consistency in Active Transitives	101
Method	104
Results	106
Experiment Eight: Patient Consistency in Active Transitives	108
Method	109
Results	109
Experiment Nine: Patient Consistency in Passive Transitives	111
Method	112
Results	113
General Discussion	115
Additional Information	122
<u>CHAPTER FOUR: DISCUSSION</u>	<u>123</u>
<u>REFERENCES</u>	<u>137</u>

## **RATIONALE FOR USING AN ALTERNATIVE THESIS FORMAT**

This thesis has been prepared following the *alternative paper format*, in accordance with the guidelines provided by the University of Liverpool for including research papers in a doctoral thesis. This alternative format was selected for the purpose of facilitating the publication of this research in scientific journals. Specifically, chapters 2 and 3 represent separate manuscripts and are structured in a manner suitable for submission to a peer-reviewed journal. At the time of writing, chapter 2 has been accepted with revisions by the *Quarterly Journal of Experimental Psychology*, whereas chapter 3 is in preparation to be submitted. For consistency, the formatting of these papers match the common font and style used throughout the thesis. No reference section is provided after each paper, with all citations presented in a single bibliography at the end of the thesis. For continuity, neither the experiment numbers nor figure indices reset between the chapters. Otherwise, the chapters are presented in the same format as the manuscripts that would be submitted for publication, with an additional summary at the outset to explain how the papers fit within the broader narrative of the thesis. This means that each chapter starts with a review of the relevant literature to introduce an informed reader to the topic and ends with a discussion of the implications of the results. Both of the papers were prepared in collaboration with my primary PhD advisor, Dr. Franklin Chang. Together, we discussed the design of the studies and how to interpret the data. I constructed the stimuli, recruited the participants, collected, coded, and analysed the data, and prepared the manuscripts presented in chapters 2 and 3. Dr. Chang provided guidance throughout this process in his capacity as my mentor. As with a traditional thesis, a general introduction describing the background literature is

provided in chapter 1, with a discussion and interpretation of the studies in chapter 4.

## ABSTRACT

A central feature of language is the ability to generalise beyond our previous experience to understand unfamiliar input and produce novel sentences (Hockett & Hockett, 1960). It is suggested that thematic roles support this ability by labelling the participants of an event with functions such as the agent and patient (Jackendoff, 1972). However, the nature of these roles is unclear, as language-based theories (e.g., Dowty, 1991) have often relied upon abstract conceptual features that are difficult to define or are not always available in the input. Yet, connectionist models of language have been unable to demonstrate systematic behaviour without including such role variables (Chang, 2002). This thesis aimed to address this issue by considering whether thematic roles in language are supported by the visual system, which allows viewers to track multiple objects in parallel (Pylyshyn & Storm, 1988) and identify their roles in causal events from their movement features (Scholl & Tremoulet, 2000).

To test this connection, a total of nine experiments were conducted with adult participants using a modified version of the multiple object tracking paradigm. In this task, pairs of white circles engaged in causal push events. The participants were required to track the visually indistinguishable targets from these push events while they moved in random patterns for a short period, before producing a sentence to describe their interaction.

Chapter 1 provides an introduction to these experiments. The function of thematic roles in reaching abstract representations of language is discussed, as well as issues with current theories regarding the nature of these roles. The visual perception of causality and the psychophysical features underpinning

these impressions is presented as a potential solution, followed by a summary of the multiple object tracking literature. Chapter 2 reports the first three experiments of the thesis, which collectively provide evidence that the multiple object tracking system can support agent and patient features, even when the attentional demands of the task are manipulated. These studies found that the participants were able to retain the target identities to produce transitive sentences at test that accurately labelled the agent and patient of the interaction. Chapter 3 presents the results of a further six experiments, which used the same multiple push tracking task to examine the processing asymmetry between subject-extracted and object-extracted relative clauses (Gordon & Lowder, 2012). Here, the participants tracked visually identical targets that played the same role in two pushes or switched roles between the events. The results of these six experiments showed that visual events with consistent role-referent bindings are easier to retain and describe in language than those with no repetition or inconsistent bindings. This suggests that subject-extracted relative clauses may be easier to process since the subject referent plays the same role in both events. It also indicates that non-linguistic processing may contribute to the relative clause asymmetry, as the effect was observed when the studies also varied the sentence structures the participants used to describe the events. Finally, chapter 4 discusses the implications of these findings and whether thematic roles are supported by multiple object tracking in vision.

## CHAPTER ONE: LITERATURE REVIEW

To become a competent language user, speakers need to learn the rules of their language to communicate effectively and be understood by other speakers. However, a core feature of language is that users can go beyond the specific sentences they have heard to express new ideas or describe unique events (Hockett & Hockett, 1960). One aspect of language that may support this ability are thematic roles, which represent the function that different entities serve in the events being described (Fillmore, 1967; Gruber, 1965; Jackendoff, 1972, 1987). For instance, in *the girl pushed the boy*, we know that the girl is causing the pushing action, so we call her the *agent*. We also know that the boy is being pushed, so he is the *patient* of the event. Agents and patients are examples of thematic roles, which help us to understand who did what to whom. In this thesis, I will examine some of the problems with existing theories of thematic roles and propose a new account based on the visual processing of events.

### Language-based Theories of Thematic Roles

In English active transitive sentences, agents and patients are arranged around the verb in a consistent order; specifically, agents always appear before the verb, with the patient always succeeding it. Once speakers have learned this general AGENT-ACTION-PATIENT sequence, they can fill each of these sentence positions with any noun or verb to produce their own active transitive sentences. The abstract nature of this understanding means that they can describe any transitive interactions, from typical and familiar events such as *the cat chased the mouse*, to highly unusual and novel events such as *the thesis ate*

*the student*. These sentence frames could even be used with unfamiliar non-sense words that have not been previously encountered. For instance, participants will accept *the rom mecked the zarg* as a grammatically valid sentence and infer that something called a *rom* is the agent, that it is performing an action called *mecking*, and that something called a *zarg* is the patient (Kako, 2006). Importantly, comprehenders cannot simply rely on the ordering of the words to make this inference, as speakers can convey the same meaning using other sentence structures that arrange thematic roles in different ways. For example, although English speakers typically use active transitive structures to describe events, such as *the girl pushed the boy*, they could alternatively use a passive transitive structure like *the boy was pushed by the girl*, with the agent appearing before the verb and the patient after (i.e., a PATIENT-ACTION-AGENT arrangement). The same events could even be described in other languages with structures that arrange the event participants in different ways; for instance, the national language of Madagascar, Malagasy, typically places patients before agents in the standard word order (Guilfoyle, Hung, & Travis, 1992; Pearson, 1998). Therefore, thematic roles can encode the meaning of the sentence independent of the chosen structure or the word order of a particular language.

Linguistic theories have often regarded thematic roles as a type of language representation, forming a finite list of semantic functions that can be identified by specific features. However, it is difficult to define the features that uniquely identify each thematic role (Dowty, 1991; Fillmore, 1967; McRae, Ferretti, & Amyote, 1997). For example, Fillmore (1967) suggested that John is an agent in *John broke the window*, because John causes the window to break.

On the other hand, in the sentence *the hammer broke the window*, the hammer is the cause of the window breaking, but it is said to be an *instrument* as it does not have its own volition and instead is a means of effecting an action.

Therefore, it is unclear what role should be assigned to the zombie in *a zombie broke the window*; presumably the zombie does not have mental states like volition, but at the same time, it is also not a means to an action. While there are options for dealing with this issue within a traditional thematic role framework, each of them has limitations. For example, one could devise new thematic roles to capture intermediate cases on the boundaries of existing roles, but this would require a large proliferation of roles to handle all atypical cases. Another option would be to use more abstract or higher-level roles like *cause*, with broad enough definitions to cover both agents and instruments. But for a sentence like *John broke the window with the hammer*, you would have two causes – *John* and *the hammer* – and such ambiguity would make it difficult to recognise who did what to whom. Therefore, while the appeal of thematic roles comes in part from their power to characterise a wide range of relationships in language using a small list of functions, there is currently no agreement on the number of roles necessary to achieve this or how they should be defined.

In response to these problems with traditional linguistic theories, Dowty (1991) hypothesised that nouns are mapped into sentence arguments using only two conceptual prototypes: the proto-agent and proto-patient. According to Dowty, these proto-roles consist of general features typical of agents and patients, allowing speakers to assign entities to the subject and object position of the sentence based on how much they resemble these proto-roles, thereby avoiding the need for them to perfectly fit into discrete thematic role categories.

Proto-agent features include being event-independent, sentient, volitional, causally responsible and moving. When deciding on the subject, each entity in the event is checked against these features. For example, in *John broke the window*, John has all of the Proto-agent features, while the window is only event-independent, hence John should be in the subject position of the transitive sentence. In the case of the zombie in *the zombie broke the window*, it is event-independent, causally responsible, and moving, explaining why it appears as the subject despite not necessarily being volitional. For the object argument of the sentence, Dowty proposes that Proto-patients are defined by a change of state, event-dependence, being causally affected, and stationary. So, while John and the zombie have no Proto-patient features, the window is stationary, undergoing a change of state, and is causally affected, hence it occurs in object position after *broke*. The account does not directly explain passive transitive sentences but instead assumes that some additional processes are needed. Thus, Dowty's Proto-roles approach argues that thematic roles are collections of features, none of which are defining, but work together to highlight the relative prominence of different arguments in sentences.

However, the proposed features of the Proto-roles account are abstract and just as problematic to define as traditional thematic roles. Concepts like volition and sentience are difficult to identify in others since we only have access to our own mental states (Descartes, 1637). For example, the distinction between manslaughter and murder depends on the intentions of the killer, and an entire jury is needed to make this distinction in trials since individual judgements of these concepts cannot be trusted.

An alternative account is that long-term knowledge about the entities in the event can guide thematic role assignment (Ferretti, McRae, & Hatherell, 2001; McRae et al., 1997). For example, our accumulated world knowledge informs us that *managers* typically do things to *employees*, *cooks* tend to do things to *food*, and *gardeners* typically to do things to *flowers*. It is suggested that this type of conceptual knowledge about different categories could be used to select appropriate thematic roles for each entity. For example, for an event involving a *teacher* and a *student*, world knowledge of these concepts would lead to an expectation that the teacher is the agent of a verb like *questioned* and the student is the patient. Consistent with this proposal, experiments have reported that participants can make real-time use of their broader world knowledge to guide thematic role assignment (e.g., Ferretti et al., 2001) and will generate expectations for the upcoming patient based on the specific actions described by the verb (e.g., Altmann & Kamide, 1999). Also, complex sentences have shown to be easier to process and understand when the relationship between the concepts is pragmatically typical (King & Just, 1991; Traxler, Morris, & Seely, 2002). Thus, McRae and colleagues have suggested that speakers form general thematic role concepts from an aggregation of verb-specific exemplars (Ferretti et al., 2001; McRae et al., 1997). However, the problem with this approach is that many events are novel, where entities occupy roles that they do not typically occur in. For example, if the teacher fell, then the student might be the agent of the verb *questioned* (e.g., *the student questioned the teacher about his leg*). This approach also has problems dealing with novel concepts; if you hear the sentence *the Zundit loves the Wimlet*, then you can

understand that the Zundit is the agent and the Wimlet is the patient, even though you have no world knowledge related to these imaginary creatures.

### **The Visual Perception of Thematic Roles**

There are many issues with accounts that rely on linguistic features or conceptual knowledge to identify thematic roles. One is that comprehenders often need to assign roles without having extensive knowledge of an entity's long-term behaviour (i.e., no world knowledge) or short-term behaviour (i.e., their internal states such as volition or sentience). Yet, perception research has provided extensive evidence that viewers are adept at identifying role-related information in visual scenes. In their seminal work using moving displays with simple shapes, Heider and Simmel (1944) found that viewers interpret the movement of the objects as those of animate beings with goals and intentions, producing descriptions such as "the big triangle is chasing the circles". This tendency to perceive animacy in the movement of simple shapes has been observed across different cultures (Morris & Peng, 1994) and age groups (Berry & Springer, 1993; Berry, Misovich, Kean, & Baron, 1992) using a variety of visual actions. For example, Barrett, Todd, Miller, and Blythe (2005) showed that German adults, German four-year-olds, and Amazonian Shuar adults could correctly distinguish six different forms of intentional interaction (chasing, fighting, courting, following, guarding, and playing) from videos of two moving arrow shapes. Furthermore, developmental research has found that infants younger than 12 months appear to interpret the movement of basic objects as those of goal-directed agents (Csibra, 2008; Csibra, Gergely, Bíró, Koos, & Brockbank, 1999; Gergely, Nádasdy, Csibra, & Bíró, 1995). Specifically, Gergely

et al. (1995) habituated 9-month-olds to visual displays where a small circle repeatedly approached a large circle by *jumping over* a rectangular obstacle. When the obstacle was removed at test, the infants looked for significantly longer when the exact same leap was performed in the absence of the rectangle compared to when the object moved in a novel but rational straight-line vector across the screen. This implies that the 9-month-old infants interpreted the small circle as a rational agent that was performing the jump for the purpose of reaching the larger circle. Thus, it appears that the features relevant for identifying thematic roles can be extracted in perceptual processing by pre-linguistic infants.

Related research has also considered how people understand causal interactions between objects. Using a range of different manipulations, Michotte (1946) examined the perception of causal relationships in displays with two moving shapes (e.g., square A and square B). In a widely replicated demonstration called the *launching effect*, he observed that when square A moves towards square B, and then B moves directly away from A after being contacted, the perception is that A is causally responsible for B's movement (see figure 1).

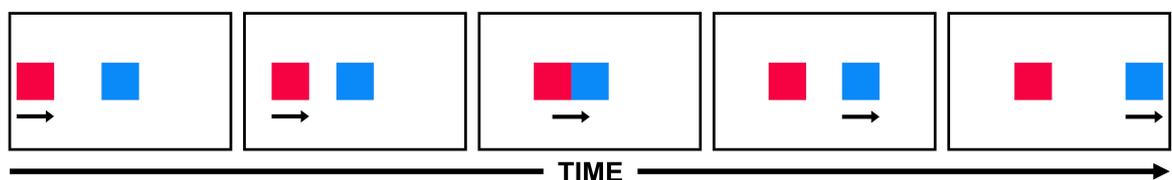


Figure 1. An illustration of Michotte's (1946) launch effect.

Many studies have found that these events are perceived as being causal when the appropriate perceptual indicators are present, namely when there is

physical contact and an immediate reaction (Rips, 2011; Schlottmann, Ray, Mitchell, & Demetriou, 2006; Scholl & Tremoulet, 2000; White, 2014; Young & Sutherland, 2009). This ability to identify causality appears early in development before 12 months of age, as habituation studies have consistently found that infant participants can distinguish causal from non-causal events based on the spatial and temporal features of these launch sequences (Leslie & Keeble, 1987; Oakes, 1994; Oakes & Cohen, 1990). For example, Oakes and Cohen (1990) presented infants with causal launching events or matched non-causal events with either no direct contact (a spatial gap) or a delay before the second object started moving (a temporal gap). The 10-month-olds that were habituated with the causal launching displays showed a significant increase in looking times and renewed interest when presented with a non-causal display at test, regardless of whether it contained a spatial or a temporal gap. However, when they were habituated with one of the non-causal displays (e.g., launching with a spatial gap), the infants only dishabituated and returned to looking at the screen when presented with a causal display and not when presented with the alternative non-causal event (e.g., launching with a temporal gap). Thus, features relevant for identifying thematic roles, such as the agent of a causal interaction, can be identified in perceptual processing by pre-linguistic infants, so it is possible that thematic roles are partially implemented within visual processing systems.

Collectively, the perception literature has provided strong evidence that viewers can perceive *who did what to whom* in their visual environment solely from movement cues and without needing information about the entity's internal states (e.g., volition) or the typical actions that they perform (e.g., world

knowledge: blocks and circles do not move on their own). Therefore, it is possible that visual perception could also provide a basis for thematic roles in language that does not rely on abstract conceptual definitions. The use of these low levels features does not necessarily contradict previous language-based treatments of thematic roles but instead offers a means of assigning these roles using concrete visual heuristics, which could be the origin of the concepts proposed in Dowty's (1991) proto-role theory. For instance, speakers may possess a proto-agent concept grounded in perceptual features such as self propulsion (see Luo & Baillargeon, 2005; Luo, Kaufman, & Baillargeon, 2009) rather than volition and sentience. However, the traditional view of cognition argues that language and vision belong to separate subsystems or modules with minimal contact (Fodor, 1983). Fodor's modularity hypothesis described a number of properties of these modules, but there are two features of primary importance. The first is *information encapsulation*, which posits that each distinct module does not have immediate access to the results of the online information processing conducted by other modules. The second is *domain specificity*, whereby each module is highly specialised and therefore only has access to a limited range of information directly related to its specific function. Under this view of cognition, the visual system would compute a representation of the scene and then pass this information to the language system at a later stage of processing.

In addition to these core properties, Fodor also argued that modules are innately specified and exist in the cognitive system from birth, consistent with Chomsky's earlier proposal that language users possess an innate universal grammar (Chomsky, 1959, 1965). Support for this idea has largely stemmed

from the observation that adults appear to possess abstract syntactic representations. As previously discussed, speakers can create grammatically valid sentences with any of the words in their lexicon to convey an infinite range of possible messages, while other language users can comprehend these novel utterances as long as they conform to grammatical rules. This suggests that adult syntactic knowledge is largely *abstract*, since the use of these structures does not rely on their meaning or the specific words they contain. A robust source of empirical evidence for the abstract nature of adult syntactic representations comes from experiments observing syntactic persistence effects; after reading or hearing sentences, speakers are more likely to reuse the syntactic structures they have recently experienced, even when the specific words that appear in these structures are different (Bock, 1986; Bock & Griffin, 2000; Bock & Loebell, 1990; Pickering & Branigan, 1998). A recent meta-analysis of over 70 experiments has confirmed the reliability of syntactic persistence (Mahowald, James, Futrell, & Gibson, 2016), providing strong evidence that speakers possess abstract representations of the different sentence structures available in their language.

### **Thematic Roles in Connectionist Models of Language**

There has been much debate on how speakers arrive at an abstract understanding of language. While many have postulated that children must be born with an innate universal grammar that becomes activated via experience (cf. Ambridge, Pine, & Lieven, 2014; Chomsky, 1965), psycholinguistic research has examined the aspects of language that can be acquired solely from input and the learning mechanisms necessary to *construct* a language without pre-

existing knowledge. There is considerable developmental evidence to suggest that language is acquired through a gradual learning process, starting from very a restricted and item dependent understanding that evolves into more general and abstract representations via continuous input (Lieven, 2016; Tomasello, 2000). Connectionist models have played a critical role in this discussion. These are a form of computational model that aim to simulate the language learning process, using statistical learning as the basis for acquiring linguistic representations. This follows substantial empirical evidence that humans possess powerful domain-general statistical learning abilities from early in development, which can be used to implicitly acquire the syntactic structure of artificial grammars or patterns in non-linguistic visual sequences in experimental settings (e.g., Aslin, Saffran, & Newport, 1998; Bulf, Johnson, & Valenza, 2011; Cleeremans & McClelland, 1991; Kirkham, Slemmer, & Johnson, 2002; Kirkham, Slemmer, Richardson, & Johnson, 2007; Saffran, Aslin, & Newport, 1996). Importantly, individual differences in performance on statistical learning measures can predict language learning abilities in both infants and adults (Conway, Baurnschmidt, Huang, & Pisoni, 2010; Kidd, 2012; Kidd & Arciuli, 2016; Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010), and elements that occur more frequently in natural language appear to be easier to learn and process (Ambridge, Kidd, Rowland, & Theakston, 2015; Bybee, 2006; Diessel, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Therefore, connectionist models, such as the Simple Recurrent Network (SRN), have used statistical learning to demonstrate that a system with no prior knowledge can learn reoccurring syntactic sequences (Chang, 2002; Christiansen & Chater, 1999; Cleeremans & McClelland, 1991; Cleeremans,

Servan-Schreiber, & McClelland, 1989; Elman, 1990, 1993), suggesting that innate universal grammar is not strictly necessary to explain the abstract nature of human language.

SRN models, as originally proposed by Elman (1990), learn to predict the next item (e.g., words) in the sequence based on previous input using a set of units organised into an input layer, a hidden layer, and an output layer. As the model is presented with sequences, the input layer processes them one item at a time and attempts to predict the next item in the sequence, which is held in the output layer. For each incoming item, the input layer projects to the output layer via a chain of hidden units, creating an activation pattern. Through a process called back-propagation, this activation pattern is used to adjust the strength of the (initially random) connection weights between the input and hidden layers. The SRN also has a context representation, where a copy of these hidden layer activation patterns are held for the previous input, providing a memory representation for the earlier parts of the sequence and thereby allowing the model to use context to guide future activation choices. Critically, adjustments in an SRN are driven by error-based learning, where the difference between the model's prediction of the next item and the actual input (the error signal) is used to guide adjustments to the connection weights. As particular connection routes between the input units and specific nodes in the output layer become stronger, they are more likely to be activated, which helps to improve the model's predictive abilities. For instance, after processing "the girl pushed..." in the input layer, the model may strongly activate "the boy" in the output layer if it has had previous experience with similar sentences. In this way, SRNs can acquire

linguistic knowledge purely as a consequence of processing language and adapting to the statistics of the input.

While connectionist research has provided explicit demonstrations of the potential learning mechanisms underlying language development and processing, a critical limitation of traditional SRN models is their inability to generalise in the same variable-like ways as human learners (Chang, 2002; Fodor & Pylyshyn, 1988; Marcus, 1998). In a critique of connectionist approaches, Fodor and Pylyshyn (1988) argued that mental representations in language are systematic and display certain symbolic symmetries, in that a language system capable of processing *the Zundit loves the Wimlet* should also be able to process *the Wimlet loves the Zundit*. As previously discussed, once speakers of English have learned that thematic roles can be arranged in a general AGENT-ACTION-PATIENT sequence, they can fill these slots with any word to produce novel transitive sentences. Although traditional connectionist models can learn the reoccurring sequences of their input, these systems are often unsuccessful in generalising words to new sentence positions and continue to be highly dependent on the specific concept-role combinations they experienced in training (Chang, 2002; Marcus, 1998). When presented with an unfamiliar sentence fragment (e.g., *the Wimlet loves...*), the model would activate only the words that it has experienced in the upcoming sentence position (e.g., *Wimlet*), as it received positive evidence in the training input that *loves* is followed by *Wimlet*. Since the frequency of *loves* followed by *Zundit* is zero in the input, it would consider *the Wimlet loves the Zundit* to be an ungrammatical sentence. Chang (2002) demonstrated this problem with a model of production called the Prod-SRN. This model combined an SRN with a

message that had different sets of units for each thematic role. For the above sentence, there would be a unit that represented ZUNDIT as an agent (ZUNDITAGENT) and another for this concept as a patient (ZUNDITPATIENT). Likewise, there were two units for Wimlet (WIMLETAGENT, WIMLETPATIENT). In training, ZUNDITAGENT and WIMLETPATIENT were presented, allowing the model to generate the sentence *the Zundit loves the Wimlet*. But since WIMLETAGENT and ZUNDITPATIENT were not trained, the same model could not produce *the Wimlet loves the Zundit*. The problem with this model is that separate units represent each combination of roles and concepts, so the model cannot activate roles and concepts separately (ZUNDITAGENT and ZUNDITPATIENT are completely unrelated). Even with thematic roles in the message, SRN models would need to be exposed to every possible concept in every possible slot to exhibit systematicity. However, these networks would still be unable to make the leap to a general understanding of AGENT-ACTION-PATIENT that could be used with any additional words that are introduced later. Consequently, traditional SRN models are unable to explain the symbolic productivity of human language.

To address this problem with systematicity, Chang (2002) developed a new connectionist model that used variables to encode thematic role-concept bindings. In this model, there were a set of units for roles and a separate set of units for concepts. The bindings between these units were represented with weighted connections, where a high weight between the AGENT role and the ZUNDIT concept would mean that the Zundit was the agent in the message. This approach makes it possible to learn how to activate roles (independent of the concept) and also learn how to map concepts to words (independent of the

role). These representations were developed using a new connectionist architecture called the Dual-path model (illustrated in figure 2). This model consisted of a *meaning pathway* that contained the message being conveyed, which was separate from the *sequencing pathway* that learned the statistical regularities of the language like a traditional SRN. This meaning pathway was comprised of a network of *where* and *what* units. The *what* units represent the concepts being discussed (e.g., the GIRL and BOY in *the girl pushed the boy*) and are connected to separate word nodes that hold the lexical labels for these concepts, allowing the model to learn the underlying meaning of specific words by activating the possible concepts attached to the labels. Alternatively, the *where* units provided the model with a set of thematic roles (agent, patient, and goal) and a separate node to hold action information. Using these two features, the meaning pathway can encode the message of any sentence by creating temporary variable bindings between the concepts held in the *what* units and the thematic roles and actions held in the *where* units. For example, if the *girl* was the agent, then the agent *where* unit would be connected to the GIRL concept in the *what* units, but if she was the patient, then the patient *where* unit could instead be connected to the same GIRL concept in the *what* layer. In Chang (2002), this Dual-path model was trained with 501 English sentences randomly sampled from a grammar that could formulate over 75,000 possible messages. After training, the model's accuracy in predicting upcoming words was tested using a new set of 2000 randomly generated sentences from the grammar.

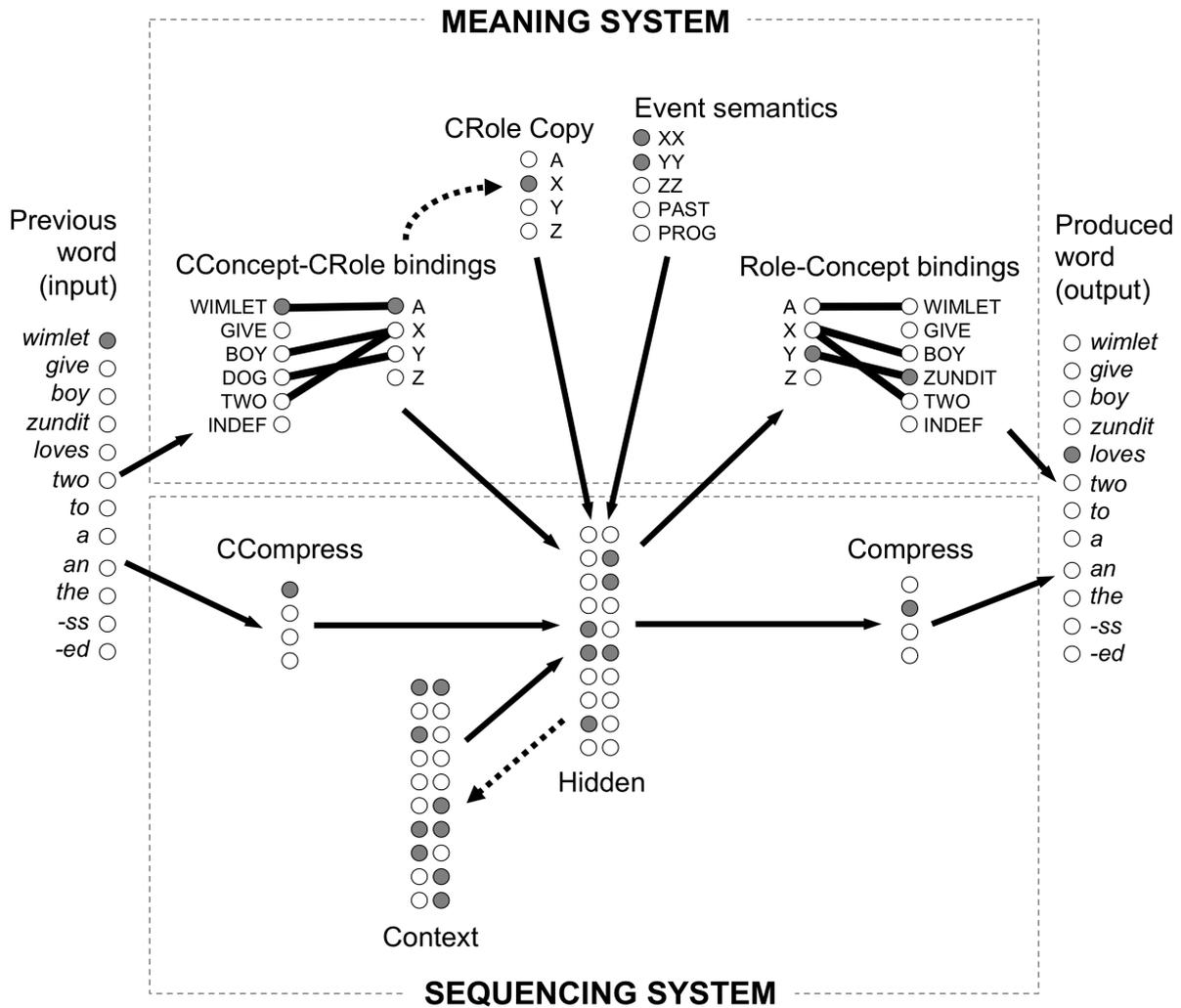


Figure 2. An illustration of the Dual-path model of language (Chang, 2002).

While these sequencing and meaning pathways are separate networks in the Dual-path model, the two are also connected via the *where* units (or thematic roles). Thus, the aim of the sequencing pathway was to learn how the *where* units are arranged in the input language using their statistical frequencies. Critically, the sequencing system does not have direct access to the content of the *what* units and the specific message being conveyed in the sentence, but instead only learns from the activation patterns of the *where* variables. This enables the Dual-path model to demonstrate more symbolic

abilities, as it is trained with the regularities of thematic role variables such as AGENT-ACTION-PATIENT rather than lexically-specific sequences like GIRL-PUSHED-BOY. Consequently, the model can generalise nouns to other thematic roles and sentence structures and produce novel adjective-noun pairs. These role variables also allow unfamiliar elements to be temporarily bound to thematic roles (e.g., ZUNDIT-AGENT; LOVES-ACTION; WIMLET-PATIENT) and thus be used in constructing or comprehending novel sentences (e.g., *the Zundit loves the Wimlet*). By using thematic roles as a fast variable-binding mechanism to connect meaning to syntactic representations, the Dual-path model has been able to account for a vast range of psycholinguistic phenomena, including structural choice and syntactic persistence in English and German (Chang, Dell, & Bock, 2006), the accessibility hierarchy and the relative clause asymmetry (Fitz, Chang, & Christiansen, 2011), heavy noun phrase shifts in English and Japanese (Chang, 2009), the acquisition of verb bias classes (Twomey, Chang, & Ambridge, 2014), auxiliary inversion rules in forming questions in English (Fitz & Chang, 2017), the acquisition of morphosyntactic rules in second language learners (Janciauskas & Chang, 2017), and event-related potentials in sentence processing experiments (Fitz & Chang, 2018).

An important claim of the Dual-path model is that fast and temporary bindings between concepts (*what* units) and thematic role variables (*where* units) are required to produce sentences where the message involves using concepts in role configurations that the model has not previously experienced. To be productive in language, it is necessary to be able to bind roles and concepts that have never appeared together in the input. However, it is not possible to develop this ability using only the statistical learning mechanisms of

an SRN, because all of these mechanisms require at least one input example. Before you can understand the novel sentence *the Chihuahua bit Trump* in a statistical learning system, you would need to have similar sentences like *the Chihuahua bit the bone* and *the boxer bit Trump* paired with their target meanings in the training input experience, but this is not always the case. The Dual-path model argues that language users can generalise to novel sentences due to a built-in variable-like ability to bind roles and concepts. Since these fast binding abilities cannot be explained by statistical learning systems or the slowly changing biological mechanisms that support neural learning, Chang and colleagues have argued that these abilities must have evolved independently for another function and were later adapted by language (Chang, 2002; Chang et al., 2006). In particular, the visual system is capable of tracking multiple objects by quickly attaching pointers to entities in the visual field and updating these pointers as they change location (Pylyshyn & Storm, 1988).

There is considerable evidence showing that the visual system has evolved specialized mechanisms and representations to support this tracking ability. One important requirement is to separate location tracking from the visual properties of the object. Research suggests that the neural structure of the visual system consists of two separate cortical pathways: the dorsal (or *where*) stream that focuses on object motion and location to allow viewers to interact with objects, and the ventral (or *what*) stream which extracts detailed features to build high-level perceptual representations of objects so that they can be identified (Goodale & Milner, 1992; Milner & Goodale, 2008). Imaging studies have observed activation patterns to suggest that tracking the location of a moving object primarily involves regions of the parietal lobe (Howe, Horowitz,

Akos Morocz, Wolfe, & Livingstone, 2009), which encompasses the dorsal pathway. Patients with brain lesions to cortical areas along this dorsal pathway show an impaired ability to track objects in the visual field contralateral to their damaged hemisphere (Battelli et al., 2001). Conversely, those with damage to the ventral pathway typically have normal motion perception, but often display signs of visual agnosia, a condition characterised by an inability to visually identify objects (Karnath, Rüter, Mandler, & Himmelbach, 2009). This dissociation is further supported by studies showing that viewers can track the location of a target without always being able to identify it (Pylyshyn, 2004). Critically, after input is analysed in these separate processing pathways, these two information streams can be quickly integrated to provide a complete visual experience (Bullier, 2001; Karnath, 2001). Without this ability to bind concepts to specific locations, viewers would be unable to complete tasks such as retaining the specific identity of a moving entity (Horowitz et al., 2007; Oksama & Hyönä, 2016) or recognising specific actions (e.g. walking) from relative movement patterns (Giese & Poggio, 2003). Therefore, the Dual-path model's ability to flexibly attach concepts (*what*) to thematic roles (*where*) in the meaning network could exist as an adaptation of the mechanisms used to rapidly bind object identity to spatial locations in visual processing.

### **Multiple Object Tracking**

To summarise, thematic roles provide a potential method of representing the meaning of events in a way that is independent of specific concepts, words, or sentence structures in language. Linguistic theories have attempted to explain thematic roles in terms of abstract features (e.g., volition; Dowty, 1991), but

these are difficult to define. However, statistical approaches have also struggled to explain the how learners arrive at abstract generalisations of language from their input without using thematic roles (Chang, 2002; Fodor & Pylyshyn, 1988; Marcus, 1998). The solution to these issues may be in spatial processing. Visual perception research has demonstrated that viewers can identify thematic relationships from measurable motion features, such as self-propulsion (Luo & Baillargeon, 2005; Luo et al., 2009), physical contact between the event participants (Young & Sutherland, 2009), the timing of their movement patterns (Oakes, 1994), the directness of agent's movement towards the patient (Gao, Newman, & Scholl, 2009), and many others. Thus, a vision-based approach could provide role-related features that are more concrete than those used in current theories of thematic roles. It would also lend support to connectionist models that have used thematic role variables to achieve novel generalisations and explain many findings reported in the psycholinguistic literature (Chang, 2002; Chang et al., 2006). Therefore, the aim of this thesis was to link thematic roles in language to perception studies reporting that participants can infer these roles from the movement of basic shapes (Scholl & Tremoulet, 2000). To test this connection, a series of experiments were conducted with adult participants using a multiple object tracking task, which required them to monitor randomly moving white circles that could only be distinguished by tracking their movement throughout the trial. The experiments tested whether thematic roles can be extracted and bound to visually identical objects as they change location, and whether this can have an immediate impact on language production. Specifically, the task depicted causal pushing events between the objects and required the participants to describe these interactions in a sentence, after a

period of tracking their movement. Before describing these studies in more detail, the object tracking literature will be reviewed to provide a clearer understanding of the hypotheses being tested in this thesis.

Multiple object tracking (MOT) abilities were first studied extensively by Pylyshyn and Storm (1988), who presented participants with scenes in which ten white crosses moved in a random manner. At the start of the experiment, a subgroup of these objects would be designated as the targets before all the objects started moving randomly. Later, one object would flash and participants reported whether it belonged to the target group. Even though the objects were indistinguishable in their visual appearance, the participants were successful in tracking the location of up to five targets simultaneously.

Following these findings, a plethora of research has been conducted into MOT, providing many insights into the nature of this ability. One well established (and somewhat intuitive) finding is that there is a limit to the number of objects we can track. While Pylyshyn and Storm showed that participants can track several objects at once, they also observed a significant linear decline in accuracy as the number of targets increased. Early estimates suggested the tracking system had a fixed capacity of approximately four objects (Pylyshyn, 1989), but more recent findings have observed a flexible tracking limit that is dependent on a number of different variables. Specifically, increasing the speed (Alvarez & Franconeri, 2007; Feria, 2013; Holcombe & Chen, 2012; Tombu & Seiffert, 2008, 2011), object crowding (Franconeri, Jonathan, & Scimeca, 2010; Franconeri, Lin, Enns, Pylyshyn, & Fisher, 2008; Tombu & Seiffert, 2008), the number of distractor objects (Bettencourt & Somers, 2009; Sears & Pylyshyn, 2000), or adding concurrent response tasks (Allen, McGeorge, Pearson, & Milne,

2006; Tombu & Seiffert, 2008) has shown to make object tracking more difficult and effectively reduce the number of targets that can be tracked. Tracking performance also appears to improve throughout childhood into early adulthood, but then decline in later adulthood (Ryokai, Farzin, Kaltman, & Niemeyer, 2013; Sekuler, McLaughlin, & Yotsumoto, 2008; Trick, Jaspers-Fayer, & Sethi, 2005; Trick et al., 2005). Furthermore, studies of video game players and radar operators have reported greater tracking performance in these populations, which appears to be due to experience, since providing such training to novices has shown to improve their tracking accuracy (Allen, McGeorge, Pearson, & Milne, 2004; Green & Bavelier, 2006; Sekuler et al., 2008; Trick et al., 2005). Regardless, while there may not be a definitive number of objects that humans can track, only a small number of objects can be monitored simultaneously, with some studies showing an upper limit of around 8 objects even in the least demanding conditions (Alvarez & Franconeri, 2007).

There is also substantial evidence that viewers can track several objects at the same time, without needing to constantly switch their attention between each target. Many experiments have observed that participants do not need to use eye movements to track targets among distractors (Luu & Howe, 2015; Pylyshyn & Storm, 1988). Instead, several eye-tracking studies have reported that most participants tend to fixate their gaze in a central position between all of the target objects and are often more successful in the tracking task compared to those using a gaze-switching strategy (Fehd & Seiffert, 2008, 2010; Huff, Meyerhoff, Papenmeier, & Jahn, 2010; Oksama & Hyönä, 2016; Zelinsky & Neider, 2008). Furthermore, Howe, Cohen, Pinto, and Horowitz (2010) report a series of experiments where participants tracked multiple objects that moved

either simultaneously or one at a time, observing that tracking accuracy in the simultaneous conditions was consistently equal to or higher than the sequential tracking conditions. Collectively, such findings suggest that MOT does not require the use of overt focal attention and may even be hindered when viewers try to switch their attention between the targets.

To explain these abilities, many theories of MOT have argued that the visual system must contain a set of pointers that connect to objects to continuously track their location as they move (e.g., Alvarez & Franconeri, 2007; Cavanagh & Alvarez, 2005; Leslie, Xu, Tremoulet, & Scholl, 1998; Pylyshyn, 1989). This was first suggested in the highly influential FINST (Fingers of Instantiation) account (Pylyshyn, 1989; Pylyshyn & Storm, 1988), which proposed that object tracking is carried out by a parallel mechanism containing four or five indexes that 'stick' to objects and provide a pathway their location independent of high-level memory representations and feature recognition. Kahneman, Treisman and Gibbs (1992) similarly argued that temporary episodic representations called *object files* are necessary for perceiving a dynamically changing visual scene. These object file representations are created when viewers focus on specific targets and are used to store information about the objects and bind it to their location in the visual field, while remaining separate from the long-term recognition network. However, a system with fixed pointers cannot explain evidence showing that our tracking capacity is flexible. Thus, more recent models have argued that object pointers are a form of attention that can be divided into independent focal points and assigned to different target locations in the visual field (Cavanagh & Alvarez, 2005). It is argued that these attentional resources are limited but can be flexibly distributed (Alvarez &

Franconeri, 2007), allowing viewers to assign additional attention to individual targets when tracking their location is more difficult (e.g., due to an increase in speed), or alternatively attend to a larger number of separate objects in less demanding displays. Within this framework, poor performance in retaining the location of multiple targets arises from the depletion or excessive division of the tracking resource, preventing the viewers from attaching pointers to all of the objects in the target group (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012). This approach helps to explain the effect that attentionally demanding properties of the task can have on tracking capacity (Alvarez & Franconeri, 2007; Holcombe & Chen, 2012; Tombu & Seiffert, 2008), and the individual variability in tracking capacity reported in many experiments (e.g., Oksama & Hyönä, 2004). Regardless of their precise nature, object pointers are widely considered to be a critical part of visual cognition, allowing viewers to attend to several objects simultaneously and explaining many important findings from adult and infant object perception research (Leslie et al., 1998; Pylyshyn et al., 1994).

Research using object tracking tasks with chasing interactions has connected the perceptual causality and object tracking literature, while demonstrating the perceptual saliency of such causal relations. Gao, Newman, and Scholl (2009) showed participants displays containing several identical circles, in which one object (the wolf) is chasing another (the sheep). Participants were highly accurate in detecting the presence of a chasing relationship and in identifying the wolf and sheep targets among the distractors. The saliency of such chasing relations appeared to be a function of the wolf's angle of pursuit. Participants could readily perceive the chase when the wolf

moved in a somewhat direct path towards the sheep, but detection became increasingly impaired as this angle of approach became less direct and the chase was more subtle. These findings were consistent with those of Dittrich and Lea's (1994) earlier work, who found that chasing directness was a primary determiner in whether the wolf object was perceived to be an intentional agent. In a subsequent study, Gao and Scholl (2011) gave participants control of the sheep and tasked them with avoiding the wolf. They found that participants were able to detect the wolf object in the display and avoid it, but that this was impaired in the presence negative evidence; episodes of random movement or deviations from direct chasing reduced rates of successful escape from the wolf, whereas periods in which the wolf was static or jiggled did not. Collectively, the findings from these studies show that viewers are able to assign agent and patient roles to animate entities in the visual environment based entirely on features extracted from tracking their movement patterns. Furthermore, such abilities have been observed in infants, who will preferentially attend to chasing scenes over those with non-interacting motion (Frankenhuis, House, Clark Barrett, & Johnson, 2013; Galazka & Nyström, 2016; Rochat, Morgan, & Carpenter, 1997). This provides evidence supporting the current hypothesis that thematic roles could be supported by the ability to track the location of different objects.

### **The Present Work**

Since we are able to communicate our perception of the visual world through language, there is undoubtedly a cognitive relationship between language and vision. For most, language is acquired in a complex visual

environment, and evidence suggests that children are able to use language to dissect their visual world from a very young age (Dessaiegn & Landau, 2013). It is becoming increasingly apparent that many mental processes involve the multimodal convergence of sensory input (Driver & Spence, 1998; Hidaka, Teramoto, & Sugita, 2015; Lalanne & Lorenceau, 2004), with substantial evidence that language and vision interact in real-time and in non-trivial ways (Anderson, Chiu, Huette, & Spivey, 2011). For instance, experiments using the visual world paradigm have consistently observed participants' eye movements around the visual environment to be closely linked to linguistic input (Huettig, Rommers, & Meyer, 2011). Despite such findings, the mechanisms of the visual system are rarely incorporated into psycholinguistic theories, which typically treat language processing as being mostly separate from visual processing systems (Levelt, 1989; Pickering & Garrod, 2013; van Gompel & Pickering, 2007). These approaches assume that language is a formal system that can be understood independently of visual-spatial processing in the brain, deriving from the assumption that language and vision are modular with respect to each other (Fodor, 1983).

The experiments in this thesis test whether thematic roles can be identified and supported by tracking object movement, then passed directly to the language system for use in sentence production, as hypothesised in the Dual-path model of language (Chang, 2002; Chang et al., 2006). The studies manipulated a range of visual features in a tracking task and examined how those visual manipulations affect sentence production accuracy. The specific features that were manipulated were based on previous work in visual object tracking and sentence processing. If the language system incorporates fast-

binding variables that are derived from those used to track objects in vision, then we should expect a range of visual manipulations to influence accuracy.

The thesis provides data from a total of nine experiments that investigated whether thematic roles can be supported by the visual object tracking system. Chapter two introduces the novel multiple push tracking task used throughout all the experiments presented in this thesis. The task is a modified form of the widely-used multiple object tracking paradigm devised by Pylyshyn and Storm (1988). In the adapted version, the participants were required to track the agents and patients of causal push events among nine randomly moving white circles that were identical in appearance. At test, two circles from one of the push events and a foil object were highlighted and the participants had to describe how these circles interacted using an active transitive sentence, such as *red pushed blue*. Following extensive evidence that viewers can track the location of several targets in parallel (e.g., Oksama & Hyönä, 2016; Pylyshyn & Storm, 1988), this chapter presents the results of three experiments that investigated whether the thematic role features of identical targets can be tracked without relying on overt serial attention. Specifically, these studies tested whether participants can track agents and patients and generate accurate descriptions under difficult conditions, such as when tracking multiple push events (experiments 1-3), fixating their gaze (experiment 1), performing a concurrent speeded-response task (experiment 2), and when tracking objects that were temporarily invisible (experiment 3). The primary goal of this chapter was to establish whether thematic roles in language could be an extension of perceptual abilities that take place in the visual object tracking system.

In chapter three, the multiple push tracking task is applied to an important issue in psycholinguistics: the asymmetry in the comprehension of subject-extracted and object-extracted relative clauses (SRC/ORC). Many studies have observed that sentences containing an SRC, such as 1, are easier to read and understand than those containing an ORC, such as 2 (Gordon & Lowder, 2012; Traxler et al., 2002; Wanner & Maratsos, 1978).

1. The reporter that attacked the senator admitted the error [SRC].
2. The reporter that the senator attacked admitted the error [ORC].

Since an SRC advantage has been observed across many languages with different properties (English: King & Just, 1991; Korean: Kwon, Gordon, Lee, Kluender, & Polinsky, 2010; Dutch: Mak, Vonk, & Schriefers, 2002; Japanese: Miyamoto & Nakamura, 2003), chapter three investigated whether there is a non-linguistic bias for these events arising from differences in the consistency of their role-referent bindings. This was examined using the same multiple push tracking task as chapter 2. However, the trials always involved two pushes and there was an overlapping circle that appeared in both events. In the SRC-target conditions, one of the circles in the display played the agent in both pushes. In the ORC-target conditions, the patient of the first push was the agent of the second. In experiment four, the participants were required to describe the trials using an SRC (e.g. *the red that pushed blue pushed green*) or an ORC (e.g., *the red that blue pushed pushed green*). In experiment five, the ORCs were substituted with a passive relative clause (e.g. *the red that was pushed by blue pushed green*), which are a preferred alternative form in English (Gennari & MacDonald, 2009; Gennari, Mirković, & MacDonald, 2012). Experiment six had

the participants describe the same stimuli with active transitive sentences (e.g., *red pushed blue*), to remove the response differences between the two conditions. Since an SRC bias was observed throughout these studies, three additional experiments were conducted to examine how tracking thematic role switches affects sentence accuracy. Here, the participants produced active transitive (experiment seven/eight) or passive transitive sentences (experiment nine) to describe scenes in which a target agent or patient played the same role in both pushes (i.e., a double agent/double patient), switched to the alternative role between the two events (agent-patient/patient-agent), or where there were no overlapping circles between the pushes (control condition). The overall aim of chapter three was to examine whether tracking thematic role switches in visual events can affect description accuracy in language, and whether this could explain the persistent SRC preference reported in comprehension research (Gordon & Lowder, 2012).

In this thesis, nine experiments are reported that manipulate thematic role-related variables within an MOT task to examine their influence on language use. If these manipulations of purely visual factors can change the accuracy of sentence production, this would support the idea that visual object tracking can provide features that can be used to identify thematic roles in language.

## CHAPTER TWO: TRACKING THEMATIC ROLES IN THE VISUAL SYSTEM

### Fit Within the Thesis

Although thematic roles are often used to characterise language meaning and can support systematic generalisations in connectionist models (Chang, 2002; Chang et al., 2006), it is unclear how they are identified from input. One possibility is these roles are computed during spatial processing; when viewers track the movement of simple shapes, they will often interpret the sequences as intentional interactions involving animate agents based on relational patterns between the objects (Barrett et al., 2005; Heider & Simmel, 1944; Leslie & Keeble, 1987; Oakes, 1994; Schlottmann et al., 2006; Young & Sutherland, 2009). Therefore, chapter two presents three experiments that examine whether role-related features can be supported by the visual object tracking system. The participants were required to track the agents and patients of multiple push events (between 1-3) in visual scenes involving nine identical randomly moving circles. At test, two circles from one of the push events and a foil object were highlighted and the participants had to describe their interaction with an active sentence, such as *red pushed blue*. These studies found that the participants could track the agent and patient targets and generate descriptions that identified their thematic roles at above chance levels, even under difficult conditions, such as when tracking multiple push events (exp. 1-3), fixating their gaze (exp. 1), performing a concurrent speeded-response task (exp. 2), and when tracking objects that were temporarily invisible (exp. 3). Furthermore, accuracy in identifying agents was around 6% higher than patients in all three experiments. Collectively, the studies in chapter two demonstrate that thematic role information can be maintained when tracking the random motion of visually

identical objects and can then be used map role fillers (e.g., the agent of a push event) into their appropriate sentence positions. This suggests that core thematic role processes may take place in the visual object tracking system.

## **Introduction**

An important function of language is to express who did what to whom in an event. For instance, in the sentence *the girl pushed the boy*, the girl is the agent that is causing the push and the boy is the patient that is being pushed. Agents and patients are examples of thematic roles (Fillmore, 1967; Gruber, 1965; Jackendoff, 1972, 1987), which are functions that describe the relationships between entities in events and capture the similarity in meaning between different utterances. For example, English speakers can describe an event using an active transitive structure such as *the girl pushed the boy*, with the agent appearing before the verb and the patient after. Alternatively, other word orders could be used such as a passive structure (*the boy was pushed by the girl*), or the same event could be described in another language with an entirely different word order. Therefore, thematic roles provide a way of encoding language meaning that serves as an interface between the perception of scenes and language-specific word orders.

Despite their importance, it has been difficult to define the specific features that reliably identify thematic roles such as the agent and patient in different contexts (Dowty, 1991; Fillmore, 1967; Jackendoff, 1972; McRae et al., 1997). To address this issue, Dowty (1991) hypothesised that nouns are mapped into sentence arguments using two conceptual prototypes: the proto-agent and proto-patient. Proto-agent features include being event-independent,

sentient, volitional, causally responsible and moving. When deciding on the subject of the sentence, each entity in the event is checked against these features. For example, in *John broke the window*, *John* has all of the Proto-agent features, while the *window* is only event-independent, hence John should be the subject. Thus, Dowty's Proto-role theory argues that conceptual features can be used to directly determine the prominence of different arguments without explicitly identifying thematic roles.

Although feature-based linguistic accounts of thematic roles are popular (e.g., Dowty, 1991), the proposed features such as sentience and volition are themselves difficult to define. However, many perception studies have shown that participants appear to understand the movement of simple shapes in ways that are related to their thematic roles (e.g., Barrett et al., 2005; Heider & Simmel, 1944). One paradigm that examines the visual perception of agents and patients is Michotte's (1946) launching display, which involves two moving shapes (e.g., square A and square B). Michotte observed that when square A moves towards square B, and then B moves directly away from A after being contacted, the impression is that agent A is causally responsible for patient B's movement (Scholl & Tremoulet, 2000). Subsequent research has shown that this effect is strongest when there is physical contact and an immediate reaction (e.g., Schlottmann et al., 2006; Young & Sutherland, 2009). Even young infants in their first year are capable of distinguishing causal from non-causal events in habituation studies based on these movement properties (Leslie & Keeble, 1987; Oakes, 1994; Oakes & Cohen, 1990). This suggests that features relevant for identifying agents, like causality, can be computed in perceptual processing by non-linguistic infants.

A necessary prerequisite for identifying thematic roles in visual events is to track the objects involved and accumulate the required evidence to support role identification. Gao, Newman, and Scholl (2009) examined this ability by presenting videos of identical circles that moved in semi-random paths. One of these circles was a wolf and would move towards and chase another circle (the sheep). Participants were highly accurate in detecting the chase and in identifying the wolf among the distractors, with performance deteriorating as the wolf's angle of approach became less direct. A variety of other chasing studies have reported similar results (Dittrich & Lea, 1994; Gao & Scholl, 2011) with developmental work showing that infants are able to perceive chasing relationships from as young as 3 to 4-months (Frankenhuis et al., 2013; Galazka & Nyström, 2016; Rochat et al., 1997). Interestingly, chase detection appears to be negatively affected by the number of potential wolves and sheep in the display (Meyerhoff, Huff, & Schwan, 2013), consistent with findings that there is a limit to the number of objects that can be tracked (e.g., Pylyshyn & Storm, 1988). Merging the wolf and sheep with the distractor objects by connecting them with solid lines has also shown to severely disrupt the ability to detect chasing relationships between the objects (van Buren, Gao, & Scholl, 2017), mirroring the findings of object tracking studies similarly showing that target-merging significantly reduces accuracy in identifying the target objects (Howe, Incledon, & Little, 2012; Scholl, Pylyshyn, & Feldman, 2001). Since the wolf is the agent of the chasing action and the sheep is the patient, these findings suggest that relational features, such as angle of approach, are maintained between pairs of objects as they are tracked and could help identify thematic roles. Other work has similarly suggested that force dynamics and

psychophysical features are important for both causal perception and causal language (e.g., Talmy, 1988; Wolff, 2007).

Research on chasing interactions suggests that thematic roles can be computed by tracking the movement of the various objects in the scene simultaneously, an ability that has been studied extensively using the multiple object tracking (MOT) paradigm. In their seminal MOT work, Pylyshyn and Storm (1988) showed participants a set of identical objects (white crosses) with a subset briefly identified as the target objects. The objects then moved in a random manner for a short period, before the participants were queried on whether a particular object was a target. They found that participants achieved high accuracy when tracking up to five crosses simultaneously, demonstrating that the visual system can maintain multiple objects even when they are visually indistinguishable. To explain this, they proposed that object tracking is carried out by a parallel mechanism containing four or five pointers that 'stick' to objects (Pylyshyn, 1989; Pylyshyn & Storm, 1988). Neuroimaging research has suggested that regions of the dorsal visual pathway are the primary cortical areas responsible for both multiple object tracking (Battelli et al., 2001; Howe et al., 2009) and the perception of causality in launch events (Blakemore et al., 2001; Fugelsang, Roser, Corballis, Gazzaniga, & Dunbar, 2005; Straube, Wolk, & Chatterjee, 2011; Woods et al., 2014). Thus there is behavioural and neuroimaging evidence for the link between object tracking system and the systems that store role-related relational features.

Subsequent MOT studies have confirmed that only a small number of targets can be monitored simultaneously, but there does not seem to be a definitive limit. Instead, tracking capacity appears to be flexible and largely

determined by both the attentional demands of the task (Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009; Franconeri et al., 2010; Tombu & Seiffert, 2008) and individual capabilities (Green & Bavelier, 2006; Oksama & Hyönä, 2004; Sekuler et al., 2008; Trick et al., 2005). Furthermore, tracking accuracy falls linearly as the number of targets is increased (Oksama & Hyönä, 2004; Pylyshyn & Storm, 1988) and viewers sometimes appear to use serial attention switching strategies instead of parallel tracking (Oksama & Hyönä, 2004). However, participants can separate randomly moving targets from distractors without using conscious eye movements (Luu & Howe, 2015; Pylyshyn & Storm, 1988) and many eye-tracking studies have reported that viewers typically prefer to fix their gaze in a position between all of the targets (the centroid) rather than switching their gaze from object to object (Fehd & Seiffert, 2008, 2010; Huff, Papenmeier, Jahn, & Hesse, 2010; Oksama & Hyönä, 2016; Zelinsky & Neider, 2008). Interestingly, tracking accuracy has shown to be higher when the viewers attend to the targets simultaneously rather than sequentially (Fehd & Seiffert, 2010; Howe et al., 2010; Zelinsky & Neider, 2008). Therefore, many researchers have concluded that a small number of objects can be monitored in parallel (Alvarez & Cavanagh, 2005; Cavanagh & Alvarez, 2005; Howe et al., 2010; Oksama & Hyönä, 2016; Pylyshyn, 1989; Yantis, 1992), as the available data cannot be entirely explained by attention switching or a single focus over the entire display. In the present research, the general term *object pointers* is used to refer to the parts of the visual system that provide a pathway to objects in the scene. Recent models have characterised these pointers as a form of multifocal attention, where each target simultaneously receives an independent focus within the limits of our available resources (Alvarez & Cavanagh, 2005;

Cavanagh & Alvarez, 2005). These attentional resources appear to be flexibly allocated (Alvarez & Franconeri, 2007), which helps to explain the variability in tracking capacity reported in many experiments (e.g., Oksama & Hyönä, 2004). Thus, MOT research suggests that we have a limited capacity for tracking multiple objects in parallel. It is possible that these limits also apply to the role-related features that are supported by the object tracking system.

Although object tracking is necessary to accumulate perceptual information for identifying thematic roles (e.g., the angle of approach for identifying the agent of chasing), many features of the scene are not automatically bound to object pointers. MOT studies have observed that participants will often fail to detect colour or shape changes on target objects and cannot always identify specific targets even when they can successfully track their location (Bahrami, 2003; Horowitz et al., 2007; Pylyshyn, 2004; Saiki, 2003). Binding features from different perceptual dimensions and tracking these object representations appears to require focused serial attention (Oksama & Hyönä, 2008; Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989). This distinction can be observed using eye-tracking; while participants favour centroid fixating when tracking only the location of the targets, they often utilise active gaze switching when tracking the identities of visually distinct objects (Oksama & Hyönä, 2016). This raises the question of whether tracking the features that encode the thematic role of an object requires focused serial attention, or whether they can be tracked using parallel processes within the capacity limitations of the visual system.

Although there is extensive research on thematic role-related motion feature processing in visual perception, this research has not been connected to

linguistic or psycholinguistic theories. Literature reviews of linguistic semantics or meaning in language processing typically do not integrate work in perception (e.g., Ferreira & Slevc, 2007; Lappin & Fox, 2015; van Gompel & Pickering, 2007). In addition, psycholinguistic studies of thematic roles often look at the semantic properties of their fillers (e.g., doctors are typical agents of verbs like *operate*; Ferretti et al., 2001; Hare, Jones, Thomson, Kelly, & McRae, 2009; McRae & Matsuki, 2009; McRae et al., 1997), resulting in theories that are not able to explain the identification of roles in scenes with visually identical objects, such as those commonly used in MOT studies. To bridge between these two literatures, the studies evaluate how visual variables in MOT videos influence thematic role accuracy in linguistic descriptions of causal pushing actions. In these videos, participants see push events in isolation, so it is clear that the action is encoded. After the event is shown, the objects move around randomly among distractors, and it is examined whether people can remember which of the multiple identical objects was the agent and patient in the previously seen pushing action. To do this, they need to maintain the relational features that encode causality with the pointers that track the objects. If this is the case, then it would suggest that thematic role features are encoded in the object tracking system, which in turn would highlight a need for the visual system to be included in psycholinguistic theories of meaning.

### **Experiment One: Tracking Roles Without Eye Movements**

To examine whether visual variables influence the ability to track thematic roles and encode them in descriptions, a challenging MOT task was developed that required participants to describe a single push interaction out of up to three

events. In this task, participants saw displays where nine identical circles moved around randomly (see figure 3A). Occasionally, these circles would temporarily stop moving and a push event would take place (figure 3B). All the circles would then resume their random movement patterns (figure 3C). At test, two circles from one of the push events and an extra foil object were given random colours (figure 3D) and the participants were asked to describe the action that occurred between these objects in an active sentence like *green pushed red*. Since the objects were identical and the test phase was separated from the push events by periods of random motion, the only way to correctly describe the push relationship was to track the agent and patient.

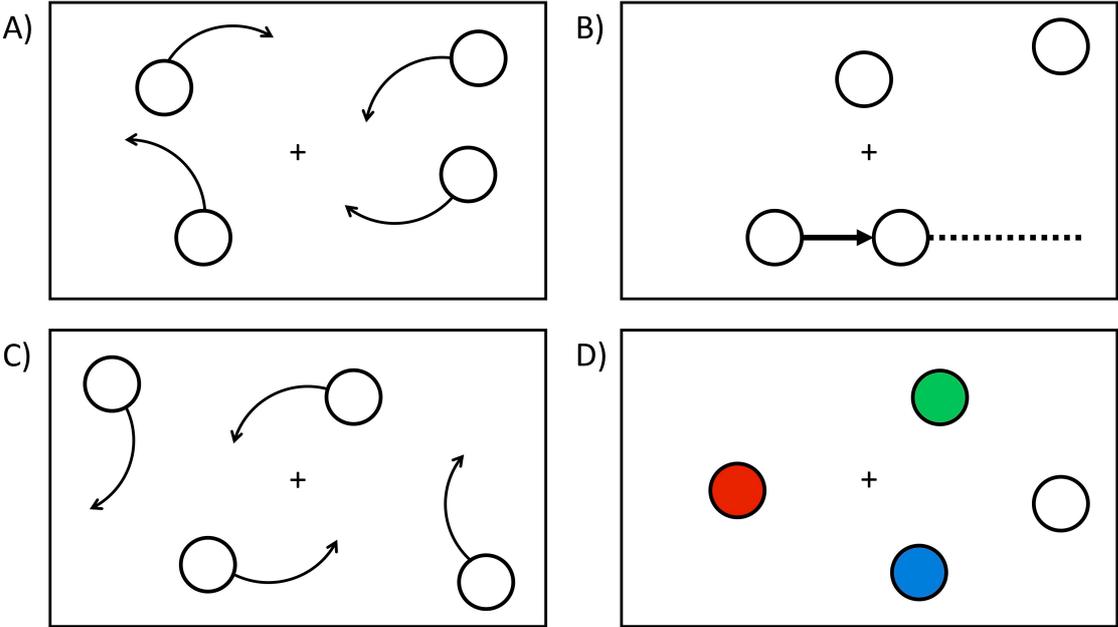


Figure 3. An illustration of the multiple push tracking task showing (a) random movement, (b) the push event, (c) random movement, and (d) test. The actual stimuli contained 9 circles and the arrows were not present.

If role-related features are maintained in the object tracking system, then it is possible that the capacity limitations of the system might influence

performance in this task. Based on previous findings using similar display parameters to the present stimuli (Alvarez & Franconeri, 2007), it was estimated that viewers would have an object tracking capacity of around four objects. Therefore, we varied the number of agent and patient targets using trials with one, two, or three push events. Scenes with two push actions require viewers to track four targets, whereas those with three pushes involve tracking six distinct objects (three agents, three patients), which is beyond the calculated tracking capacity and should be more difficult. Thus, it was predicted that role assignment accuracy would remain consistently above chance until the tracking capacity is surpassed, which was estimated to be two push events (or four objects).

To limit overt shifts of attention, we required participants to fix their gaze on a central cross as they completed the task, which was monitored using eye-tracking. However, it remained possible for the participants to shift their internal focus of attention to follow particular objects (Yantis, 1992). Hence, we analysed whether performance varied with the push event that was highlighted at test (the test push). For example, in an event with three pushes, the third push might be recalled more accurately due to recency. However, if the participant decides to focus solely on the first push rather than tracking all three events, then there would be a tendency for this to be the most accurate. A significant effect of the test push or an interaction with another factor would provide evidence that participants were adopting strategies to complete the task, rather than treating all of the objects equally.

Finally, we examined agent and patient accuracy separately to test for role related differences in performance, as previous work has shown that

viewers often preferentially attend to agents over patients in events (Cohn & Paczynski, 2013). However, others have found that participants sometimes apply a centroid grouping strategy to track target pairs (Fehd & Seiffert, 2008, 2010; Huff et al., 2010; Oksama & Hyönä, 2016; Zelinsky & Neider, 2008) and this would suggest that there will be no differences in accuracy between agents and patients, as these roles are defined relative to each other. Thus, if the participants perform above chance, then differences in accuracy between the agent and patient would provide additional information about how they completed the task.

## Method

### *Participants*

Participants were recruited from the undergraduate population of the University of Liverpool ( $N = 24$ ). All participants were required to be native English speakers with normal language and cognitive abilities, as well as normal or corrected-to-normal vision. The sample size was selected based on the results of an external pilot study, which indicated that a sample greater than 14 participants would provide sufficient power ( $\beta > 0.8$ ) to detect the effects in our analysis (see *Analysis* section). A larger sample was recruited to account for methodological adjustments made after the pilot and the potential need to exclude trials where the participants did not fixate their gaze.

### *Apparatus*

Eye movements were recorded using an EyeLink 1000 system at a sampling frequency of 500 Hz and saccade sensitivity set to high. The stimuli

were created using the Processing 3 programming language (<https://processing.org/>) and were presented on a 17" LED monitor with a screen resolution of 1280 × 1024 pixels and a 60-Hz refresh rate. The participants were positioned approximately 57cm in front of the display ( $\sim 37.6^\circ \times 30.2^\circ$ ) without a head restraint.

### *Stimuli*

The task consisted of animated display sequences in which nine identical objects moved randomly against a black background. These objects were white unfilled circles  $0.8^\circ$  in diameter. A red fixation cross ( $0.4^\circ \times 0.4^\circ$ ) was positioned in the centre of the display.

Each trial lasted 25 seconds. During the first three seconds, all nine circles moved randomly (figure 3A). Unique patterns of unpredictable motion for each circle were generated by an algorithm that reassigned the objects with a random direction within a  $120^\circ$  window approximately every 250 ms. The circles moved at a constant speed of  $6^\circ/\text{sec}$ . If objects were closer than  $4.2^\circ$  (centre to centre), their direction was changed so that they moved away. At these levels, expected tracking capacity is approximately four objects (Alvarez & Franconeri, 2007).

After 3 seconds of random movement, two of the objects were selected to be the agent and patient and engaged in a push event (figure 3B). These roles were assigned pseudo-randomly, with the algorithm only selecting objects that had not featured in previous pushes. The push event was an implementation of Michotte's (1946) launch effect; the agent travelled along a direct vector towards the patient, where, upon contact, the agent immediately stopped and the patient

moved away along the same vector and at the same velocity. During the push event, the other circles remained stationary. This entire sequence lasted approximately 3 seconds, following which, all nine objects reverted to random motion (figure 3C). For trials with two or three push events, the objects experienced a second and third push event respectively, with 1 second of random movement between each push.

After 25 seconds, object movement was terminated and three of the nine circles were highlighted in red, blue and green (figure 3D). Two of the coloured objects were the agent and patient of one of the push events, while the third was a foil randomly selected from the objects that had not been involved in any pushes.

### *Procedure*

The participants were guided through an example trial and were verbally instructed to track all the objects involved in all the push events, remembering the agent and patient of each push. They were also asked to fixate their gaze on the marker in the centre of the screen and were informed that this would be monitored by the eye-tracker. After being calibrated with 9-point calibration, the participants completed a total of 60 trials, with the opportunity to take breaks when needed. At the beginning of every trial, the word “READY” appeared in the centre of the screen, with the scene commencing once the participant fixated on the text for more than 3 seconds. When the agent, patient, and foil objects changed colour at the end of the trial, this prompted the participants to describe the interaction that occurred between two of the coloured circles on the screen. They were required to provide their description using an active transitive

structure, such as *red pushed blue*. The identified agent and patient (e.g., *red* and *blue*) were coded online by the experimenter, before advancing to the next trial. The participants' responses were also audio recorded and transcribed, which were used to verify the online coding. The final data showed whether their utterances had correct agents and patients and any errors that they made.

### *Analysis*

The dependent measure of the analysis was the mean correct labelling of the agent and patient for each participant in each condition. Linear mixed-effects models were applied using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in *R* version 3.5.1 (R Core Team, 2018). Since there were three test items, 0.3333 was subtracted from the dependent measure so that the intercept of the model encoded a comparison with chance. The fixed effects structure of the model consisted of three fully crossed variables: the number of push events (1/2/3) and the test event (1/2/3) as centred continuous predictors, with thematic role (agent/patient) as an effect coded fixed factor. The random effects structure of the model represented the maximal model (Barr, Levy, Scheepers, & Tily, 2013): subject was entered as a random intercept with the random slopes for the main effects and interactions of the three fixed predictors. If necessary, the random effects structure was simplified until model convergence was achieved, starting with the highest order interactions. The hypothesized effects of the fixed predictors were tested via likelihood-ratio ( $\chi^2$ ) comparisons through the sequential decomposition of the model. The marginal and conditional  $R^2$  statistics are also reported as effect sizes (Johnson, 2014; Nakagawa & Schielzeth, 2013; Nakagawa, Johnson, & Schielzeth, 2017). These provide

measures for assessing the goodness-of-fit of generalised linear mixed-effect models, representing the variance explained by the model with the random effect structure included (conditional  $R^2$ ) or excluded (marginal  $R^2$ ) from the calculation. Bootstrap resampling was applied to obtain 95% confidence intervals (CIs) and accurate p-values for all the model estimates and likelihood ratio tests ( $R = 1000$ ).

## Results

Consistent with the criteria applied in previous MOT work (e.g., Pylyshyn & Storm, 1988), trials were rejected when the participant fixated on an area more than  $2^\circ$  away from the central point during the random movement periods following the push events. This led to 12.5% of the trials being excluded from the analysis.

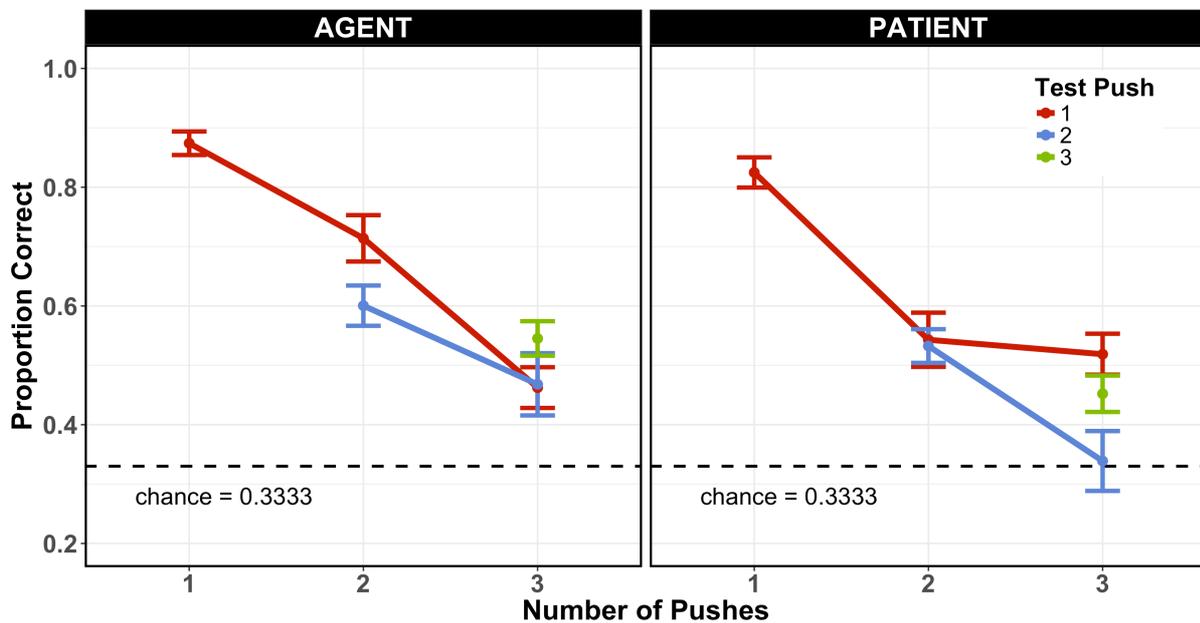


Figure 4. Role assignment accuracy in experiment 1 with error bars to show the standard error adjusted for the random effect structure of the mixed model.

The maximal model that converged contained subject as a random intercept, with random slopes for the main effects of the three predictors. As illustrated in figure 4, the mixed-effects model found that the participants' overall accuracy in assigning the agent and patient roles was significantly above chance (intercept  $\beta = 0.215$  [0.176, 0.2517],  $SE = 0.0193$ ,  $t = 11.12$ ,  $p < .001$ ). However, a significant negative effect of the number of pushes in the scene was also observed ( $\beta = -0.1348$  [-0.1961, -0.0727],  $SE = 0.0315$ ,  $\chi^2 = 32.59$ ,  $p = .001$ ). This effect of the number of pushes interacted with the test event ( $\beta = 0.0886$  [0.0247, 0.1543],  $SE = 0.033$ ,  $\chi^2 = 6.96$ ,  $p = .012$ ), as figure 4 shows that accuracy was highest for the first event in the two push trials but not in the three push trials. The results also revealed a significant agent advantage, with accuracy for the agent roles being 6.59% [2.63%, 10.47%] higher than for the patient roles ( $\beta = 0.0317$  [9e-04, 0.0614],  $SE = 0.0154$ ,  $\chi^2 = 10.42$ ,  $p = .001$ ). Overall, the model accounted for 29.52% of the variance in the data without the random effect structure and 46.39% of the variance when it was included ( $R_m^2 = 0.2952$ ,  $R_c^2 = 0.4639$ ).

Surprisingly, we found that participants were able to track multiple identical circles while fixating their gaze in a central position and identify agents and patients at above chance levels in this challenging task. Since the circles were identical, the participants were required to retain their role-related features as they moved around randomly. Interestingly, we found that accuracy decreased with additional pushes, which suggests that attention may have been involved in maintaining the push information and was taxed by additional pushes. The interaction of test event with the number of pushes also suggests that different attentional strategies were applied depending on how many

pushes occurred. For trials with two events, the participants would still track the first push, even after seeing the second. However, when a third push occurred, they switched some of their focus to this new event at the expense of tracking earlier pushes. Thus, one possibility is that the above chance performance overall was due to a combination of an object tracking system with a capacity of around 4 items (for the present stimuli) combined with attentional strategies to support the more difficult trials, particularly those with three pushes. Since tracking accuracy was high, even for three push events, we decided to tax attention with a concurrent distraction task to clarify the contribution of the object tracking system in the maintenance of role information.

## **Experiment Two: Distraction Task**

Experiment one demonstrated that observers can track the agents and patients of multiple push events while fixating their gaze on the centre of the display. The results also suggested that the participants may have used covert shifts of attention to support their tracking. To attenuate such covert switching, we had participants complete the same multiple push tracking paradigm as experiment one, while also performing a secondary task to capture their focal attention. It has been found that participants are effectively blind to many aspects of their visual surroundings when engaged in specific activities (Drew, Vö, & Wolfe, 2013; Hyman, Boss, Wise, McKenzie, & Caggiano, 2009; Mack & Rock, 1998; Simons, 2010; Simons & Chabris, 1999; Ward & Scholl, 2015). Thus, our second experiment examined whether responding to a colour change in the centre of the display interferes with the maintenance of thematic role

features like causality, or whether these features can be sustained without overt attention.

There is a large body of evidence showing that MOT performance is attention-sensitive, as a reduction in object tracking abilities has been observed when participants must also engage in a concurrent task (e.g., auditory tone monitoring, telephone conversations, finger tapping, or visual/verbal category judgements; Allen et al., 2006, 2004; Kunar, Carter, Cohen, & Horowitz, 2008; Tombu & Seiffert, 2008; Trick, Guindon, & Vallis, 2006). The effects of such tasks have shown to mirror changes that variation in speed or proximity can have on tracking performance (Alvarez & Franconeri, 2007; Tombu & Seiffert, 2008), demonstrating that object tracking itself has an attentional component (Cavanagh & Alvarez, 2005). Furthermore, there is some evidence that even “pop-out” features like colour or shape are more often noticed for tracked objects than distractors (Alvarez & Cavanagh, 2005; Tran & Hoffman, 2016), so there may not be a clear distinction between automatic and attention-dependent features in MOT tasks. However, it is well-established that location tracking can be carried out in parallel without the need to focally attend to the target objects (e.g., Oksama & Hyönä, 2016; Pylyshyn & Storm, 1988) so we selected a distraction task in which success depends on focal attention. The participants provided a speeded response (via key-press) whenever a static cross in the centre of the display changed colour, following evidence that viewers will often miss coloured objects travelling past their fixation point when attending to moving objects elsewhere in the display (Most et al., 2001). The colour changes occurred randomly and frequently, so success in this secondary task required continuous attention.

This study is similar to the first experiment, except a concurrent speeded-response task was used to occupy attention and eye-tracking was not performed. If role-tracking remains above chance, then it would suggest that the object tracking system is maintaining role-related features. Whereas, if the participants are unable to track multiple agent and patient roles while simultaneously responding to the distraction task, it would suggest that serial attention is critical to the maintenance of role information.

## **Method**

### *Participants*

An additional group of undergraduate participants ( $N = 24$ ) were recruited.

### *Apparatus*

The study used animated display sequences that were designed and presented using the Processing 3 programming language (<https://processing.org/>) and shown in fullscreen on a high-resolution monitor ( $2880 \times 1800$ ;  $\sim 36.5^\circ \times 23.2^\circ$  visual angle).

### *Stimuli*

The stimuli involved the same multiple push tracking paradigm used in experiment one (see figure 3), with the addition of a simple distraction task (see *Procedure*).

## *Procedure*

The study followed the same overall procedure as experiment one, with two key differences. First, participants' gaze was not monitored with an eye-tracker. Second, the fixation cross in the centre of the display would switch colours between blue and pink during the random movement parts of the trial. These colour changes occurred at random intervals every 1-2 seconds ( $M = 11.44 \pm 2$  changes per trial) and never occurred during the push events. The participants were instructed to respond to the colour changes in the distraction task as fast as possible via a key press.

## **Results**

Performance on the colour change task was considered accurate for a given trial if average response time to the changes was less than 1 second, consistent with other object tracking studies using a speeded-response task (Tombu & Seiffert, 2008). Based on this criterion, 15.83% of the trials were excluded from the analysis.

The maximal model that converged included random slopes for the main effects of the pushes, test event, and role variables. It revealed that the participants' overall accuracy in assigning agent and patient roles was above chance (intercept  $\beta = 0.1013$  [0.059, 0.1411],  $SE = 0.021$ ,  $t = 4.83$ ,  $p < .001$ ). However, there was a significant decline in response accuracy as the number of pushes increased ( $\beta = -0.0666$  [-0.1216, -0.0101],  $SE = 0.0285$ ,  $\chi^2 = 16.32$ ,  $p = .001$ ). A significant interaction between the number of pushes and the test event variables also occurred ( $\beta = 0.1257$  [0.0558, 0.1976],  $SE = 0.0362$ ,  $\chi^2 = 11.6$ ,  $p = .002$ ), due to accuracy being highest for the final event of the three push trials,

but no event preferences in the two push trials. The analysis also showed a significant agent advantage, with accuracy for agent roles being 6.02% [2.25%, 9.82%] higher than patient roles ( $\beta = 0.0282$  [-0.0094, 0.0641],  $SE = 0.0188$ ,  $\chi^2 = 4.77$ ,  $p = .023$ ). This mixed-effects model accounted for 14.62% of the variance without the random effect structure and 32.78% when it was included ( $R_m^2 = 0.1462$ ,  $R_c^2 = 0.3278$ ).

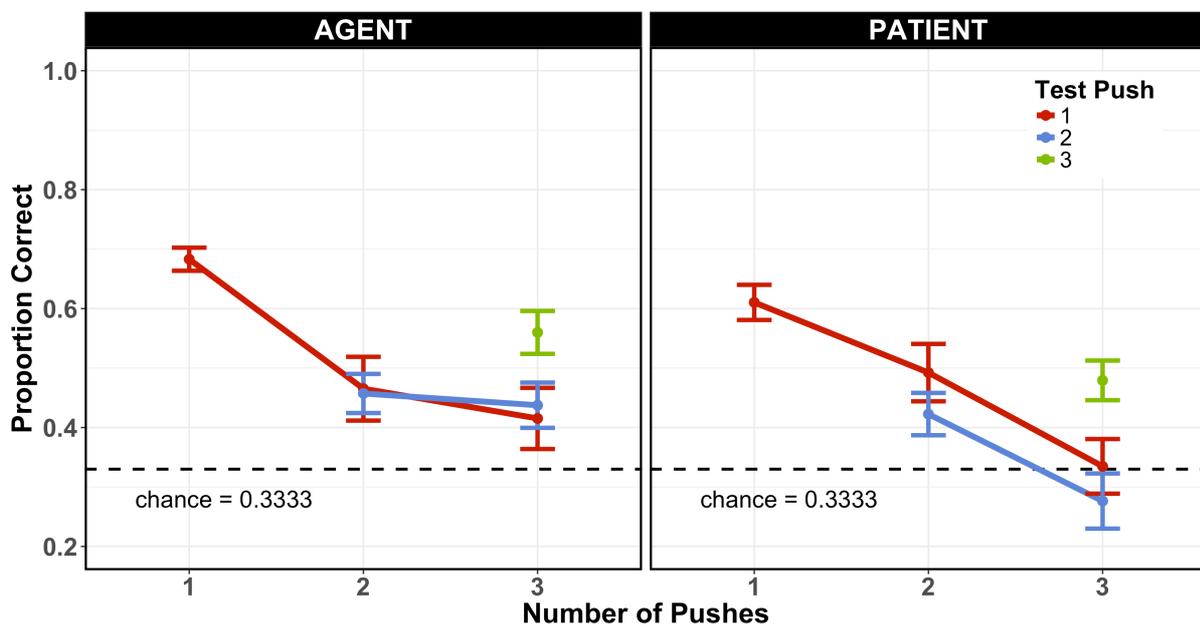


Figure 5. Role assignment accuracy in exp. 2 with error bars to show the standard error adjusted for the random effect structure of the mixed model.

Experiment two showed that participants can accurately track push events at above chance levels while simultaneously responding to a distraction task. The ability to track the agents and patients degraded with additional pushes, which suggests that some attentional processing was being used for tracking. Consistent with this, an interaction between the number of pushes and test event was found, with participants appearing to focus their attention on the last event in the difficult three push trials rather than following all six targets. This

suggests that the ability to identify agents and patients was not strongly impacted by the introduction of a distraction task and thus may involve features that can be tracked in parallel with objects pointers.

### **Experiment Three: Temporarily Invisible Objects Task**

A key assumption of the object-tracking hypothesis tested in the present research is that the visual system contains object pointers that can track the location of several objects in the visual world. A strong source of evidence for the existence of these pointers comes from MOT studies reporting that participants can track multiple targets even when they are temporarily occluded or invisible (Alvarez & Cavanagh, 2005; Flombaum, Scholl, & Pylyshyn, 2008; Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006; Scholl & Pylyshyn, 1999). For example, Scholl & Pylyshyn (1999) observed that the ability to track multiple randomly moving objects was unaffected by having the items travel behind occluding surfaces that completely concealed them. This ability appears early in development (e.g., 12 months; Spelke, Kestenbaum, Simons, & Wein, 1995), suggesting that the object tracking system is inherently capable of dealing with occlusion.

Tracking objects during occlusion is thought to involve simple distance-based heuristics (Fencsik, Klieger, & Horowitz, 2007; Franconeri, Pylyshyn, & Scholl, 2012; Keane & Pylyshyn, 2006) and motion features like velocity (Fencsik et al., 2007; Howe & Holcombe, 2012; Iordanescu, Grabowecky, & Suzuki, 2009; Luu & Howe, 2015). While these cues may be similar to those used for role identification, one important difference is that these features are used to support individual objects during occlusion, but causal interactions and

chasing often involves identifying a relationship between multiple objects (e.g., the velocity of the agent in relation to the patient). It is possible that participants can track motion features across occlusion but lose track of the relationships between objects. This was tested in experiment three by removing the avoidance constraint in the stimuli and instead allowing the circles to simply pass through each other during the periods of random movement. Whenever this happened, both circles would temporarily (<500ms) vanish. Since tracking across occlusion is a critical ability of the object tracking system, this study tests whether relational features between occluded objects are maintained.

## **Method**

### *Participants*

An additional group of undergraduate participants ( $N = 18$ ) were recruited.

### *Stimuli*

The stimuli used the same core multiple push tracking task as experiment one (see figure 2), with three important changes. First, the randomness of the objects' movement patterns was greatly reduced, with direction changes occurring much less frequently (approximately every 1000ms). Second, there was no restriction on the distance between objects, so objects could pass through each other. When the distance between two (or more) objects was less than  $2.5^\circ$  (centre to centre), the colour of those circles would change to black, effectively making them invisible. These blackouts only occurred during the periods of random movement and would never last more than 500 milliseconds

at a time. Although the movement patterns were randomly generated by the experimental program, each target object was occluded for approximately 0.8 seconds per trial on average ( $M = 836.89\text{ms} \pm 477.04\text{ms}$ ).

### *Procedure*

The procedure was identical to experiment two, with the exception that the participants did not have to respond to a distraction task with a keypress, but instead only had to keep track of the agent and patient in the push events.

## **Results**

The maximal model that converged included random slopes for the main effects of the number of pushes, test event, and role variables (but not their interactions). It revealed that participants' overall accuracy in assigning thematic roles was significantly above chance (intercept  $\beta = 0.0647$  [0.0251, 0.104],  $SE = 0.0201$ ,  $t = 3.21$ ,  $p = .001$ ). However, there was a significant decrease in accuracy as the participants were required to keep track of additional push events ( $\beta = -0.0181$  [-0.0749, 0.0368],  $SE = 0.0285$ ,  $\chi^2 = 5.52$ ,  $p = .012$ ). The test event variable yielded no notable effects: neither its main effect ( $\beta = -0.028$  [-0.0846, 0.0307],  $SE = 0.0294$ ,  $\chi^2 = 1.18$ ,  $p = .290$ ) nor interaction with the number of pushes approached significance ( $\beta = 0.0187$  [-0.0502, 0.0853],  $SE = 0.0217$ ,  $\chi^2 = 0.31$ ,  $p = .588$ ). A significant agent advantage was observed, with accuracy for agent roles being 5.28% [1.21%, 9.33%] higher than patient roles ( $\beta = 0.0243$  [-0.0061, 0.0543],  $SE = 0.0154$ ,  $\chi^2 = 4.88$ ,  $p = .023$ ). This maximal model with pushes, test event, and role as fixed effects accounted for 4.69% of

the variance without the random effect structure and 22.94% with it included ( $R_m^2 = 0.0469$ ,  $R_c^2 = 0.2294$ ).

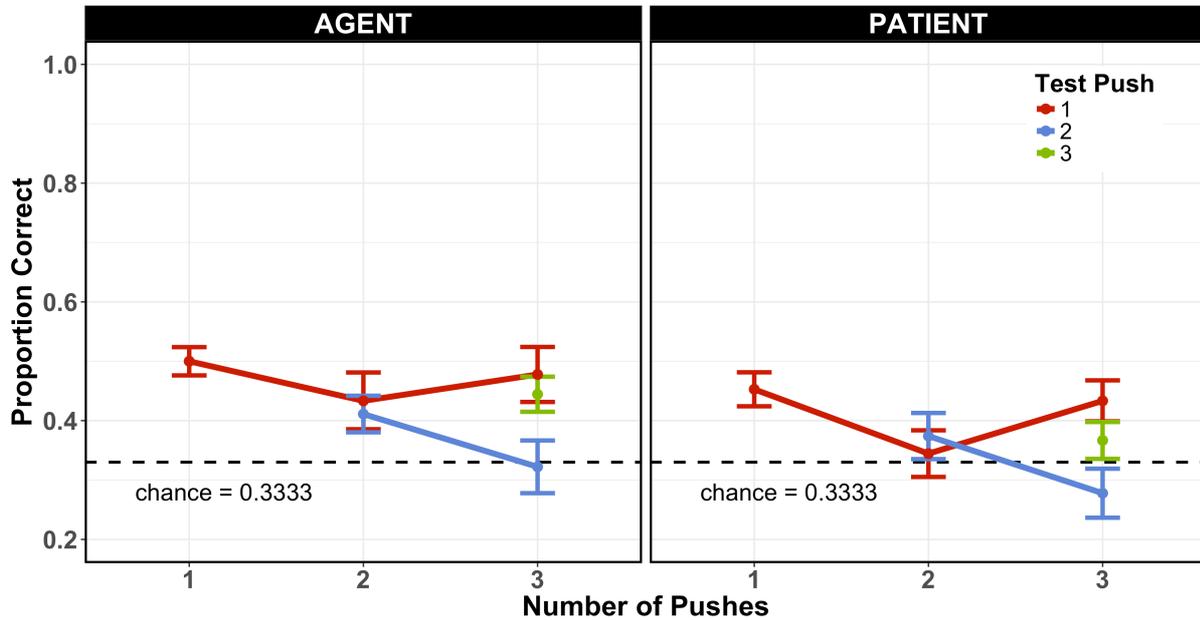


Figure 6. Role assignment accuracy in exp. 3 with error bars to show the standard error adjusted for the random effect structure of the mixed model.

Although overall accuracy in experiment three was lower than in the previous two studies, the participants were still able to identify agents and patients at above chance levels overall. While there was a negative effect of the number of push events, there was no main effect or interaction with test event. This suggests that the challenge of tracking across occlusions made it less beneficial to apply specific strategies for trials with multiple push events. The negative effect of test push can be explained if we assume that tracking was occasionally lost when occlusion occurred, and the more pushes that must be tracked, the more likely that these losses of tracking will affect one of the objects in the push events.

## Combined Analysis

To compare performance across the three tasks, the data from all three studies were combined into one analysis. The same mixed-effects model was fitted to the data, with the addition of experiment (one/two/three) as a Helmert coded factor. The maximal model that converged included random slopes for the main effects of the pushes, test event, and role variables. An overall agent advantage of 6% [3.59%, 8.41%] was observed in the combined data ( $\beta = 0.0271$  [0.007, 0.0477],  $SE = 0.0104$ ,  $\chi^2 = 20.21$ ,  $p = .001$ ). There were significant differences in performance between the three studies. Specifically, accuracy was significantly higher in the first experiment that used eye-tracking ( $M = 63.06\%$ ,  $SE = 3.53\%$ ) than in the second study ( $M = 52.1\%$ ,  $SE = 3.92\%$ ) which involved a distraction task ( $\beta = 0.0568$  [0.0278, 0.0853],  $SE = 0.0147$ ,  $\chi^2 = 6.69$ ,  $p = .013$ ). Compared to these two studies combined, accuracy was significantly lower in the third experiment ( $M = 41.99\%$ ,  $SE = 3.64\%$ ) where the objects became temporarily invisible ( $\beta = -0.0234$  [-0.037, -0.0102],  $SE = 0.0068$ ,  $\chi^2 = 21.49$ ,  $p < .001$ ). The analysis confirmed the negative main effect of the number of pushes ( $\beta = -0.0594$  [-0.0984, 0.0231],  $SE = 0.0192$ ,  $\chi^2 = 40.36$ ,  $p < .001$ ) and showed a significant interaction with experiment; a steeper negative slope was observed for the number of pushes variable in experiment one than in experiment two ( $\beta = -0.0341$  [-0.0749, 0.0043],  $SE = 0.0202$ ,  $\chi^2 = 9.67$ ,  $p = .002$ ), while the negative slope observed in experiment three was significantly flatter than in the first two experiments ( $\beta = 0.0206$  [0.0016, 0.0399],  $SE = 0.0098$ ,  $\chi^2 = 16.71$ ,  $p < .001$ ). Although, there was no main effect for the test event variable ( $\beta = -0.0288$  [-0.0709, 0.0141],  $SE = 0.0217$ ,  $\chi^2 = 0.23$ ,  $p = .637$ ), there was a significant interaction between the number of pushes and the

test event ( $\beta = 0.0629$  [0.0175, 0.1083],  $SE = 0.0231$ ,  $\chi^2 = 16.17$ ,  $p < .001$ ).

Collectively, this combined analysis demonstrates that the negative effect of the number of pushes differed between the experiments, but similar tracking strategies were applied to deal with difficult trials.

## **General Discussion**

Linguistic accounts of thematic roles often focus on how to identify roles when dealing with entities that differ in sentience, volition, animacy, and other semantic features (Dowty, 1991; Fillmore, 1967; Jackendoff, 1972, 1987; McRae et al., 1997). Consequently, these theories do not provide a clear account of how multiple agents and patients in the same kind of event (e.g., pushes) can be distinguished among visually identical referents. Studies of multiple object tracking show that viewers can monitor a small number of objects in parallel (e.g., Pylyshyn & Storm, 1988) and track a role-related interaction among these objects (e.g., Gao et al., 2009), but since there is only one interaction, it is possible that the whole event is used to activate role-related features that support behavioural responding. In the three studies presented here, there were multiple agents and patients, so correct responding depended on knowing about the particular binding between individual circles and their role in each push event. Therefore, the object tracking system needed to maintain the agent and patient-related features for different circles as they moved around randomly. The findings indicate that people can maintain some of these bindings for multiple objects, as accuracy in describing the interactions was above chance, even though the participants did not know which push event would be tested. This suggests that the linguistic system used to map thematic role fillers into

sentence positions can query relational role features available in the visual object tracking system.

The findings of all three experiments showed that they were able to identify both agents and patients, despite each study using variations of the task that taxed their ability to track the objects in different ways. In the first experiment, the participants' gaze position was monitored with an eye-tracker and they were instructed to fixate on a central point. However, as this does not control for covert shifts of attention, a second study was performed where the participants responded to a concurrent distraction task. Overall accuracy in this study was significantly lower than in the first study, congruent with earlier reports that object tracking is hindered when other attentionally demanding tasks are performed simultaneously (Tombu & Seiffert, 2008). Finally, to provide a stronger test that object tracking is involved in maintaining thematic roles, a third experiment had the objects disappear when in close proximity to each other, forcing the participants to track the motion of momentarily invisible objects in order to ensure continuity of tracking. Relative to the first two studies, accuracy in the third study was reduced. This can be explained by experiments showing that motion extrapolation can only be carried out for two objects at a time (Fencsik et al., 2007; Howe & Holcombe, 2012; Luu & Howe, 2015), suggesting that the participants were limited to tracking only one push event in this version of the task.

There was a negative effect of the number of pushes in all three studies, indicating that the participants found it harder to track role-related information as the number of pushes increased. For difficult trials with three pushes, the participants appeared to apply attentional strategies rather than tracking all six

target objects in parallel, consistent with the two push (or four objects) tracking capacity estimated for the present stimuli (Alvarez & Franconeri, 2007). They often shifted their attention to the final event, which is a logical strategy since this push needs to be retained for the least amount of time before test and is therefore less likely to be lost during tracking.

Across the three studies, accuracy in identifying agents was superior to that of patients, with an average agent advantage of 6%. Since agents and patients were defined by the same push action, it is not entirely clear why agents are preferred. Many of the proposed features thought to trigger the perception of causality in launch events, such as an immediate response to direct contact (Leslie & Keeble, 1987; Oakes & Cohen, 1990), should be equally useful in identifying both agents and patients. Since agents were produced before patients at test, it is possible that a temporal advantage is the basis for the agent bias. However, the length of the temporal delay between the push event and test was not a reliable predictor of performance. For instance, accuracy was always highest in trials containing only one push, but less time elapsed between the last push and the test event in trials with three pushes. Another explanation is that the participants adopted tracking strategies that created a bias for agents. However, in all three experiments, the agent advantage did not interact with any other variable, suggesting that it is not something that changes with the other aspects of the task. Instead, it is possible that the agent advantage originated in the role-related features being tracked with the targets. Causality and agency can be perceived in the relationships between objects (e.g., chasing; Gao et al., 2009) and from their individual properties (e.g., unprovoked changes in speed and direction; Tremoulet & Feldman, 2000). However, these features may not be

equally distributed among the objects and could create attentional biases. For example, Gao et al. found that chasing subtlety (i.e., the agent's angle of motion relative to the most direct path to the patient) could explain the identification of chasing in their study, suggesting that chasing is primarily defined by the attributes of the agent rather existing in a perceptually symmetrical relationship with the patient (see also Cohn & Paczynski, 2013). Similarly, the push actions in our task involved a self-propelled agent moving towards a static patient, and developmental research has shown that self-propulsion is an important cue for detecting agency (e.g., Luo & Baillargeon, 2005; Luo & Johnson, 2009). Therefore, the availability and salience of relational and non-relational thematic role cues might produce biases for particular roles and could explain the consistent preference for agents in this work. Further research is needed to determine the exact features that create the linguistic agent advantage, but these findings suggest that the object tracking system is involved.

In summary, the present study demonstrates that the thematic roles of multiple objects can be tracked across periods of random motion, even under difficult test settings. The simplest explanation is that the linguistic processes of the language production system can directly query perceptual features in the object tracking system to support the mapping of agent and patient into appropriate sentence positions. This suggests that linguistic meaning is not a separate system that encodes abstract features, such as volition and sentience, but rather, it may be grounded in the visual perception of scenes. This idea revisits an older approach where visual processing played a role in production (e.g., Osgood, 1952; Osgood & Bock, 1977). However, it suggests a new paradigm for studying linguistic meaning, where messages are presented in

formats compatible with perception studies, and visual variables can be manipulated to test the degree that language production is sensitive to the machinery in the visual system.

### **Additional Information**

Chapter two has been accepted for publication with revisions by *The Quarterly Journal of Experimental Psychology*<sup>1</sup>. I was responsible for all aspects of the design and implementation of this work, under the supervision of Dr. Franklin Chang as my primary academic advisor. There were no other significant contributors to this research.

---

<sup>1</sup> Jessop, A. & Chang, F. (2018). Thematic role information is maintained in the visual object tracking system. [under revision]

## CHAPTER THREE: THEMATIC ROLE TRACKING IN RELATIVE CLAUSES

### Fit Within the Thesis

Chapter two presented the results from three experiments using a multiple push tracking task. These studies consistently found that adult participants can track agents and patients among visually identical distractors to produce descriptions that identify these referents and their thematic roles. By manipulating the demands of the task, these studies also provided evidence that speakers can use perceptual features extracted via the object tracking system to support the mapping of agent and patient referents into appropriate sentence positions. Chapter three extends these findings to an important issue in language processing.

Across many different languages, subject-extracted relative clauses (SRCs) appear to be easier to process and understand than object-extracted relative clauses (ORCs). Since the events corresponding to these sentences differ in their thematic role consistency, chapter three examined whether it is easier to track the agent and patient bindings of push interactions matching SRCs than ORCs. Six studies applied the same multiple push tracking task used in chapter two, but there were always two pushes in the trials and an overlapping circle that appeared in both events. In the SRC conditions, the overlapping object played the agent in both pushes, whereas in the ORC conditions, the patient of the first push was the agent of the second. In three experiments, production accuracy was found to be higher for SRC than the ORC events when they were described using an either SRC/ORC (exp. 4), an SRC/passive relative (exp. 5), or active transitive sentences (exp. 6). These results suggest that role-referent consistency creates a production bias for SRC

events, since a significant difference in accuracy was repeatedly observed in a tracking task that controlled for others cues, even when relative clause structures were not used.

Three additional experiments were conducted to further examine the effects of thematic role consistency on production accuracy. Here, the participants used transitive sentences to describe scenes in which a target agent or patient played the same role in both pushes, or switched to the alternative role between the two events. These studies observed consistently higher production accuracy when the overlapping target played the same role in both events compared to when they switched roles or only appeared in one of the pushes. This was demonstrated when the trials were described using active (exp. 7-8) and passive transitive sentences (exp. 9). This provides strong evidence that role consistency in visual events can create biases in sentence production, as a significant difference in accuracy was consistently observed for a range of event configurations and various language structures.

Within the broader context of this thesis, chapter three provides converging evidence from six experiments to show that the respective difficulty in tracking the thematic roles of visually identical objects can directly affect sentence production and the ability to map referents into appropriate sentence positions. This is consistent with the theory that thematic roles in language may arise from spatial processing mechanisms, while also offering a potential explanation for the persistent SRC bias that has been observed across many different languages, even in those where linguistic theories would predict superior processing for ORCs (see Vasishth, Chen, Li, & Guo, 2013).

Throughout this chapter, the previous experiments presented in chapter two are referenced as Jessop and Chang (2018). Since I have opted for the alternative thesis format, I have decided to preserve these citations in their current form, rather than replacing them with references to chapter two. The earlier studies are summarised when necessary, as this chapter was written to be a self-contained manuscript, rather than as part of the continuous narrative of this thesis.

## **Introduction**

Many studies have found that subject-extracted (1) relative clauses are easier to process and understand than object-extracted (2) relative clauses (for a review, see Gordon & Lowder, 2012).

1. The reporter that attacked the senator admitted the error [SRC].
2. The reporter that the senator attacked admitted the error [ORC].

A typical experiment of relative clause processing involves presenting participants with a series of sentences containing either an SRC or an ORC and measuring their reading times or subsequently testing their comprehension with probe questions (Gordon, Hendrick, & Johnson, 2001; Just & Carpenter, 1992; King & Just, 1991; MacWhinney & Pleh, 1988; Mak et al., 2002; Traxler et al., 2002; Wanner & Maratsos, 1978; Warren & Gibson, 2002). Studies using this approach have observed superior processing for SRC sentences (such as 1) compared to ORC sentences (such as 2) across a range of postnominal (English: King & Just, 1991; Dutch: Mak et al., 2002) and prenominal languages (Korean: Kwon et al., 2010; Japanese: Miyamoto & Nakamura, 2003). While

early attempts to explain this asymmetry focused on syntactic differences and the varying memory demands they impose (Gibson, 1998; Gordon et al., 2001; Just & Carpenter, 1992), a plethora of evidence suggests that these biases are also influenced by experience and the frequency that the structures appear in natural language (e.g., MacDonald & Christiansen, 2002; Reali & Christiansen, 2007; Wells et al., 2009). Consequently, many mechanisms have been proposed to explain how the asymmetry arises within memory-based or experience-based processing architectures. However, the purpose of maintaining an ORC structure that is more difficult to comprehend is unclear, especially since these models argue that the language system could modify the grammar or the input distribution of these structures to make them easier to process. In the present work, we will suggest that the persistent difficulty with the ORC arises not from language processing, but from the greater complexity in tracking thematic roles in non-linguistic events. Before describing this account, we will first review some of the existing theories of the relative clause asymmetry in more detail.

One explanation of the relative clause asymmetry is that ORC structures demand more cognitive resources to process correctly. This view is supported by experiments demonstrating that the amount of working memory available while processing the relative clause verb region of the sentence – either through experimental manipulation or individual differences – predicts the magnitude of the SRC-ORC asymmetry (Gordon & Hendrick, 2005; Just & Carpenter, 1992; King & Just, 1991; Wanner & Maratsos, 1978). A prominent memory-based account is the dependence-locality theory (Gibson, 1998; Grodner & Gibson, 2005; Warren & Gibson, 2002), which argues that the relative clause asymmetry

is predominantly driven by the distance and the number of intervening dependent referents between the head noun phrase (NP) and the relative clause verb (or more precisely, the filler and gap). From this perspective, processing ORCs requires both the head NP (e.g., the reporter) and the embedded NP (e.g., the senator) to be held in working memory before they can be integrated with the verb phrase (e.g., attacked). However, in an SRC structure, the head NP can be attached to the verb before the embedded NP is encountered, making them easier to comprehend. Interestingly, memory-based models suggest that the syntactic arrangement of relative clauses could be adjusted to neutralise the asymmetry or introduce a bias in favour of ORCs. Support for this prediction comes from passive relatives in English, which are semantically identical to ORCs in terms of how thematic roles are assigned (i.e., *the reporter* is the *patient* and *the senator* is the *agent*), but have a shorter filler-gap distance and no intervening referents (see 3).

3. The reporter that *was attacked by* the senator admitted the error [Passive Relative].

Experiments have reported greater comprehension accuracy and faster reading times for passive relatives than ORCs (e.g., Gennari & MacDonald, 2008), suggesting they are a less demanding sentence structure. Thus, one way that the language system could remove the comprehension difficulty associated with the ORC is for the grammar to stipulate that the passive relative be used instead. If the asymmetry is due to cognitive resources or memory, it is not clear

why the language system does not change in some way to reduce resource expenditure.

An alternative perspective is that the level of experience speakers have with particular forms determines how difficult they are to process. Analyses of English corpora have reported differences in the frequency that SRCs and ORCs appear in natural language (e.g., Gordon & Hendrick, 2005; Roland, Dick, & Elman, 2007) and speakers appear to have less difficulty with ORCs in the contexts that they most frequently appear. For instance, Reali and Christiansen (2007) observed faster reading times for SRCs than ORCs when the embedded NP was an impersonal pronoun (e.g., *it*), but slower times when it was a personal pronoun (e.g., *I, you, she*). A comparison with corpus frequencies showed that embedded personal pronouns were more common in ORCs, whereas impersonal pronouns appeared more frequently in SRCs. Therefore, many have argued that the reason ORCs are often difficult to understand is because they are rare in natural language and only appear in specific contexts. Although various mechanisms have been proposed, the global claim of these experience-based models is that comprehenders possess linguistic representations reflecting the statistics of their language that help identify the potential meaning of an incoming utterance as it incrementally unfolds (Fitz et al., 2011; Gennari & MacDonald, 2008; Hale, 2001; Levy, 2008; MacDonald, 2013; MacDonald & Christiansen, 2002). Work with humans and connectionist models have demonstrated that more experience with ORCs, or even the sub-parts of these structures, can lead to a reduction in the observed asymmetry (Fitz et al., 2011; MacDonald & Christiansen, 2002; Wells et al., 2009). Thus, it might be possible to manipulate verb bias, topicalization frequencies, or the

frequency of similar structures to reduce this difficulty. If the asymmetry is due to input experience, then it is unclear why the distribution of structures does not change to make communication easier.

Empirical comparisons suggest that memory-based and experience-based theories provide distinct but complementary models that explain different aspects of relative clause comprehension behaviour (see Levy & Gibson, 2013; Staub, 2010; Staub, Dillon, & Clifton, 2017). Importantly, both frameworks predict variance in the asymmetry between languages based on frequency distributions or syntactic properties, with neither precluding the possibility that ORCs will be easier to process than SRCs in certain languages. While such instances may exist (e.g., Basque: Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía, & Laka, 2010), the majority of cross-linguistic work has found that ORCs are harder to comprehend and produce than SRCs across a diverse spectrum of languages (e.g., Korean: Kwon et al., 2010; Dutch: Mak et al., 2002; Chamorro: Wagers, Borja, & Chung, 2018). Thus, although the memory-based or experience-based theories can explain many of the behavioural patterns observed with relative clauses, these models would predict more variation in the SRC advantage associated with the grammar and linguistic distributions of different languages.

To explain why the relative clause asymmetry seems to be insensitive to language properties and persists in the face of processing difficulties, we argue that there is a non-linguistic bias for SRC structures in tracking thematic roles. In the ORC sentence below (4), the boy is the *patient* (i.e., the entity affected by the action) being pushed by the girl in the relative clause, but then switches to being the *agent* (i.e., the entity carrying out the action) that chases the dog in the

main clause. Whereas in a matched SRC sentence (5), the boy is consistently the *agent* that both pushes the girl and then chases the dog.

4. The boy the girl pushed chased the dog [ORC]
5. The boy that pushed the girl chased the dog [SRC]

Early linguistic theories suggested that such switches in thematic roles or perspective makes ORCs harder to understand (MacWhinney & Pleh, 1988; Sheldon, 1974). Although these switches cannot account for all relative clause behaviour (Traxler et al., 2002), they highlight an important divergence in the complexity of the interactions described in these two relative clause forms. The role-referent bindings in ORCs (AGENT-boy1 and PATIENT-boy1) are more complicated than for SRC sentences (AGENT-boy1). It is possible that binding two roles to one referent in ORC events can result in the memory for these interactions becoming distorted and more difficult to retain. Since event processing is non-linguistic (see Papafragou, 2015), this account offers an explanation of why the SRC bias exists across a diverse range of languages and why opting for less demanding word orders would not be sufficient to neutralise the bias. Furthermore, difficulties in understanding and producing ORC events would reduce the frequency of these structures in natural language, promoting the distributional biases for SRC structures highlighted by experience-based accounts. Therefore, a non-linguistic preference for SRC events could help to explain a range of findings in the sentence processing literature.

Relative clauses are often used to provide additional information to help characterise and identify the head NP (Fox & Thompson, 1990), which is particularly important when the intended referent is potentially ambiguous

(Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). For the examples given above in 4 and 5, if there was only one boy referent in the context (e.g., boy1), then the main clause alone would be sufficient to describe the event (e.g., the boy chased the dog). However, a relative clause can be included when there is another entity (e.g., boy2) that might be mistaken as the intended referent. To distinguish between these candidates, an earlier event that one of the referents uniquely participated in could be given, such as the SRC in 5 (e.g., the boy pushed the girl) or the ORC in 4 (e.g., the girl pushed the boy). However, a non-linguistic understanding of these event sequences is required before they can be described in language. Specifically, the observers need to track the referents (e.g., boy1, boy2, girl1, and dog1) throughout their various interactions, while identifying and remembering the agent and patient of each event. Multiple object tracking (MOT) tasks (Pylyshyn & Storm, 1988) offer a controlled method of assessing this role tracking ability, as all the possible referents are visually identical objects without any conceptual or typicality information that could support memory for their roles (McRae et al., 1997). Studies using MOT tasks have found that participants can follow the location of several objects in visual scenes as they move in unpredictable patterns (e.g., Alvarez & Franconeri, 2007; Oksama & Hyönä, 2016; Pylyshyn & Storm, 1988). By tracking such movement, observers also appear to automatically extract information that allows them to recognize the roles of the objects in different interactions (Barrett et al., 2005; Heider & Simmel, 1944; Michotte, 1946; Scholl & Tremoulet, 2000). For example, Gao, Newman, and Scholl (2009) showed participants scenes where multiple circles moved in a random manner, except for a particular *wolf* circle that was chasing a *sheep* circle around the screen. The observers were

readily able to detect the wolf whenever it moved in a fairly direct manner towards the sheep, showing that they can identify the agent of a chasing event from movement patterns alone (see also Dittrich & Lea, 1994; Gao & Scholl, 2011). The current work similarly adapted the MOT paradigm to examine how role-referent bindings are maintained in the visual system.

To study thematic role tracking in relative clauses, the present research used visual sequences containing two separate causal push events. Unlike a chase that could be interrupted and resumed, or continue indefinitely, push events are discrete in time and can be configured to have a clear correspondence with SRC or ORC sentences. The perceptual significance of pushing events (and many other interactions) was originally demonstrated by Michotte (1946) and has since become one of the most widely studied and robust instances of causality in the perception literature (see Rips, 2011; Scholl & Tremoulet, 2000). The standard pushing (or launching) display involves one object (e.g., square A) moving directly towards a second stationary object (e.g., square B) and stopping when it makes physical contact. If square B immediately moves away along the same vector upon contact, then observers will interpret the sequence as square A *pushing* square B and causing it to move (e.g., Schlottmann et al., 2006). Critically, infants younger than 12 months can also differentiate causal pushing events from non-causal events with similar movement configurations (e.g., Cohen & Amsel, 1998; Leslie & Keeble, 1987; Oakes, 1994; Oakes & Cohen, 1990). For instance, Oakes and Cohen (1990) habituated infants to either causal pushes or matched non-causal events containing either a spatial gap (no direct contact) or a temporal gap (a delay before the second object started moving) in the sequence. After being

habituated with the causal displays, 10-month-olds showed a significant increase in looking times when presented with either of the non-causal configurations. However, when they were habituated with one of the non-causal interactions (e.g., a spatial gap), the infants only dishabituated to a causal display and not to the alternative non-causal event (e.g., a temporal gap). This suggests that infants can extract features relevant for identifying thematic roles in spatial processing before they have acquired language.

Previous work has examined whether participants can remember and describe causal push events within a multiple object tracking task (Jessop & Chang, 2018). The participants were presented with a display of nine visually identical circles that were initially moving in random patterns. After a short period, this movement stopped and up to three causal push events took place between separate pairs of circles. The participants were tasked with following the agents and patients from all of the pushes when they resumed their random movement. At test, two circles from one of the pushes and a random foil object were presented in different colours and the participants described how they interacted in an active transitive sentence, such as *blue pushed red*. The study found that the participants were able to accurately describe the events at test, suggesting that they could retain the agent and patient roles of the targets while tracking their movement. Critically, sentence accuracy also linearly deteriorated as the participants were required to track more pushes. It is therefore possible that processing multiple events simultaneously can lead to interference that affects meaning representations for language, particularly when describing several events such as in relative clause sentences. However, since each push event involved an independent pair of objects, there was no burden associated

with tracking the same referent playing different roles in different events. To address this, the present research offers six experiments comparing accuracy in tracking and describing interactions between visually identical objects with different thematic role configurations. The studies involved sequences where the same referent participated in two push events, playing either the same role in both events (best described with an SRC) or different roles (best described in an ORC). If consistent role-referent bindings have an important influence on meaning representations, then events matching SRCs should be more accurately described than those matching ORCs. The language requirements of the task were also varied to examine whether the structures used to describe the events affected accuracy, as predicted by the memory and experience-based approaches.

**Experiment Four: Subject and Object Relative Clauses**

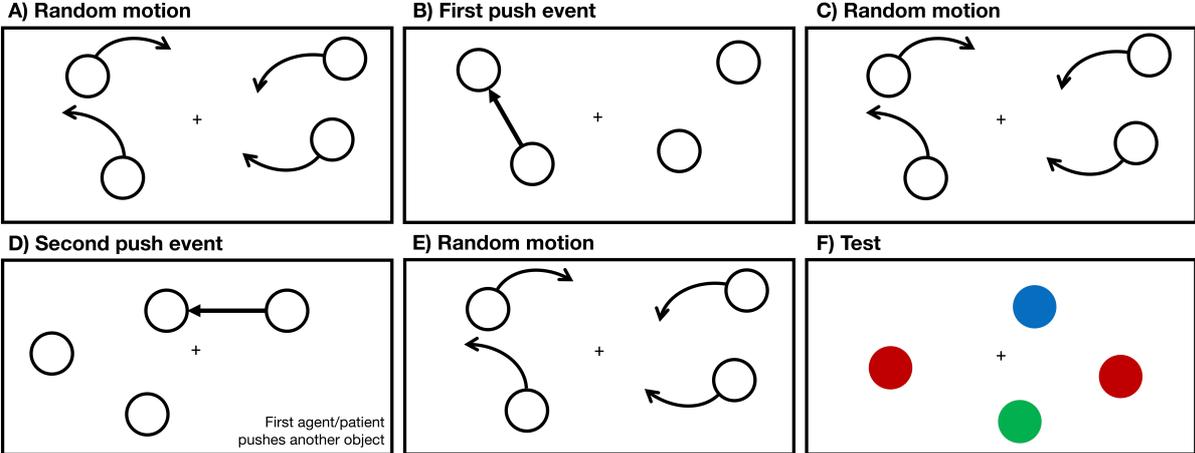


Figure 7. An illustration of the multiple push tracking task showing (a) random movement, (b) the first push event, (c) random movement, (d) the second push event, (e) random movement, and (f) the test display. The actual stimuli contained 9 circles and the arrows were not present.

In experiment four, participants were presented with visual scenes matching the thematic role configurations of SRCs and ORCs to examine whether they could track the agents and patients of two push events and describe them with an appropriate relative clause. The study adapted the multiple push tracking paradigm developed by Jessop and Chang (2018), which provides a controlled way of examining thematic role tracking within the visual system. In the present version of the task, the participants watched displays containing nine visually identical circles moving in unpredictable patterns (see figure 7A). During each trial, two push events occurred (figure 7B and 7D), separated by a short period of random movement (figure 7C). The participants were instructed to track the targets from the pushes and remember whether they played the agent or patient in the events. Importantly, one of the objects from the first event always appeared as the agent in the second push (see figure 7D). In the SRC trials, this overlapping object was the agent of the first push (i.e., the target was an agent twice), whereas in the ORC trials, it was the patient from the first event (i.e., a patient-agent). In both conditions, the participants needed to track the three targets and remember their various thematic roles for around ten seconds as they moved in random patterns for the remainder of the trials (figure 7E). At test, the three targets were highlighted in different colours, along with a random foil object in the same colour as the overlapping target (figure 7F). This meant that the participants needed to produce a restrictive relative clause structure to disambiguate which of the objects was being referenced (e.g., which of the two red circles pushed the green). Specifically, they described the interaction using either an SRC such as 6, or an ORC such as 7.

6. The *red* that pushed *blue* pushed *green*
7. The *red* that *blue* pushed pushed *green*

Since the objects were identical and the test phase was separated from the push events by periods of random motion, the participants needed to continuously track the agent and patient following the push events. It was predicted that accuracy in tracking and describing the interactions would be lower for the ORC than SRC trials, since the overlapping target in the ORC events plays different roles in the two pushes. Consistent with this hypothesis, evidence from the multiple object tracking (MOT) literature indicates that it may be harder to visually process and track event configurations matching ORCs. Many studies have reported that viewers can track only a small number of targets in parallel (e.g., Pylyshyn & Storm, 1988), but also that this tracking capacity diminishes as the risk of confusing the targets with the distractors increases. Specifically, manipulating display parameters such as the speed of the objects (Alvarez & Franconeri, 2007; Feria, 2013; Holcombe & Chen, 2012; Tombu & Seiffert, 2008, 2011), their proximity to each other (Franconeri et al., 2010, 2008; Tombu & Seiffert, 2008), the number of distractor objects (Bettencourt & Somers, 2009; Sears & Pylyshyn, 2000), or whether the targets cross visual hemifields (Alvarez & Cavanagh, 2005), has shown to significantly affect the viewers' ability to track the targets. Therefore, it is possible that when an agent becomes a patient in a subsequent visual event, the risk of confusing the thematic role features attached to the target increases, weakening the overall representation of the interaction.

As performance was assessed via sentence production rather than the traditional approach of measuring reading times during comprehension (Gordon

et al., 2001; Just & Carpenter, 1992; King & Just, 1991; MacWhinney & Pleh, 1988; Mak et al., 2002; Traxler et al., 2002; Wanner & Maratsos, 1978; Warren & Gibson, 2002), there was no ambiguity in the speakers intended message. Therefore, the existence of any performance asymmetries between the different event configurations would suggest a divergence in encoding the meaning of the event or in producing the message itself. The effects of sentence meaning and complexity on the relative asymmetry has previously been demonstrated by Fitz et al. (2011) using the Dual-path connectionist model (Chang, 2002; Chang et al., 2006). When trained with word sequences without access to their underlying meaning, the model was unable to fully learn transitive relative clause structures such as 1 and 2. Also, to completely neutralise the relative clause processing hierarchy, the authors needed to adjust the frequency of different structures in proportion to the number of thematic roles they contained. Alternatively, perception research has shown that both adults and infants in their first year can understand causal visual sequences composed of several discrete events purely from the movement of simple shapes (see Rips, 2011; Scholl & Tremoulet, 2000). Therefore, since most participants can track four targets (or two push events) with the display settings used in the present stimuli (Alvarez & Franconeri, 2007; Jessop & Chang, 2018), they may have the sufficient attentional resources available to accommodate role switches with no significant differences in accuracy between the SRC and ORC events. This would suggest that visual processing may not have a strong impact on the formation of the widespread relative clause asymmetry.

Experiment four used the multiple push tracking task to test the prediction that visual interactions matching SRC sentences are easier for participants to

track and then describe in a relative clause sentence at test than those corresponding ORCs. By focusing on non-linguistic processes, this study attempts to explain the SRC comprehension bias that has been observed in many languages with different properties without relying on shared linguistic constraints and heuristics (e.g., MacDonald, 2013).

## **Methods**

### *Participants*

22 participants were recruited from the undergraduate population of the University of Liverpool. The participants were required to be native English speakers with normal language and cognitive abilities, plus normal or corrected-to-normal vision. To ensure the sample size was sufficient to detect the hypothesised effects in the analysis, posthoc power simulations were conducted for the main analysis (see *Analysis*) and are reported with the model results.

### *Design*

The experiment used a univariate design with event configuration (SRC/ORC) as the predictor. These events differed in which of the two circles from the first push appeared as the agent in the second event (SRC = first push agent; ORC = first push patient). The participants completed 30 trials with each configuration, with the same event type never appearing more than twice in a row. This was controlled using two diametrically ordered counterbalancing lists, with half the participants randomly assigned to each group.

### *Stimuli and Apparatus*

Each trial lasted 25 seconds and involved animated sequences where nine identical white circles ( $0.8^\circ$  in diameter) moved randomly on a black screen. Throughout the entire trial, a red fixation cross ( $0.4^\circ \times 0.4^\circ$ ) appeared in the centre of the display. For the first three seconds, all nine circles moved in unpredictable random patterns at a constant speed of  $6^\circ/\text{sec}$  (figure 7A). This was controlled by an algorithm that reassigned the objects with a random vector within a  $120^\circ$  window approximately every 250 ms. Their direction was also changed whenever they moved closer than  $4.2^\circ$  to other objects (centre to centre), thereby forcing them to move away from each other.

After 3 seconds of random movement, the first push event occurred (figure 7B). Two of the objects were selected at random and assigned the roles of agent and patient. Then, all of the circles stopped moving as the agent and patient objects engaged in a causal launch event that lasted approximately 3 seconds (Michotte, 1946). Here, the agent directly approached the patient and immediately stopped upon contact, then the patient moved away along the same vector and at the same velocity. Afterwards, all nine circles resumed their random motion for one second (figure 7C) before another push event occurred (figure 7D). A third object was randomly selected to be the patient, while one of the objects from the first push served as the agent, as determined by the event type condition. In the SRC events, the agent from the first push also carried out the second. However, in the ORC events, the patient from the first push became the agent in the second. Once both pushes were completed, all nine objects continued to move in random patterns for the remainder of the trial (approximately 15 secs; figure 7E).

After 25 seconds, all movement was terminated and four of the nine circles were highlighted in red, blue and green (Figure 7F). Two of the circles were highlighted in the same colour (e.g. *red*). One of these was a foil randomly selected from the objects that had not been involved in any push event. The other was the circle that appeared in both of the push events; specifically, as either an agent in both (SRC events) or as a patient then an agent (ORC events). The purpose of having two circles highlighted in the same colour was to prompt the participants to produce this target in the head NP position of their utterance and encourage the use of a relative clause to disambiguate the target from the otherwise identical foil. The other two circles highlighted were the remaining targets that appeared in the first and second push actions.

The stimuli were shown in full screen on an LCD display (2880 × 1800; ~36.5° × 23.2° visual angle). The Processing 3 programming language (<https://processing.org/>) was used to design the stimuli, present the trials, and record the response data.

### *Procedure*

The participants were guided through example trials for both the SRC and ORC event configurations. They were instructed to track all the objects involved in all the push events, while remembering the agent and patient of each push. It was explained that they were to describe how the objects interacted in a sentence such as *the red that pushed blue pushed green* (SRC) or *the red that blue pushed pushed green* (ORC), using the appropriate colour words for the given trial. The participants were instructed to start their utterance (i.e., the head NP) with the target highlighted in the same colour as the foil. They were also

asked to fixate their gaze on the marker in the centre of the screen, although this was not monitored. After being randomly assigned to one of two counterbalance groups, the participants completed a total of 60 trials, with the opportunity to take breaks when needed. When the four circles changed colour at the end of the trial, they described the interaction before clicking the objects in the order they were spoken (head NP, embedded NP, matrix object NP). This provided a means of checking that all of the targets were being tracked, discouraging a strategy of relying on the head NP being highlighted in the same colour as the foil at test. Once all three circles had been clicked, the program recorded the participants' selections and advanced to the next trial. The participants' verbal descriptions were audio recorded to be later transcribed and coded offline by the experimenter. With the exception of the head NP, these verbal descriptions were used to determine their accuracy rather than the clicked responses, which were used only to verify the transcriptions. The final data showed whether the participants' utterances included an appropriate RC structure and the correct colours in each slot, as well as any errors that they made.

### *Analysis*

All of the analyses in the present work used generalised linear mixed-effects models implemented in the *lme4* package (Bates et al., 2015) in *R* version 3.5.1 (R Core Team, 2018). The dependent variable of the analysis was full accuracy in describing the visual event, which was coded on a binomial distribution (1 = correct, 0 = incorrect). For a description to be considered accurate, it needed to include the appropriate relative clause structure for the trial (e.g., an SRC for an SRC event configuration) with the correct colour words

in each position. The likelihood of achieving full accuracy in the task by chance was computed as 12.5% (structure = 50%; head NP = 50%; embedded+matrix NP = 50%), although chance comparisons were not used in the present analysis. Event configuration (SRC/ORC) was entered as an effect coded fixed factor. The random effects structure corresponded to the maximal model (Barr et al., 2013), with subject entered a random intercept and event configuration as a random slope. If necessary, the model was simplified until convergence was reached. Log likelihood-ratio ( $\chi^2$ ) comparisons obtained through the sequential decomposition of the model were used as confirmatory tests for the effect of configuration on accuracy (Bates et al., 2015). The marginal and conditional  $R^2$  effect sizes are also reported, which denote the variance explained by the model both with (conditional  $R^2$ ) and without (marginal  $R^2$ ) the random effect structure included (Johnson, 2014; Nakagawa & Schielzeth, 2013; Nakagawa et al., 2017). Finally, parametric bootstrapping ( $R = 1000$ ) was used to obtain 95% confidence intervals (CIs) and reliable p-values for the model estimates (Luke, 2017).

## Results

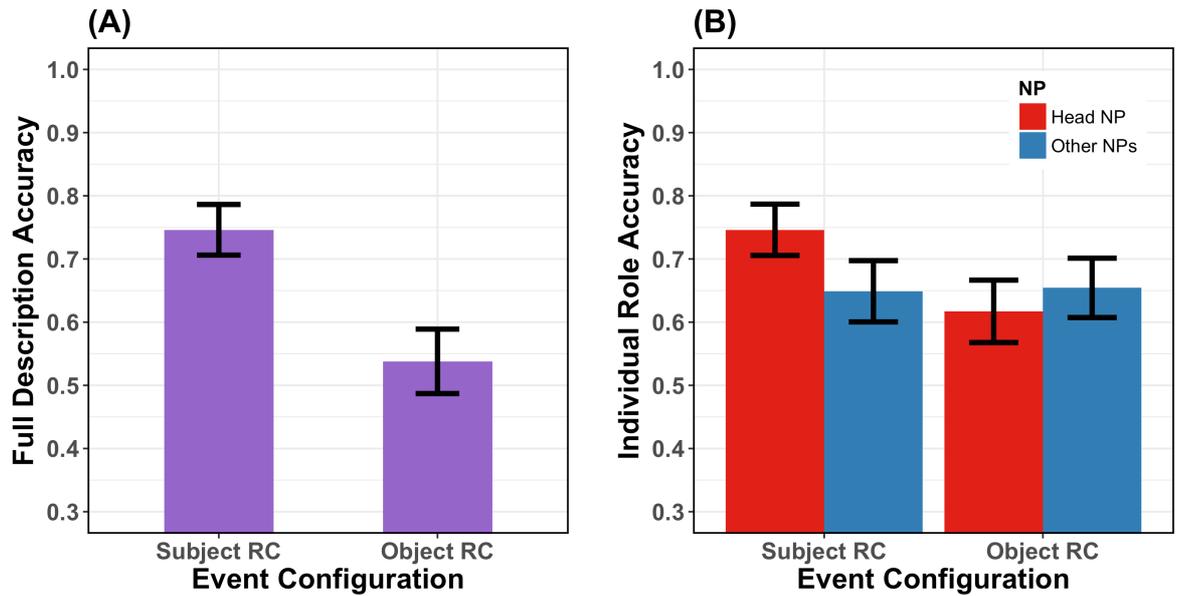


Figure 8. Plots to show the results of experiment 4.

Before fitting the model, trials where the participants did not produce the target highlighted in the same colour as the foil in the head NP position were removed, resulting in 1.29% of the total data being excluded from the analysis. The maximal model containing event type (SRC/ORC) as a random slope for subject converged without the need for simplification. This model showed significantly higher full response accuracy for the SRC ( $M = 0.7462$ ,  $SE = 0.04$ ) than the ORC ( $M = 0.538$ ,  $SE = 0.051$ ) event configurations ( $\beta = -1.1151$  [-1.3901, -0.8505],  $SE = 0.1377$ ,  $\chi^2 = 74.24$ ,  $p > .001$ ), accounting for 6.75% of the variance in the data without the random effect structure and 28.59% when it was included ( $R_m^2 = 0.0675$ ;  $R_c^2 = 0.2859$ ). These results are illustrated in figure 8A. Posthoc power analyses showed that the recruited sample of 22 participants provided a high level of statistical power ( $\beta = 1$ ) to detect effects of event type in this model.

To further investigate this difference, an exploratory mixed-effects model was fitted to the accuracy scores of the individual sentence positions. This model used a subset of the data containing only the trials where the appropriate RC phrase was produced at test, since thematic role assignment differs between these two structures. As in the previous analysis, event type (SRC/ORC) was entered as an effect coded fixed factor, with the addition of sentence position (head NP/other NPs) also as an effect coded factor. After producing the head NP, the participants were left with two coloured circles to assign to the embedded and matrix object NP positions. Consequently, accuracy for both of these positions was identical; they were either both correct or both incorrect. Therefore, accuracy for the head NP was compared with the other NP slots (embedded+matrix) combined. The maximal model that converged contained the main effects (but not interaction) of both factors as random slopes for subject. A significant interaction was observed between the event type and sentence position, with accuracy for the head NP being higher than the other NP positions in SRC but not ORC configurations ( $\beta = 0.7233$  [0.3617, 1.0805],  $SE = 0.1834$ ,  $\chi^2 = 16.11$ ,  $p < .001$ ). This model found no significant main effects and accounted for 6.74% of the variance in the data without the random effect structure and 28.56% when it was included ( $R_m^2 = 0.0674$ ;  $R_c^2 = 0.2856$ ), and suggests that tracking is facilitated by having the target appear consistently as the agent of the events (see figure 8B).

The results of experiment four showed that performance in tracking visual interactions and then producing an appropriate relative clause sentence was higher when the thematic role configurations of the events matched an SRC than an ORC sentence. In particular, the participants were more accurate in

identifying the head NP when the correct referent was the agent in both events (matching an SRC) compared to when it switched from being a patient to an agent (as in an ORC). These findings suggest that object tracking and differences in the non-linguistic comprehension of visual events could contribute to the persistent SRC comprehension bias that has been observed across many languages. Tracking role switches in ORC-type events could weaken the link between meaning and ORC forms. However, it is possible that the observed asymmetry was also affected by difficulties in using an ORC structure to describe an interaction between three animate targets. Previous work has shown that English speakers very rarely produce ORCs with animate head NPs, and would mostly favour a passive relative in these contexts (e.g., Gennari & MacDonald, 2009; Gennari et al., 2012; Roland et al., 2007). Thus, it is possible that the viewers were able to visually track and comprehend ORC events, but experienced difficulties in mapping the referents into these structures. Experiment five aimed to address this by substituting the ORC with a passive relative.

### **Experiment Five: Subject and Passive Relative Clauses**

Experiment four demonstrated that SRC event configurations are easier to visually track and describe than interactions matching ORCs. However, the task required the participants to produce an SRC or ORC, which may have contributed to the large divergence in accuracy. While ORCs are broadly more difficult to comprehend than SRCs, there are a number of circumstances in which this asymmetry is reduced. For instance, ORCs appear to be easier to process when they contain an inanimate head NP (see 8; Gennari &

MacDonald, 2008; Mak et al., 2002; Mak, Vonk, & Schriefers, 2006; Traxler et al., 2002; Traxler, Williams, Blozis, & Morris, 2005), a pronominal subject (see 9; Gordon et al., 2001; Gordon, Hendrick, & Johnson, 2004; Reali & Christiansen, 2007; Warren & Gibson, 2002), a proper noun (see 10; Gordon et al., 2001, 2004; Gordon, Hendrick, Johnson, & Lee, 2006), or a plausible semantic relationship (see 11; King & Just, 1991; Traxler et al., 2002).

8. *The movie that the director watched received a prize* [inanimate head NP and animate embedded NP]
9. *The dog that I stroked chased a cat* [pronominal subject]
10. *The team that Jonathan supports scored a goal* [proper noun as embedded NP and common noun as head NP]
11. *The robber that the fireman rescued stole the jewellery* [prototypical semantic events]

The events presented in the multiple push tracking task involved three animate targets that were randomly assigned colours at test to provide the participants with distinct common nouns to describe the scene. Under these circumstances, an asymmetry in the use of SRC and ORC forms may be expected. Previous research has shown that ORC sentences rarely appear with animate head NPs and are difficult to understand when they are encountered (Gennari & MacDonald, 2008, 2009). Instead, when describing instances where an animate referent switches thematic roles between events, evidence suggests that speakers prefer to produce passive relative clauses such as 12 (Gennari & MacDonald, 2008, 2009; Gennari et al., 2012).

12. *The green that was pushed by red pushed blue* [passive relative]

Consistent with experience-based models, corpus analyses have reported that passive relatives are broadly more common than ORCs in English (e.g., Roland et al., 2007). This preference for passive relatives also appears to increase with the level of conceptual similarity between the animate referents of the sentence (Humphreys, Mirković, & Gennari, 2016), congruent with memory-based accounts arguing that similarity between the head and embedded NP in ORC sentences causes interference during memory retrieval, leading to confusion with their thematic roles and ordering (Gordon et al., 2001, 2004; Van Dyke, 2003; Van Dyke & McElree, 2006). Thus, passive relatives may be favoured since they are more frequent and semantically similar to ORCs, but carry less risk of similarity-based competition as the two relative clause NPs are separated with a verb phrase (e.g., *was pushed by*) and so do not need to be held in memory at the same time.

Experiment five was identical to the fourth study, except the participants were instructed to describe the interactions with either an SRC or a passive relative. If accuracy continues to be significantly higher for visual interactions matching SRCs than ORCs, then it would suggest that difficulties in visually tracking the role switches in ORC events may be contributing the relative clause asymmetry. However, if no differences in accuracy are observed, this would suggest that ORC configurations are not more difficult to visually comprehend and that the results of the first experiment may have been due to the language requirements of the task.

## Method

The design, stimuli, apparatus, and analysis were identical to experiment four. Following the same criteria as the previous study, an additional sample of 20 undergraduate participants were recruited at the University of Liverpool. The procedure also matched the fourth study with one exception: the participants were instructed to describe the events using either an SRC (e.g. *the red that pushed blue pushed green*) or a passive relative clause (e.g. *the red that was pushed by blue pushed green*).

## Results

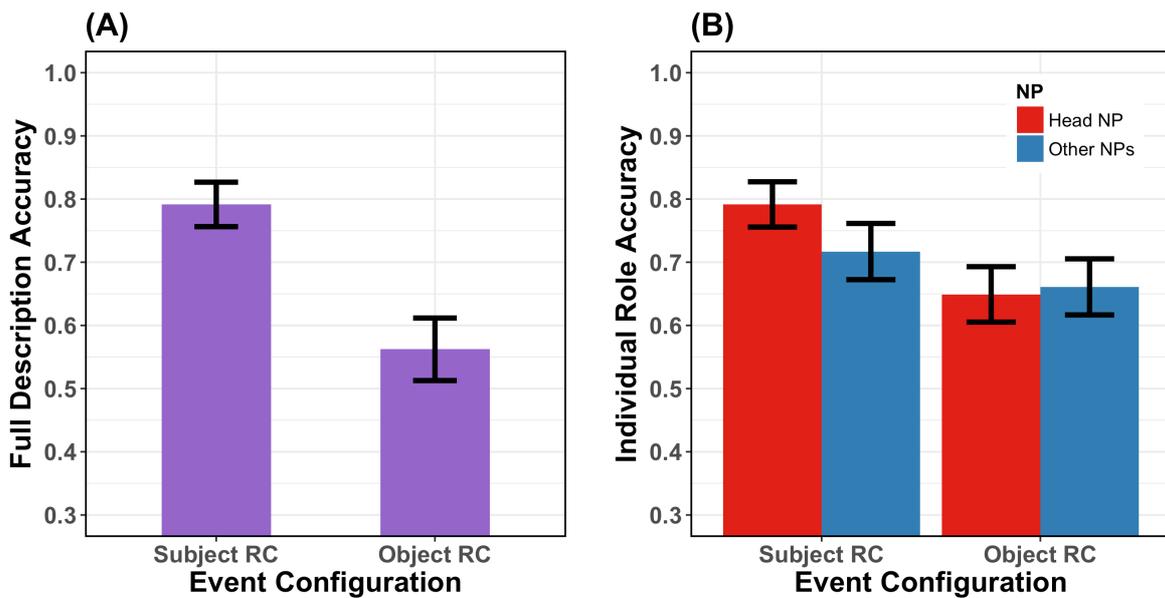


Figure 9. Plots to show the results of experiment 5.

Trials where the participants did not produce the target highlighted in the same colour as the foil in the head NP position were removed, resulting in 1.92% of the data being excluded from the analysis. The model that converged represented the maximal random effects structure with event type included as a random slope for subject. The results revealed significantly higher full response

accuracy for the trials with SRC ( $M = 0.7915$ ,  $SE = 0.0352$ ) than ORC ( $M = 0.5622$ ,  $SE = 0.0495$ ) event configurations ( $\beta = -1.3603$  [-2.0231, -0.7015],  $SE = 0.3344$ ,  $\chi^2 = 11.98$ ,  $p < .001$ ), as illustrated in figure 9A. The model accounted for 8.43% of the variance without the random effect structure and 40.1% when it was included ( $R_m^2 = 0.0843$ ;  $R_c^2 = 0.401$ ). Posthoc power analyses showed that the recruited sample size of 20 participants provided a high level of statistical power ( $\beta = 1$ ) to the detect effect of event type in this model.

As in experiment four, a follow-up model was fitted to the accuracy scores of each individual sentence position in trials where the appropriate RC structure was produced, with event configuration (SRC-target/ORC-target) and sentence position (head NP/other NPs) included as effect-coded fixed factors. The maximal model that converged contained the main effects of event type and sentence position as random slopes for subject. The results showed an interaction between the two factors, with accuracy for the head NP being significantly higher than the other NPs in SRC events but not ORC events ( $\beta = 0.5593$  [0.1601, 0.9604],  $SE = 0.2042$ ,  $\chi^2 = 7.56$ ,  $p = .006$ ). This effect is illustrated in figure 9B. The model found no significant main effects and accounted for 2.05% of the variance in the data without the random-effect structure, but 38.82% when the random variance was removed ( $R_m^2 = 0.0205$ ;  $R_c^2 = 0.3882$ ).

The findings of experiment five closely replicated those of experiment four, despite the language demands of the tasks being different. First, accuracy in tracking the pushes and describing them in an appropriate relative clause – specifically, an SRC or a passive relative – was higher for events matching the thematic role configurations of SRCs than those matching ORC (or passive

relative) sentences. Second, the follow-up analysis showed a boost in accuracy for the head NP in the SRC trials, where the target referent was an agent in both pushes, but no such advantage for the head NP in ORC events, where the target switched from being a patient to an agent. These results indicate that it is easier to track objects and their role features in interactions with consistent role-referent bindings, supporting the theory that the widespread SRC comprehension bias may originate in non-linguistic event processing. However, although passive relatives are more common than ORC structures in English (Roland et al., 2007) and are often produced with animate referents (Gennari & MacDonald, 2009), it is difficult to separate the contribution of visual object tracking from language processing in the current data as the participants were required to provide different linguistic responses at test depending on interaction that occurred during the trial. Since the thematic role features are being mapped into different relative clause forms, it is possible that performance in the task could have been influenced by variance in the natural frequency or the memory demands of each structure. Experiment six attempts to isolate the visual differences in this task by having participants produce simple active transitive sentences in both conditions.

### **Experiment Six: Relative Clause Events in Active Transitives**

The two previous experiments provided evidence that the preference for SRC structures observed across many different languages may have a non-linguistic basis. The studies observed that tracking and describing the causal relationships between visually identical targets was superior for events matching SRCs than ORCs. This suggests that a divergence in event comprehension

could influence the formation of language representations for these structures. However, in both of these studies, the participants needed to describe the sequences using an appropriate relative clause (i.e., an SRC or an ORC/passive relative), which meant that there were differences in the linguistic response provided in each condition. This leaves open the possibility that the asymmetry observed in these studies were due to the linguistic demands of the task, rather than differences in tracking the role-related features of the targets. However, while variation in the memory demands of different relative clause structures may contribute to the asymmetry (see Staub, 2010; Staub et al., 2017), cross-linguistic work shows that syntactic characteristics cannot provide a complete explanation. ORCs appear to be harder to comprehend and produce than SRCs across a diverse spectrum of languages, including those with head first (Dutch: Mak et al., 2002; German: Schriefers, Friederici, & Kuhn, 1995), head final (Korean: Kwon et al., 2010; Japanese: Miyamoto & Nakamura, 2003), flexible word order (Hungarian: MacWhinney & Pleh, 1988; Chamorro: Wagers et al., 2018), and ergative properties (Mayan: Clemens et al., 2015). Many investigations have also reported an SRC bias in Chinese (Jäger, Chen, Li, Lin, & Vasishth, 2015; Lin & Bever, 2006; Vasishth et al., 2013), despite that the ORC structures tested had shorter fill-gap distances and thus impose fewer memory demands than SRCs (cf. Chen, Ning, Bi, & Dunlap, 2008; Gibson & Wu, 2013; Hsiao & Gibson, 2003). This suggests that the difference in accuracy observed in experiments four and five were not due to the structures used to describe the events, as the relative clause asymmetry has repeatedly appeared in languages with different syntactic characteristics. Instead, the bias may emerge from non-linguistic processes, such as difficulties in tracking thematic

role switches in visual events. If this is the case, then higher accuracy for SRC events should still be observed even when they are not being described in relative clause sentences.

Experiment six aimed to extend the findings of the previous two studies by removing the differences in how the participants responded at test. Rather than using a relative clause to describe both push events, the participants were tested on only one of pushes (alongside two foil objects) and were instructed to use a simple active transitive sentence, such as *blue pushed green*. The thematic role configurations of the events during the trial remained the same. By isolating the visual differences in the task, this study tested whether accuracy would remain higher for events matching SRCs than ORCs, which would suggest that it is difficult to track the role switches in ORC events. However, no significant differences in accuracy would suggest that ORC configurations are not more difficult to visually comprehend and that the language requirements of the previous experiments produced the observed differences in accuracy.

## **Method**

### *Participants*

An additional sample of 20 undergraduate participants were recruited at the University of Liverpool.

### *Design*

Experiment six used a similar design to the previous two studies, with the addition of test push as an independent variable (first/second event). This determined which of the two pushes in the trial were presented to the

participants at test. The study consisted of 60 trials: 30 with each event configuration (SRC/ORC), each of which was subdivided into 15 trials testing the first push and 15 testing the second. Two counterbalance lists were generated to fix the order of these four combinations so that they were unpredictable and to prevent the same event configurations or test conditions appearing more than twice in a row.

### *Stimuli*

The stimuli involved the same multiple push tracking paradigm with identical SRC-target and ORC-target conditions as experiments four and five (see figure 6). However, the participants were tested on only one of the push events that occurred during the trial, rather than having to describe both interactions. At test, the four circles were highlighted in all different colours (red/blue/green/pink) as it was not necessary to elicit a relative clause. Two of these circles were the agent and patient of one of the events, while the other two were unrelated foil objects that did not feature in any of the pushes.

### *Procedure*

The procedure was the same as the previous experiments, except the participants described the events at test using an active transitive sentence such as *red pushed blue*. They were instructed to track all of the circles from both push events and were told that either of them could be tested.

## Results

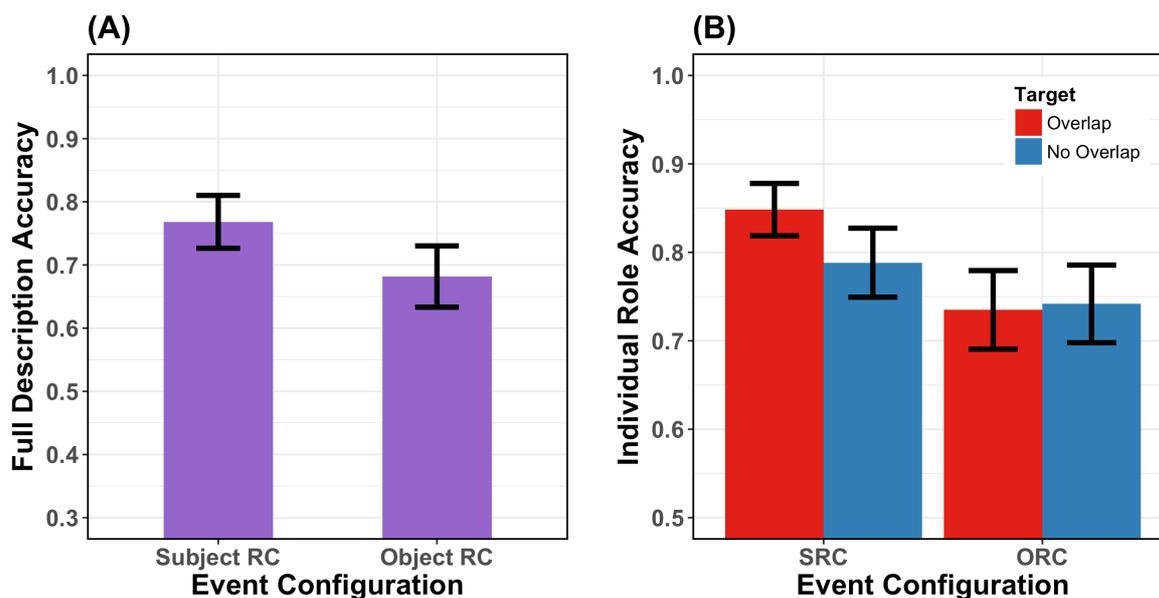


Figure 10. Plots to show the results of experiment 6.

The model that converged represented the maximal random effects structure with event type included as a random slope for subject. The results revealed significantly higher full response accuracy for the trials with SRC ( $M = 0.7683$ ,  $SE = 0.0418$ ) than ORC ( $M = 0.6817$ ,  $SE = 0.0485$ ) event configurations ( $\beta = -0.4774$  [ $-0.7836$ ,  $-0.1694$ ],  $SE = 0.1567$ ,  $\chi^2 = 7.88$ ,  $p = .005$ ), as illustrated in figure 10A. This accounted for 1.4% of the variance without the random effect structure, but 19.35% when it was included ( $R_m^2 = 0.014$ ;  $R_c^2 = 0.1935$ ). A posthoc power analysis found a high level of statistical power ( $\beta = 0.902$ ) to detect the main effect of event configuration in this model with a sample size of 20 participants.

A separate model tested for differences in accuracy in assigning the target objects appropriate thematic roles at test. As in the previous analysis, event configuration (SRC/ORC) was entered as an effect coded fixed factor. Target overlap was also entered as a fixed factor, which was effect coded for

whether the target appeared in two pushes (overlap) or just one (no overlap). For the SRC events, the overlapping target was the agent of both pushes, whereas in the ORC trials, the first patient reappeared in the second event as the agent. The maximal model that converged contained the main effects (but not interaction) of event type and target overlap as random slopes for subject. The model confirmed that accuracy was significantly higher for SRC than ORC events ( $\beta = -0.5403$  [-0.7492, -0.3229],  $SE = 0.1087$ ,  $\chi^2 = 25.13$ ,  $p < .001$ ) and found marginally higher accuracy for targets that appeared in two pushes compared to just one event ( $\beta = 0.2066$  [0.0014, 0.4104],  $SE = 0.1044$ ,  $\chi^2 = 2.81$ ,  $p = .093$ ). Importantly, there was a significant interaction between event configuration and push overlap ( $\beta = -0.4918$  [-0.9186, -0.0707],  $SE = 0.2163$ ,  $\chi^2 = 5.38$ ,  $p = .020$ ), with accuracy for overlapping targets being higher than non-overlapping targets in the SRC configurations, but no such differences in ORC events. This suggests that targets appearing as an agent twice are easier to track and describe in language than those that only appear in one of the pushes or switch roles between the events. This model accounted for 2.65% of the variance without the random effect structure, but 22.39% when it was included ( $R_m^2 = 0.0265$ ;  $R_c^2 = 0.2239$ ).

Finally, a comparison analysis was conducted on the full response accuracy scores from the combined data of experiments four, five, and six. As in the main analyses, the model included event configuration (SRC/ORC) as an effect coded predictor, with the addition of experiment (four/five/six) as a fixed factor with two Helmert contrasts. The first contrast compared experiments four and five to assess whether switching from an ORC to a passive relative had a significant impact on the asymmetry. For the second contrast, experiments four

and five were combined and then compared with accuracy in experiment six, to examine whether the observed asymmetry was significantly different when the participants used a transitive rather than a relative clause to describe the events. This analysis confirmed that full accuracy was significantly higher when describing the displays matching SRCs than ORCs ( $\beta = -0.8483$  [-1.0931, -0.6067],  $SE = 0.1241$ ,  $\chi^2 = 41.12$ ,  $p < .001$ ). The magnitude of this asymmetry did not significantly differ between experiments four and five ( $\beta = -0.1878$  [-0.7422, 0.3493],  $SE = 0.2784$ ,  $\chi^2 = 0.25$ ,  $p = .618$ ), suggesting that switching from an ORC to a passive relative did not affect the difficulty of the task. However, there was a significant interaction between event type and experiment ( $\beta = 0.3671$  [0.1265, 0.6186],  $SE = 0.1255$ ,  $\chi^2 = 8.18$ ,  $p = .004$ ), as the size of the asymmetry was significantly smaller in the sixth experiment compared to the other two studies combined, which appears to have been driven by higher accuracy levels for the ORC trials in experiment six. The model accounted for 6.28% of the variance in three studies without the random effect structure, but 29.98% when it was included ( $R_m^2 = 0.0628$ ;  $R_c^2 = 0.2998$ ).

Experiment six confirmed that visual events matching SRC sentences are easier to track and describe than those with ORC configurations, even when the corresponding relative clauses are not produced. However, the effect size was significantly lower than in the previous studies, as greater accuracy in the ORC trials was observed when the participants were not required to describe these events in a relative clause. This difference could have arisen from difficulties in mapping thematic roles into the relative clause forms, or the increased demands of having to report both pushes in experiment four and five, but only one of the pushes in experiment six. Collectively, the results of these studies suggest that

the relative clause asymmetry may be affected by how easy it is to track thematic roles and extract accurate non-linguistic event representations that can be mapped into language. The data also indicate that thematic role consistency may be the source of the divergence between these events, similar to the suggestions of previous models of language comprehension (MacWhinney & Pleh, 1988; Sheldon, 1974). Specifically, tracking was facilitated when the targets participated in multiple pushes with the same thematic role, whereas accuracy for targets that switched roles was similar to those that appeared in only one of the pushes. Since role consistency was not the main focus of the first three studies, additional experiments were conducted to more closely examine these effects under various role configurations.

### **Experiment Seven: Agent Consistency in Active Transitives**

The previous three experiments demonstrated that the relative clause asymmetry may be influenced by the limits of the object tracking system and differences in the complexity of the interactions being referenced. The studies provided converging evidence that it is easier to track visually identical agents and patients and describe their causal interactions in a sentence when the thematic roles of the targets were consistent throughout the sequence. Experiment seven extends these findings by further investigating the effects of thematic role consistency in forming meaning representations for language.

The first aim of the experiment was to test whether thematic role consistency influences description accuracy with different event configurations. The focus of the previous experiments was on two specific event types: SRC configurations, where one of the circles was the agent of two causal pushes,

and ORC configurations, where the patient of the first push became the agent of the second. Since the participants showed greater accuracy in describing SRC events with various sentence structures, the results indicate that it may be easier to track the targets and recall their interactions when they play the same thematic roles in both events. If this is the case, then thematic role consistency should also affect accuracy in describing events using different configurations that do not directly correspond to relative clause structures. Experiment seven tested this hypothesis by focusing on the consistency of the first agent. The study used scenes where the agent of the first push was always the overlapping target, reappearing in the second event either as the agent again or as the patient. As in the sixth study, the participants always used active transitive sentences to describe the scenes, such as *blue pushed green*. Higher accuracy in trials where the target plays the same role would indicate that thematic role consistency facilitates the formation of meaning representations for language. Alternatively, if no differences are observed, then the asymmetry reported in the previous studies may have been specific to SRC and ORC configurations, potentially due to experience with each of these event types (see Rips, 2011)

The second aim of experiment seven was to examine how role consistency affects tracking performance compared to a baseline condition containing two isolated causal interactions. The results of the previous three experiments suggest that playing the same role in both pushes provides a boost in accuracy for the target, but switching roles does not lead to a decrease in accuracy compared to objects only involved in one of the events. This implies that role switches do not complicate the event meaning but rather that role repetition enhances it. Since visual pointers need to be assigned to the targets

during the first event to continue tracking their location (e.g., Cavanagh & Alvarez, 2005; Oksama & Hyönä, 2016; Pylyshyn & Storm, 1988), it may be easier for the participants to encode their thematic role features in the second push, particularly if they appear in the same role as the first event, as their attention is already drawn to these targets and their role features are in agreement between the two events. However, since the studies did not include a control condition without an overlapping target in both pushes, it is difficult to fully assess how the role switches affected the observers' comprehension of the events. Perception research has consistently found that participants can identify causality and animacy from the movement of simple shapes (see Rips, 2011; Scholl & Gao, 2013; Scholl & Tremoulet, 2000). To achieve this, the viewers need to extract features from these motion patterns to trigger the perception of causality and identify the roles of the objects involved in the event. Only the properties of individual targets are needed to recognise that it is animate, such as when it appears to be self-propelled and makes unprovoked changes in velocity (Luo, Kaufman, & Baillargeon, 2009; Tremoulet & Feldman, 2000). However, causal interactions often require the participants to analyse relationships between separate objects; for instance, the perception of chasing events depends on the directness of the agent's movement towards the patient (Gao et al., 2009). Thus, role switches could have a holistic effect on the entire event representation by confusing the relational features between the targets, which would affect accuracy with both the overlapping targets and those it interacted with. To address this, experiment seven examined the impact of thematic role consistency by comparing performance to a control condition where there is no overlapping target between the pushes.

## Method

### *Participants*

Another 24 participants were recruited from the undergraduate population of the University of Liverpool. As in the previous studies, all subjects were required to be native English speakers with normal language and cognitive abilities, plus normal or corrected-to-normal vision.

### *Design*

The experiment tested the effects of role overlap on description accuracy. The overlap factor consisted of three levels reflecting the role that the agent from the first event played in the second push: same role (agent), different role (patient), and no overlap. At test, the participants were presented with the two circles from the first push and a third foil object, all in different colours (red, blue, and green). The foil object was either a random object that did not appear in any of the pushes or the third target circle from the second push event. This was to control for potential confounding effects arising from the number of push targets the participants needed to consider when formulating their description. The participants completed 60 trials with an equal number of each of the overlap levels, which were ordered using two counterbalance lists to ensure that the same condition did not appear more than twice consecutively.

### *Stimuli*

The same multiple push tracking task as the previous three studies was used, but with notable differences in the push event configurations and test event. In the earlier experiments, either the agent (SRC) or patient (ORC) from

the first push event appeared as the agent of the second. This was reversed in experiment seven, as the overlapping circle was always the agent from the first event and its role in the second push was manipulated. The study also included trials in which the two events were independent and involved different circles. During the test events, the participants were presented with three circles highlighted in different colours (red/blue/green), whereas four circles were given in the previous experiments. Two of these were the agent and patient targets to be reported, while the third object was either a random foil distractor or the remaining target from the second push event.

### *Procedure*

The procedure was identical to the previous experiments with two critical differences. First, the participants were instructed to track only the circles involved in the first push and to ignore their roles in any subsequent events. Second, similar to experiment three, the participants were instructed to describe the interaction between the agent and patient at test using an active transitive structure, such as *red pushed blue*.

### *Analysis*

Experiment seven followed a similar data analysis protocol as the previous experiments, using bootstrapped generalised maximal mixed-effects models with full description accuracy as the outcome variable. For an utterance to be considered accurate, the participants needed to correctly identify both the agent and patient of the push event in an active transitive structure at test. Role overlap was entered as a deviation coded fixed factor, in which the same and

different role conditions were separately contrasted with accuracy in the no overlap conditions. This allowed the model to assess how thematic role consistency impacts on description accuracy compared to a neutral baseline condition. Subject and foil type were entered as random intercepts, each with random slopes for role overlap.

## Results

The maximal model that converged contained the random intercept of foil type, as well as a random intercept for subject with random slopes for role overlap. This model revealed that accuracy in the different role condition was not significantly different to the no overlap trials ( $\beta = -0.7668$  [-1.3091, -0.2019],  $SE = 0.2824$ ,  $\chi^2 = 7.74e-03$ ,  $p = .930$ ). However, significantly greater accuracy was observed in the same role trials than the no overlap trials ( $\beta = 1.4554$  [0.7594, 2.0985],  $SE = 0.3416$ ,  $\chi^2 = 13.79$ ,  $p < .001$ ). The model accounted for 5.5% of the variance without the random effect structure, but 31.73% when it was included ( $R_m^2 = 0.055$ ;  $R_c^2 = 0.3173$ ). A posthoc power analysis found a high level of statistical power to detect the effects of overlap in this model with a sample of 24 participants ( $\beta = 1$ ).

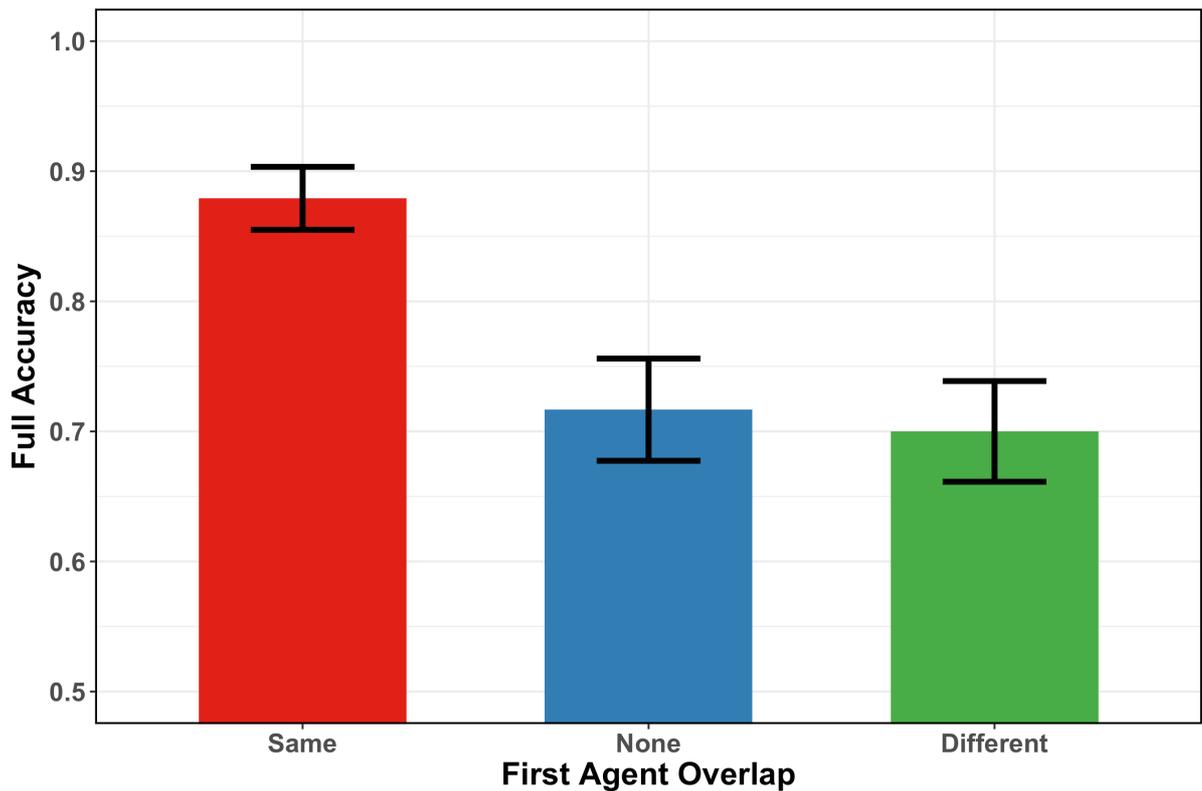


Figure 11. Plots to show the results of experiment 7.

The results of experiment seven suggest that thematic role consistency can influence description accuracy in contexts that do not directly correspond to relative clause sentences, providing further evidence that visual event processing can affect meaning representations for language. Critically, switching thematic roles between the pushes does not appear to disrupt event representations compared to a control condition with no overlapping targets in both events. Instead, role repetition makes it easier to encode the interactions and describe them in language. Similar findings were reported by Kahneman et al. (1992), who found that participants named targets faster at test when they were given a “preview” of their labels before they changed position in the display, but only when these labels were attached to the same perceptual objects between the preview and test phases. Therefore, while observers can

track the thematic role features of individual targets in the display to produce accurate sentences (see also Jessop & Chang, 2018), these role-referent bindings are enhanced through successive presentations of the same role features with the same targets. This suggests that comprehending and describing SRC configurations is facilitated by thematic role repetition, rather than ORC events being disadvantaged by role switches. However, this study focused on the agent of the interaction, which previous research suggests are easier to encode and process than other entities in the scene (Cohn & Paczynski, 2013; Jessop & Chang, 2018). Therefore, experiment eight aimed to provide a more stringent test of role consistency by using trials where the overlapping target was always the patient from the first event.

### **Experiment Eight: Patient Consistency in Active Transitives**

The four previous studies have provided evidence that thematic role consistency facilitates the comprehension of visual events and the formation of accurate meaning representations that can be expressed in different language structures. However, this evidence is derived from stimuli where one target plays the agent in two causal actions. Therefore, a further test of thematic role consistency would be to look at performance when the overlapping object is always the patient of from the first push event. Many studies have observed that agents hold an advantage over patients across different domains. Specifically, agents appear to provide more information about the structure of an event (Cohn & Paczynski, 2013), they tend to be fixated on first in visual scenes (Webb, Knott, & MacAskill, 2010), and they precede patients in the canonical word order of most languages (Kemmerer, 2012). Perhaps the most relevant

example for the present research are the findings of Jessop and Chang (2018), as accuracy was consistently around 6% higher for agents than patients in a multiple push tracking task where the participants described non-overlapping interactions. Promoting this critical event participant to a *double agent* in SRC configurations may reinforce the already salient agent features, thus it is possible that patient consistency will not offer the same boost in description accuracy. To examine this, experiment eight shifted the focus away from the agent by using a version of the task where the overlapping target was always the patient from the first event. If accuracy remains higher when the overlapping target plays the same role in both pushes, then it suggests that role consistency has general influence on event representations. Otherwise, the findings would highlight the importance of the agent variable on message formation in language.

## **Method**

The design, stimuli, apparatus, and analysis were identical to experiment seven with one key change; whereas the first agent was always the overlapping object in experiment seven (except for in the no overlap trials), it was the first patient that always appeared in the second event in experiment eight. Following the same criteria as the previous study, an additional sample of 24 undergraduate participants were recruited at the University of Liverpool.

## **Results**

The maximal model that converged contained random intercepts for subject and foil type without random slopes. The model showed that full description accuracy in the same role condition was significantly higher than

trials with no overlap ( $\beta = 0.779 [0.3207, 1.2227]$ ,  $SE = 0.2301$ ,  $\chi^2 = 12.06$ ,  $p < .001$ ), but there were no significant differences in accuracy between the different role and no overlap trials ( $\beta = -0.2037 [-0.622, 0.2175]$ ,  $SE = 0.2142$ ,  $\chi^2 = 1.23$ ,  $p = .268$ ). The model accounted for 2.1% of the variance without the random effect structure and 15.46% when it was included ( $R_m^2 = 0.021$ ;  $R_c^2 = 0.1546$ ). Posthoc power simulations showed a high level of statistical power ( $\beta = 0.887$ ) to detect the effects in this model with a sample of 24 participants.

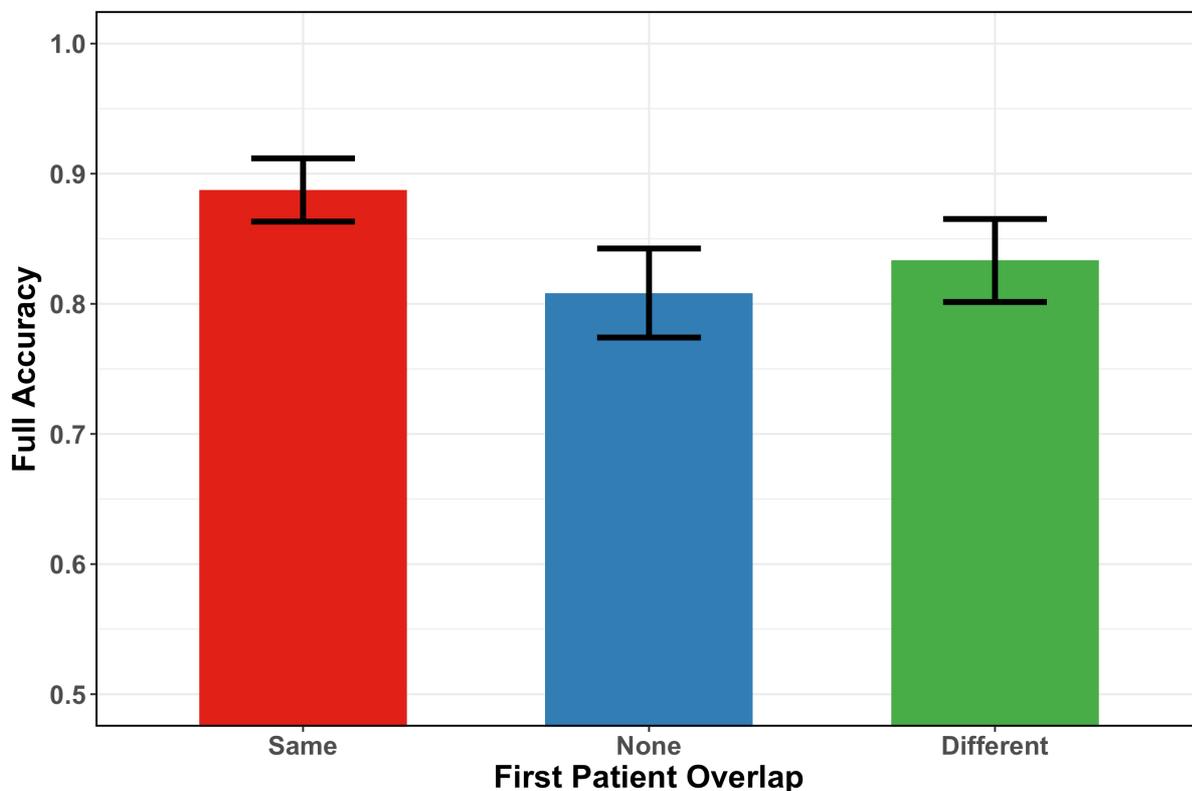


Figure 12. Plots to show the results of experiment 8.

The results of experiment eight provide further evidence that role feature repetition enhances thematic role tracking, while role switches do not necessarily make tracking more difficult. Specifically, compared to the control condition with no overlap, description accuracy was boosted when one of the

targets played the patient in both pushes, but it was not significantly affected when the first patient switched to being the agent. The findings parallel those of the seventh study, where the first agent was always the overlapping object. This implies that meaning representations for language are sensitive to event properties, such as the consistency of the role-referent bindings for both the agent and patient of the interaction. The final experiment presented an additional group of participants with the same stimuli as the eighth study, but instructed them to describe the interactions in passive transitive structures such as *blue was pushed by green*, to establish whether the results of experiment eight could be replicated when event representations are mapped into an alternative transitive form.

### **Experiment Nine: Patient Consistency in Passive Transitives**

Experiment nine uses the same event configurations as the previous study, with the overlapping target always being the patient from the first push event. However, the participants were instructed to describe the scenes using passive transitive structures, such as *red was pushed by blue*. This was to test whether production accuracy in the push tracking task is also determined by the specific structures that the role representations are mapped into. Evidence from experiments four and five showed that switching relative clause forms, specifically from an (active) ORC to a passive relative, did not significantly affect accuracy in describing the stimuli. Yet, changing from these relative clauses to a simple transitive sentence in experiment six reduced the magnitude of the observed asymmetry. This reduction was possibly driven by task demands rather than any direct interference from the structures themselves. The

participants needed to report both of the pushes when producing relative clause structures, but only one of the interactions when using a transitive sentence. They also needed to choose the correct type of relative clause in experiment four and five, but the same structure was used for both conditions in experiment six. This meant there was less opportunity for error in the sixth study.

The data presented throughout the current work points to the conclusion that accuracy in describing scenes with causal interactions, either in a relative clause or transitive sentence, is primarily determined by the ability to track the thematic roles of the entities throughout the sequence. However, other evidence suggests that speakers prefer to produce more accessible and frequently occurring referents earlier in the sentence (see MacDonald, 2013). When a scene contains a *double patient*, it may be easier to produce a passive transitive than an active sentence, as previous studies have observed preferences for passive transitives and passive relative clauses when the patient or experiencer is more salient than the agent or cause of the event (e.g., Ferreira, 1994; Gennari & MacDonald, 2009). Therefore, experiment nine aimed to replicate the eighth study with the participants describing the interactions in passive transitive structures. A second comparison model was then fitted to combined data of experiment seven, eight and nine, to examine how description accuracy was affected by the shift of focus from the agent to the patient and changes in the language requirements of the tasks.

## **Method**

Experiment nine used the same design, stimuli, and analysis as experiment eight. The procedure was also identical, except the participants were

instructed to describe the interactions at test using a passive transitive structure (e.g., *green was pushed by blue*) rather than an active sentence. For this study, a further 24 undergraduate participants were recruited at the University of Liverpool.

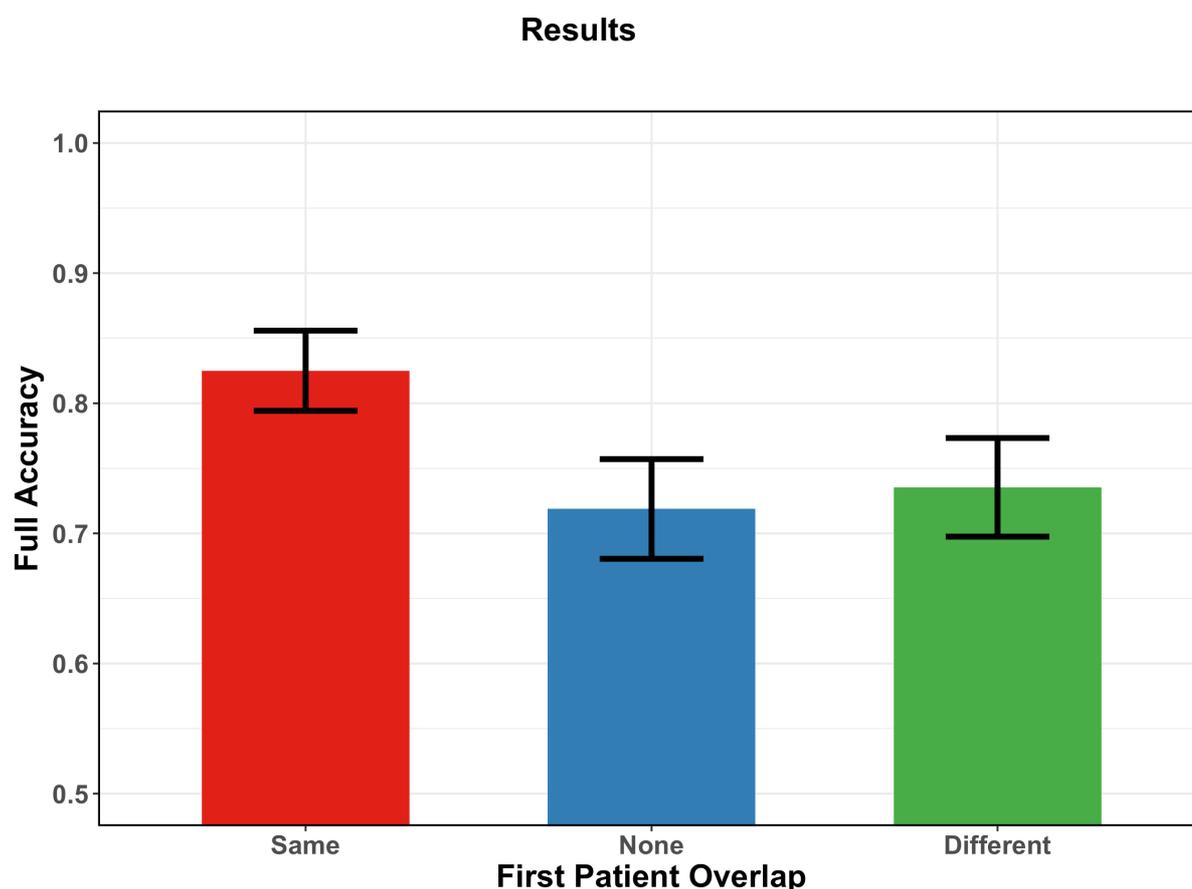


Figure 13. Plots to show the results of experiment 9.

The maximal model that converged contained random slopes for push overlap. This model revealed no significant differences in accuracy between the no overlap and different role conditions ( $\beta = -0.3585$  [-0.7374, 0.0135],  $SE = 0.1915$ ,  $\chi^2 = 0.44$ ,  $p = .505$ ). However, significantly greater accuracy was observed in the same role trials than the no overlap trials ( $\beta = 0.9248$  [0.5136, 1.325],  $SE = 0.2070$ ,  $\chi^2 = 21.21$ ,  $p < .001$ ). The model accounted for 2.39% of

the variance without the random effect structure, but 27.78% when it was included ( $R_m^2 = 0.0239$ ;  $R_c^2 = 0.2778$ ). A posthoc power analysis found a high level of statistical power ( $\beta = 0.976$ ) to detect the effects of role overlap in this model with a sample of 24 participants.

A final comparison model was fitted to the full description accuracy scores in the combined data from experiment seven, eight and nine to compare performance in these three studies of thematic role consistency. The model included role overlap (same role/different role/no overlap) as a deviation coded predictor to match the separate analyses used for these datasets. Additionally, experiment (seven/eight/nine) was entered as a Helmert coded fixed factor with two contrasts. The first contrast considered the differences in accuracy between experiment eight and experiment nine, which both involved an overlapping patient but required the participants to describe the events in either an active (experiment eight) or passive structure (experiment nine). This provided a test of whether the language requirements of the task influenced accuracy. The second contrast tested the difference between experiment seven and the combined accuracy of experiments eight and nine. This was to assess whether the effects of thematic role consistency are different for the agent and patient of the events. The model that converged included subject and foil type as random intercepts without any random slopes. The results confirmed that there were no significant differences in accuracy between the different role and no overlap conditions ( $\beta = -0.5174$  [-0.737, -0.2866],  $SE = 0.1149$ ,  $\chi^2 = 0.35$ ,  $p = .555$ ), whereas accuracy was significantly higher for the same role conditions than the no overlap trials across the three studies ( $\beta = 1.1258$  [0.8687, 1.3617],  $SE = 0.1258$ ,  $\chi^2 = 88.39$ ,  $p < .001$ ). There were no overall differences in accuracy between experiment

eight and nine ( $\beta = -0.5013 [-1.1147, 0.0652]$ ,  $SE = 0.3010$ ,  $\chi^2 = 2.78$ ,  $p = .106$ ) or between experiment seven and combined accuracy of studies eight and nine ( $\beta = -0.1319 [-0.4614, 0.2102]$ ,  $SE = 0.1713$ ,  $\chi^2 = 0.79$ ,  $p = .374$ ). However, there was a marginal interaction between experiment and the difference between the same role and no overlap conditions ( $\beta = 0.2566 [-0.0338, 0.537]$ ,  $SE = 0.1456$ ,  $\chi^2 = 3.11$ ,  $p = .078$ ), with agent consistency yielding a slightly larger boost in production accuracy patient consistency. The model accounted for 4.87% of the variance in the combined data without the random effect structure and 25.49% when controlling for the random intercepts of subject and foil type ( $R_m^2 = 0.0487$ ;  $R_c^2 = 0.2549$ ).

## General Discussion

Since subject-extracted relative clauses (SRCs) are reported to be easier to comprehend than object-extracted relative clauses (ORCs) across a range of typologically diverse languages (e.g., English: King & Just, 1991; Korean: Kwon et al., 2010), we investigated whether there is a non-linguistic bias for SRC events due to the consistency of their role-referent bindings. Participants were presented with visual sequences containing nine moving white circles that were identical in appearance. During each trial, two causal events would take place in which one circle travelled directly towards another and pushed it away, before all nine objects moved around the display in unpredictable random patterns. Critically, one of the circles from the first push event would also appear in the second, playing either same role in both events (e.g., the agent in both push one and two) or switching to the alternative role (e.g., the patient in push one and the agent in push two). The participants were tasked with tracking the targets from

the pushes and then describing their interactions when the objects were presented in different colours at test.

Across six experiments, more accurate descriptions were produced when the overlapping referent was paired with the same thematic role in both pushes, compared to when it played different roles in each event. Experiment four observed higher accuracy when the participants tracked push interactions corresponding to an SRC (with consistent roles) compared to an ORC (with role switches) and were then required to produce the matching relative clause structure at test. These results were replicated in experiment five where the participants were instructed to describe the same causal interactions in either an SRC or a passive relative. Experiment six also observed superior performance for the SRC configurations when simple active transitive sentences were used to describe only one of the pushes in the trial. The subsequent studies examined whether these differences in accuracy were driven by the consistency of the role-referent bindings, which was tested using different event configurations and sentence structures. Experiment seven focused on the consistency of the agent from the first event, with the participants producing active transitive descriptions at test. Compared to a control condition, there was no reduction in description accuracy when the first agent switched roles, but there was a significant increase in accuracy when it reappeared as the agent in the second event. Parallel results for the patient of the first push were observed in experiment eight, as description accuracy in an active transitive structure was boosted when the target reappeared as the patient in the second event, but did not decrease when it switched to being an agent. The final study also focused on the first patient and observed similar results to experiment eight when the participants

produced passive transitive descriptions at test. Collectively, the findings of the present studies suggest that role-referent consistency influences the level of difficulty that speakers experience when describing events in language.

When formulating accurate meaning representations of the interactions, the participants faced two challenges. First, they needed to track the location of the target objects for around ten seconds as they moved in random patterns after the push events. Although various tracking mechanisms have been proposed, a common position shared by most theories is that viewers can track the location of a small number of moving objects in parallel (Alvarez & Cavanagh, 2005; Cavanagh & Alvarez, 2005; Howe et al., 2010; Oksama & Hyönä, 2016; Pylyshyn, 1989). The exact number of targets that can be sustained simultaneously appears to be flexible rather than fixed, as the observed tracking capacity has shown to vary with the attentional demands of the task (Tombu & Seiffert, 2008), the risk of confusing the targets with the distractors (Bettencourt & Somers, 2009; Franconeri et al., 2010), and individual capabilities (Green & Bavelier, 2006; Oksama & Hyönä, 2004). Based on previous studies (Alvarez & Franconeri, 2007; Jessop & Chang, 2018), it was calculated that viewers would be able to track up to four objects or two push events simultaneously. Therefore, the present research taxed the participants' tracking resources, as they needed to monitor three circles from two push events. The second challenging aspect of the task was that the viewers needed to identify and attach thematic role information to specific targets and then retain and update these associations as the objects moved in unpredictable patterns. Since all of the objects were white circles until the end of the trials, there was no visual information to help distinguish and remember the targets and their

interactions following the push events. Other studies have indicated that the identity and location of an object are processed separately (Pylyshyn, 2004). Many features, such as colour and shape, are not automatically extracted during parallel object tracking (Bahrami, 2003; Saiki, 2003). Instead, additional overt attention is needed to remember the identity of specific targets (Oksama & Hyönä, 2016), which appears to reduce the number of objects that can be tracked simultaneously (Horowitz et al., 2007). Despite the challenging nature of the present task, accuracy in tracking and describing the interactions remained high throughout all six experiments. This is consistent with the earlier work of Jessop and Chang (2018) showing that participants can track the thematic role features of multiple objects, even under difficult and attentionally demanding conditions.

Experience-based theories argue that the relative clause asymmetry reflects the level of exposure speakers have had to these particular language structures (Fitz et al., 2011; Gennari & MacDonald, 2008; Hale, 2001; Levy, 2008; MacDonald, 2013; MacDonald & Christiansen, 2002). An advantage of this framework is that it can explain the role of semantic content and noun type as moderators of the asymmetry. Studies combining both experimental tasks and corpus analyses have shown that ORCs are most frequently used with an inanimate head NP or a pronominal subject and that the relative clause asymmetry is vastly reduced when participants are tested with these types of sentences (Kidd, Brandt, Lieven, & Tomasello, 2007; Mak et al., 2002; Reali & Christiansen, 2007). ORCs also appear to be easier to process and understand when only one of the NPs in the relative clause is a plausible agent of the action, or when the described relationship is pragmatically typical of the referents (see

11 above, King & Just, 1991; Traxler et al., 2002). These previous findings suggest that comprehenders draw on statistical information and conceptual knowledge during sentence processing. However, throughout the present research, the participants described randomly moving white circles using full NPs (e.g., *red, blue, green*), with no strong differences in animacy, conceptual typicality, or noun form that could potentially influence the results. Furthermore, in experiment six, the participants used simple transitive structures to describe all of the interactions, avoiding the potential differences in experience associated with each relative clause structure. Yet, an SRC advantage was still observed throughout the experiments, even after these experienced-based factors were neutralized, suggesting that other non-linguistic processes can create a production bias for SRCs.

Alternatively, memory-based theories argue that ORC structures demand more cognitive resources to process correctly than SRCs, primarily due to their differences in word order (Gibson, 1998; Gordon et al., 2001, 2004; Warren & Gibson, 2002). The present research varied the sentence structures used to describe the interactions between the different experiments to isolate the contribution of non-linguistic event processing. Specifically, the participants described the interactions with either an SRC or an ORC (experiment 4), an SRC or a passive relative (experiment 5), an active transitive (experiment 6-8), or a passive transitive (experiment 9). Despite these changes, all six experiments observed more accurate descriptions for scenes where the referents played the same thematic roles in both events (e.g., SRC events) compared to when they switched roles (e.g., ORC events). Since the asymmetry was consistently observed using a variety of sentence structures, the present

findings cannot be explained solely by memory-related syntactic processing. Instead, the results suggest that language-independent processes related to event comprehension may contribute to the persistent SRC bias.

A consistent pattern in the current data is that thematic role repetition leads to greater accuracy in tracking and describing the interactions, but role switches do not increase the difficulty of the task. Previous research in the perception literature has observed that visual attention is often deployed to integrated objects rather than individual features (see Chen, 2012). By attending to specific objects, viewers often fail to notice other salient aspects of the scene (Mack & Rock, 1998; Simons & Chabris, 1999). For instance, studies using multiple object tracking tasks have found that viewers are faster and more accurate in identifying colour changes on the attended target objects than on the unattended distractors (Bahrami, 2003). However, an advantage of this object-focused attention is that it is easier to subsequently detect or respond to these specific entities and their individual features (Kahneman et al., 1992; Kristjánsson & Campana, 2010). During the first push event, the participants needed to assign parallel visual attention to the agent and patient to continue tracking their location (Cavanagh & Alvarez, 2005; Oksama & Hyönä, 2016). To detect causality from movement, viewers need to continuously process the motion features of the scene rather than taking episodic samples, as it would not be possible to identify an interaction such as a push or a chase from a single frame. Thus, by tracking the random movement patterns of the targets from the first push event, the participants would be susceptible to further thematic role cues. This could lead to a boost in the strength of the target role bindings when subsequent thematic cues provided during the second push are in agreement

with those of the first event, since there is more *positive* evidence for the previous roles of these targets. Evidence for this comes from earlier work showing that positive and negative evidence can affect the detection of the wolf (i.e., the agent) of a chasing interaction (Gao & Scholl, 2011). Thus, the observed effects of role consistency on description accuracy suggest that viewers are utilising their object tracking pointers to extract thematic role features from the scene to support the mapping of referents into appropriate sentence positions.

While the present data suggests SRC events have a non-linguistic advantage over ORCs due to the consistency of their role-referent bindings, it is not our claim that these differences can explain all relative clause comprehension behaviour. Rather, we suggest that thematic role consistency exerts a small but persistent effect on language processing. MacDonald's (2013) Production-Distribution-Comprehension (PDC) account may help connect the current findings to previous comprehension studies. This account argues that the constraints and heuristics used to facilitate language production can influence frequency distributions and promote comprehension preferences. According to the PDC, the widespread SRC preference emerges from speakers' tendencies to produce the easiest referents first, reuse previously encountered word sequences, and attempt to avoid potential interference in memory. The present work has demonstrated that complex visual events are more accurately described when their configurations match SRCs than ORCs. A small difference in the visual comprehension of these events across many speakers may influence the evolution of the language, to the extent that the SRC becomes the preferred processing configuration. Speakers may be more likely to describe

SRC events if they are easier to understand non-linguistically, which could promote SRC-biased distributions by boosting the frequency of these structures, particularly with the types of arguments that are often involved in these interactions in the visual world (e.g., animate referents). The SRC advantage may then become magnified by a combination of memory and experience-based factors (Staub, 2010; Staub et al., 2017). As proposed in the PDC account, language learners acquire these biased representations from the input they receive from others, compounding the asymmetry in the language. Thus, even when comprehending sentences that describe abstract interactions where visual role tracking is not required, the asymmetry may still be observed since it exists in the comprehenders' language representations, partly shaped by a weak but consistent non-linguistic bias present since the beginning of language evolution.

### **Additional Information**

At the time of writing, the studies presented in chapter three are in preparation to be submitted for publication<sup>2</sup>. As in the previous chapter, I was solely responsible for designing and conducting this research under the supervision of Dr. Franklin Chang as my primary PhD advisor. There were no other collaborators that made notable contributions to this work.

---

<sup>2</sup> Jessop, A. & Chang, F. (2018). The relative clause asymmetry is shaped by role tracking difficulties. [in preparation]

## CHAPTER FOUR: DISCUSSION

The aim of this thesis was to address limitations with existing theories of thematic roles by proposing a new solution based on visual event processing. Thematic roles, such as the agent and patient, encode the function that a referent plays in an event and offer a way of representing meaning in language that does not rely on specific words, concepts, or sentence structures. Traditional linguistic theories have considered thematic roles to be a type of language representation that can be identified from specific features or how closely an entity resembles conceptual role representations (Dowty, 1991; Fillmore, 1967; McRae et al., 1997). Concept-based accounts have proposed general features to identify roles based on their short-term behaviour (e.g., their volition and sentience; Dowty, 1991), but have also suggested that our long-term world knowledge helps to guide thematic role assignment (e.g., *doctors* are typical agents of the verbs like *operate*; Ferretti et al., 2001; McRae et al., 1997). However, we are often faced with situations that require us to assign thematic roles without having access to extensive world knowledge or the immediate behavioural features of the entity (i.e., their internal states such as volition or sentience), so we cannot always rely on these abstract cues to identify *who did what to whom*.

Despite these difficulties, perception studies have shown that the movement features of simple shapes can generate impressions of animacy and causality. For example, participants will describe dynamic scenes containing simple shapes as if these objects were animate beings, producing sentences such as “the big triangle is chasing the circles” (Barrett et al., 2005; Heider & Simmel, 1944). They can also identify causal relationships between shapes

based entirely on covariation in their motion patterns (e.g, spatial and temporal features), such as one square appearing to *push* another square across the screen (Michotte, 1946; Schlottmann et al., 2006; Scholl & Tremoulet, 2000). Critically, developmental research has found that infant participants in their first year can also make animacy inferences (Csibra, 2008; Csibra et al., 1999; Gergely et al., 1995) and distinguish causal from non-causal events (Leslie & Keeble, 1987; Oakes, 1994; Oakes & Cohen, 1990) in the movement of simple shapes. Therefore, even infants that have not yet acquired language can identify *who did what to whom* in visual scenes, suggesting that thematic roles may originate in perceptual processing systems.

However, language and vision have traditionally been treated as separate cognitive processes (Fodor, 1983), with theories of language rarely considering the potential contribution of visual processing (Levelt, 1989; Pickering & Garrod, 2013; van Gompel & Pickering, 2007). Instead, a plethora of psycholinguistic research has implicated general statistical learning abilities as an essential component of language acquisition (e.g., Aslin et al., 1998; Cleeremans & McClelland, 1991; Saffran et al, 1996). However, without using thematic roles, connectionist models that implement statistical learning have been unable to account for how language learners make abstract generalisations from their input. For instance, Simple Recurrent Network (SRN) models trained with *John loves Mary* would consider *Mary loves John* as an ungrammatical sentence without additional training. To address this problem, Chang (2002) developed a connectionist framework called the Dual-path model, consisting of a *meaning pathway* and a *sequencing pathway*. The meaning pathway uses variables to encode bindings between thematic roles and concepts (AGENT-JOHN,

ACTION-LOVES, PATIENT-MARY), while the sequencing pathway attempts to learn how these role variables can be arranged in the grammar of the input language (e.g., English active transitives: AGENT-ACTION-PATIENT) without knowing the concepts that are attached to them. By using thematic roles as a fast variable-binding mechanism to map meaning into sentences, the Dual-path model can generalise nouns to novel roles and sentence structures and has been able to account for a range of findings in language research (Chang, 2009; e.g., Chang et al., 2006; Fitz & Chang, 2018; Twomey et al., 2014).

The Dual-path model suggests that fast role-concept bindings are required to generalise and achieve systematic language abilities. However, these fast binding variables are not easily derived from the properties of generic neural mechanisms that support learning, but they may have evolved separately for another cognitive function (Chang, 2002; Chang et al., 2006). Many studies have found that viewers can track a small number of targets in parallel (e.g., Pylyshyn & Storm, 1988) and that tracking the relative movement of these objects allows both infant and adult observers to detect causal interactions, such as a one circle chasing another circle around the screen (Frankenhuis et al., 2013; Gao et al., 2009). Therefore, it is possible that the ability to quickly bind and update location pointers with visual objects and extract relational features from their movement can provide a neural basis for the thematic role variables proposed in the Dual-path model of language.

In this thesis, I have provided evidence that multiple object tracking can support thematic roles for language and can have a direct and observable effect on sentence production. Before discussing the implications of these findings, I

will briefly review the methodology and results of the experiments presented in chapters two and three.

A total of nine experiments were presented, where adult participants completed a multiple push tracking task that required them to monitor visually identical objects that moved around the screen in random unpredictable patterns. This task was adapted from the multiple object tracking paradigm (Pylyshyn & Storm, 1988), which has been widely used to study location tracking and other visual processing phenomena. The precise configuration and display parameters of this task varied between each experiment, but the core procedure remained consistent. At the start of the trial, nine white circles would appear on the screen moving in random patterns. After three seconds, the objects would stop moving and a causal push event would occur between a pair of circles. These push events were an implementation of Michotte's (1946) launching display, where an agent circle moved directly towards a stationary patient circle and pushed it away (see figure 1). The number of push events that were presented in each trial varied between the nine experiments, but there were never more than three pushes in total. The participants were instructed to track the circles from these push events throughout a subsequent period of random movement, before all the circles stopped moving and the test frame was shown. Here, a subset of the objects changed colour, which always included circles that were involved in a push event and random foil objects. The participants then described the interactions that occurred between these circles in a specific sentence form (e.g., an active transitive, such as *red pushed blue*). Since the circles were identical, the participants were required to retain their role-related

features as they moved around randomly to produce an accurate description at test.

Chapter two presented the results of three studies that investigated whether adult participants could extract thematic roles from multiple push events, bind these roles to their referents as they moved around randomly among visually identical distractors, and then produce an accurate transitive description of the interactions at test (e.g., *red pushed blue*). These experiments also examined how manipulations of two core features of multiple object tracking affected sentence production accuracy. First, since object tracking has a limited capacity (Alvarez & Franconeri, 2007; Pylyshyn & Storm, 1988), the number of agents and patients that needed to be tracked was varied by including trials with one, two, or three push events. It was estimated that most viewers would have a tracking capacity of two events (or four targets). If role-related features are maintained in the object tracking system, then the capacity limitations of this system should influence accuracy in describing the interactions. Second, object tracking is widely considered to be a parallel ability, as viewers can monitor a small number of objects simultaneously without needing to use overt attention or active gaze switches (Alvarez & Cavanagh, 2005; Cavanagh & Alvarez, 2005; Howe et al., 2010; Oksama & Hyönä, 2016; Pylyshyn, 1989; Yantis, 1992). This was manipulated in different ways across the three experiments. In experiment one, the participants were unable to depend on overt shifts of attention to complete the task, as they were required to fix their gaze on a central cross and were monitored with eye-tracking. Experiment two attempted to also limit covert switches of attention by having the participants perform a secondary task to capture their focal attention; specifically, they provided a speeded key-press

whenever a static object in the centre of the display changed colour. Finally, following evidence that viewers can continue to track multiple targets during occlusion (Flombaum et al., 2008; Horowitz et al., 2006; Scholl & Pylyshyn, 1999), the third experiment had the objects move less randomly but disappear in close proximity to each other. This provided a stronger test of whether thematic roles can be maintained by the object tracking system, as it forced the participants to continuously track all the targets to accommodate periods when they became momentarily invisible.

Chapter two made three main discoveries. First, even though each study used different versions of the push tracking task that placed various demands on tracking abilities, the participants produced the correct agents and patients at above chance levels in all three experiments. There were significant differences in overall accuracy levels between the studies, but these were consistent with reports from the multiple object tracking literature. Specifically, concurrently performing an attentionally demanding task made tracking more difficult (see Tombu & Seiffert, 2008), with accuracy being reduced even further when the objects became invisible in close proximity (see Luu & Howe, 2015). The second main finding of chapter two was that role tracking accuracy was negatively affected by the number of pushes in the trial, indicating that the participants found it harder to track role-related information as they were required to support more role bindings. Specifically, the results suggest that the participants could track up to two pushes and started using attentional strategies for difficult trials with three events. This was consistent with previous multiple object tracking work showing that participants could track up to four separate objects in a task with similar parameters (Alvarez & Franconeri, 2007). The third main finding was

that accuracy in reporting agents was around 6% higher than patients. Since this was consistent across all three experiments and did not interact with any other variable, it is possible that this agent preference reflects the respective availability and salience of thematic role cues in the motion patterns of the objects (e.g., self propulsion; Luo & Baillargeon, 2005; Luo & Johnson, 2009). Collectively, these three findings suggest that the object tracking system can support the thematic roles of identical objects and can be used to map referents (e.g., *red*) into appropriate sentence positions in language.

In chapter three, the multiple push tracking task was used to investigate the relative clause asymmetry, following a plethora of evidence that subject-extracted relative clauses (SRCs) are easier to process than object-extracted relative clauses (ORCs) across many languages with different properties (Korean: Kwon et al., 2010; Dutch: Mak et al., 2002; Japanese: Miyamoto & Nakamura, 2003; English: Wanner & Maratsos, 1978). Previous theories have attributed this asymmetry to differences in the resources required to process the syntax of the structures (Gibson, 1998; Gordon et al., 2001; Just & Carpenter, 1992) or the amount of experience speakers have with these sentence forms based on their frequency in natural language (Levy, 2008; e.g., MacDonald & Christiansen, 2002). However, these models suggest that this processing asymmetry should vary with the word order and frequency distributions of different languages, so the cross-linguistic persistence of the SRC preference has yet to be fully explained. To address this, chapter three ran six experiments using the multiple push tracking task to investigate whether differences in the complexity of the role-referent bindings of events can produce a bias in sentence production.

Experiments four, five and six, compared description accuracy in trials where the event configurations matched either an SRC or an ORC. In this version of the push tracking task, there were always two push events with an overlapping circle that appeared in both pushes. This overlapping circle appeared as the agent in both events for the SRC-target conditions, but it was the patient of the first push and the agent of the second in the ORC-target conditions. This allowed the participants to describe these interactions using a relative clause when four of the circles were given different colours at test, after a period of random movement. The participants were instructed to use specific types of sentences to describe the interactions, which varied between each study. In experiment four, the events needed to be described using an appropriate relative clause, which could be either an SRC (e.g., *the blue that pushed red pushed green*) or an ORC (e.g., *the blue that red pushed pushed green*). In this study, description accuracy was significantly higher in the SRC-target trials compared to the ORC-target trials. While this demonstrates that the relative clause asymmetry can be observed in sentence production using a push tracking task, difficulties in using the ORC structure to describe an interaction between three animate targets could have contributed to the difference in accuracy. To account for this possibility, experiment five tested a new sample with the same stimuli, where the participants were instructed to describe the events using either an SRC or a passive relative clause (e.g. *the red that was pushed by blue pushed green*), as English speakers appear to prefer these structures over ORCs when describing visual events with animate referents (Gennari & MacDonald, 2009; Gennari et al., 2012). This study replicated the large significant difference in accuracy observed in experiment four, suggesting

that the production asymmetry was not due to the compatibility of the sentence structures being used to describe the interactions. However, since the participants described each event type with different structures at test, the results could still have been affected by linguistic factors. In experiment six, the viewers were presented with only one of the push events at test (alongside two random foils) and described both the SRC and ORC conditions with active transitive sentences (e.g, *blue pushed red*) rather than a complex relative clause. Although the overall difference was smaller, accuracy for events matching SRCs remained significantly higher than the trials matching ORCs. Thus, even when production differences were removed, there was a bias for SRC events. Since the conditions differed only in their role-referent bindings, the results suggest that thematic role consistency in the multiple push tracking task can have a direct effect on sentence production.

Experiments seven, eight and nine provided a more in-depth investigation into how thematic role consistency affects sentence production by using different event configurations and sentence structures. Rather than only presenting events directly corresponding to specific relative clause structures, the studies focused on the consistency of the agents or patients throughout the sequence. In experiment four, the agent from the first push event would reappear in the second push, either playing the agent again (same role) or switching to being a patient (different role), and the participants described one of these pushes in an active transitive sentence. The study found that role repetition led to significantly greater description accuracy compared to a control condition with no overlapping circles between the two pushes, but switching thematic roles between the pushes did not significantly affect it compared to the control trials.

Experiment eight aimed to replicate this study when first patient was the circle that always participated in both events. Again, this overlapping patient either played the same role in the second push or switched to being an agent. The results of this study matched those of experiment seven; production accuracy was significantly higher in the same role trials compared to the control condition, but switching roles did not hinder performance. Finally, experiment nine used the same stimuli as experiment eight, but required the participants to describe these events in a passive transitive (e.g., *the green was pushed by blue*) instead of an active transitive to provide a final test that these effects exist independently of the specific structures used to describe the interactions. As in the previous studies, significantly higher description accuracy was observed in scenes where the overlapping referent was consistently the patient, but was unaffected by role switches between the pushes.

Collectively, the experiments presented in chapter three showed that production accuracy was higher for visual events corresponding to SRCs than ORCs, which was replicated using a range of different sentence structures. This effect appears to be due to thematic role consistency, as similar differences in accuracy were observed using other event configurations, such as interactions with consistent patient referents. Importantly, there were no linguistic or conceptual cues that could support these differences, as the task involved multiple identical objects labelled with randomly assigned colours at test. Therefore, the visual nature of the task and its insensitivity to linguistic manipulations suggests that non-linguistic processes associated with tracking role-referent bindings were driving the observed SRC bias, particularly, whether the referents played consistent roles throughout the sequence.

To summarise these findings, the nine experiments presented in this work showed that (1) thematic roles can be extracted from push events involving identical objects, (2) these roles can be bound to referents and maintained during periods of random movement among distractor objects, (3) these bindings can be maintained for multiple push events simultaneously without depending on overt focal attention, (4) role-referent consistency can create production biases such as an asymmetry between subject-extracted and object-extracted relative clauses, (5) role bindings are enhanced when a referent appears in the same thematic role across multiple events, and (6) the thematic role configurations of an event can directly affect sentence accuracy independent of the precise structures produced to describe the events. Therefore, the primary deduction to be made of from this thesis is that multiple object tracking can support thematic roles for language.

These results have a number of implications. First, this research connects thematic roles in language to perception studies reporting that participants can compute these roles in spatial processing (Scholl & Tremoulet, 2000). Visual research has identified a range of motion features that observers can use to identify animate agents and causal interactions independent of the physical nature of the entities involved, such as sudden changes in speed and direction (Tremoulet & Feldman, 2000), the agent's angle of approach towards the patient (Gao et al., 2009), and spatial and temporal correlations in the objects' movement (Leslie & Keeble, 1987; Oakes & Cohen, 1990). However, linguistic and concept-based treatments of thematic roles have been unable to reach a consensus on the number of roles that exist or the features that allow them to be uniquely and reliably identified (see Dowty, 1991). Consequently, some theorists

have attempted to explain language development without relying on pre-existing thematic roles (Tomasello, 2000). However, thematic roles cannot be entirely discarded, as there is evidence that these roles exist in language and may be important to acquisition; for instance, experiments have shown that their arrangement in sentences can affect comprehension difficulty and production choices (Chang, Bock, & Goldberg, 2003; Ferreira, 2003) and the Dual-path model relies on thematic role variables to form abstract generalisations (Chang, 2002; Chang et al., 2006). The present work suggests that the solution to this problem is a vision-based theory of thematic roles, as the participants were able to use psychophysical motion features to identify agents and patients, bind these roles to specific objects, and then use these bindings to produce accurate sentences. Thus, the motion features in visual processing offer concrete role-features that allow speakers to map referent objects into language, without needing additional linguistic cues or abstract conceptual features. However, this does not disregard the abundant evidence that conceptual knowledge and experience with thematic roles can influence their assignment in language (e.g., Chang et al., 2003; Ferretti et al., 2001; Knoeferle, Crocker, Scheepers, & Pickering, 2005; McRae, Spivey-Knowlton, & Tanenhaus, 1998), but demonstrates that this knowledge may not be the primary basis of thematic roles since they can be derived purely by tracking motion cues without conceptual information about their fillers.

A second outcome of this work is that it provides empirical support for the Dual-path model of language (Chang, 2002; Chang et al., 2006). This connectionist framework uses fast role-concept bindings to encode the message of the sentence. Since this mechanism has no direct neural correlates, Chang

proposed that these role variables are derived from visual processing that allows object identity and object location to be processed separately (Goodale & Milner, 1992; Milner & Goodale, 2008) and then rapidly integrated (Karnath, 2001). Evidence to support this hypothesis was provided throughout the nine studies, particularly those in chapter 2. Specifically, the participants were able to successfully complete the multiple push tracking task, which required them to bind role-related features to objects to produce accurate descriptions of the push events at test. The Dual-path model also predicts that manipulations to role-referent bindings in visual events should directly affect sentence production. This was demonstrated in chapter three, which showed that thematic role consistency can create biases in production, such as an asymmetry in the production of different relative clause sentences. These results also suggest that a visual version of the Dual-path model could be implemented. Specifically, it may be possible to replace the thematic role variables with attentional pointers and a set of heuristics for encoding object motion features, and then train the model with visual scenarios paired with sentences. This has the potential to provide further insights into the relationship between vision and language and would allow the model to explore data obtained from visual paradigms, such as the predictive looking behaviour in the visual world tasks used to study language processing (see Huettig et al., 2011)

A more global implication of this research is the existence of a close link between visual processing and language. While some have argued that the representations and mechanisms of language are isolated from other aspects of cognition (Fodor, 1983), there now exists a substantial amount of experimental work reporting that visual input can exert a real-time influence over language

processing (see Anderson et al., 2011). This thesis contributes to this line of evidence by demonstrating that thematic roles can be extracted and supported by the object tracking system in a way that directly affects sentence production. The studies also show that this relationship has the potential to explain important phenomena in language processing, such as the persistent relative clause asymmetry that has been observed in many languages. While the present evidence suggests that powerful location tracking mechanisms that originally evolved in the visual system might also play a central role in explaining the many of the features of human language, further research is needed to identify how closely these subsystems are intertwined.

In summary, this thesis has provided evidence that manipulations of purely spatial factors in a multiple object task can affect accuracy in sentence production. Even under difficult tracking conditions, viewers can retain the thematic roles of objects from multiple events and a number of variables can affect this ability. In particular, accuracy is hindered by an increase in the number of the events that need to be tracked and the availability of attention, but it is also enhanced when the referents play the same roles in multiple events. These findings provide support for connectionist models that use fast role-concept bindings by showing that this ability may arise from perceptual processing. It also provides a potential solution to the limitations of previous language-based accounts of thematic roles. However, the research presented in this thesis cannot be consolidated within a modular view of cognition as it suggests that cognition may not be as definitively segmented as typically assumed. Instead, the primary conclusion to draw from this data is that multiple object tracking can support thematic role features for language.

## REFERENCES

- Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. (2006). Multiple-target tracking: A role for working memory? *The Quarterly Journal of Experimental Psychology*, *59*(6), 1101–1116. <https://doi.org/10/fcns38>
- Allen, R., McGeorge, P., Pearson, D., & Milne, A. B. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology*, *18*(3), 337–347. <https://doi.org/10/djhwpz>
- Altmann, G. T. (1999). Thematic role assignment in context. *Journal of Memory and Language*, *41*(1), 124–145. <https://doi.org/10/b58f8n>
- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643. <https://doi.org/10/bnzbxt>
- Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14. <https://doi.org/10/fndcq6>
- Alvarez, G. A., & Scholl, B. J. (2005). How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. *Journal of Experimental Psychology: General*, *134*(4), 461–476. <https://doi.org/10/ctcdg2>
- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(02), 239–273. <https://doi.org/10/f62n8w>
- Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2014). Child language acquisition: Why universal grammar doesn't help. *Language*, *90*(3), e53–e90. <https://doi.org/10/gdf46z>

- Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, *137*(2), 181–189. <https://doi.org/10/cr5tfw>
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. <https://doi.org/10/bmr6gd>
- Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, *10*(8), 949–963. <https://doi.org/10/fv2q48>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10/gcm4wc>
- Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, *26*(4), 313–331. <https://doi.org/10/b6fpvk>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10/gcrnkx>
- Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M. J., Hénaff, M.-A., Michèl, F., & Barton, J. J. (2001). Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron*, *32*(6), 985–995. <https://doi.org/10/cf6czw>
- Berry, D. S., & Springer, K. (1993). Structure, motion, and preschoolers' perceptions of social causality. *Ecological Psychology*, *5*(4), 273–283.

- Berry, D. S., Misovich, S. J., Kean, K. J., & Baron, R. M. (1992). Effects of disruption of structure and motion on perceptions of social causality. *Personality and Social Psychology Bulletin*, *18*(2), 237–244.
- Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, *9*(7), 9–9. <https://doi.org/10/fcd32p>
- Blakemore, S.-J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., ... Decety, J. (2001). How the brain perceives causality: An event-related fMRI study. *Neuroreport*, *12*(17), 3741–3746. <https://doi.org/10/dkg3tr>
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. <https://doi.org/10/c4gk24>
- Bock, J. K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177.
- Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*, 1–39.
- Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132. <https://doi.org/10/fptgds>
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2), 96–107.
- Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733. <https://doi.org/10/cjznxh>
- Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., & Laka, I. (2010). Subject relative clauses are not universally easier to process:

- Evidence from Basque. *Cognition*, 115(1), 79–92.  
<https://doi.org/10/bpj64p>
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9(7), 349–354.  
<https://doi.org/10/dcxnpr>
- Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26(5), 609–651.
- Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, 61(3), 374–397. <https://doi.org/10/ftwhs5>
- Chang, F., Bock, J. K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, 90(1), 29–49. <https://doi.org/10/d95nsb>
- Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272. <https://doi.org/10/cxcrx7>
- Chen, B., Ning, A., Bi, H., & Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychologica*, 129(1), 61–65. <https://doi.org/10/dm3nw8>
- Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics*, 74(5), 784–802. <https://doi.org/10/f33c66>
- Chomsky, N. (1959). Review of verbal behavior. *Language*, 35(1), 26–58.  
<https://doi.org/10/cs73sr>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2), 157–205. <https://doi.org/10/bwhgnx>
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120(3), 235. <https://doi.org/10/bjhztr>
- Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite state automata and simple recurrent networks. *Neural Computation*, 1(3), 372–381. <https://doi.org/10/d2xrm3>
- Clemens, L. E., Coon, J., Mateo Pedro, P., Morgan, A. M., Polinsky, M., Tandet, G., & Wagers, M. W. (2015). Ergativity and the complexity of extraction: A view from Mayan. *Natural Language & Linguistic Theory*, 33(2), 417–467. <https://doi.org/10/gdthtw>
- Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development*, 21(4), 713–731. <https://doi.org/10/cnhzc4>
- 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. <https://doi.org/10/gdh8xw>
- Conway, C. M., Baurnschmidt, A., Huang, S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356–371. <https://doi.org/10/ctdhhz>
- Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, 107(2), 705–717. <https://doi.org/10/c84vb7>

- Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of 'pure reason' in infancy. *Cognition*, 72(3), 237–267.
- Descartes, R. (1637). *Discourse on the method of rightly conducting one's reason and of seeking truth in the sciences*. Hackett Publishing Company.
- Dessalegn, B., & Landau, B. (2013). Interaction between language and vision: It's momentary, abstract, and it develops. *Cognition*, 127(3), 331–344. <https://doi.org/10/f4xj2x>
- Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, 25(2), 108–127. <https://doi.org/10/d6gnhx>
- Dittrich, W. H., & Lea, S. E. (1994). Visual perception of intentional motion. *Perception*, 23(3), 253–268. <https://doi.org/10/b8f2gw>
- Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67(3), 547. <https://doi.org/10/bg3ktc>
- Drew, T., Võ, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: Sustained inattentive blindness in expert observers. *Psychological Science*, 24(9), 1848–1853. <https://doi.org/10/m9s>
- Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, 2(7), 254–262. <https://doi.org/10/dsg5vc>
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. <https://doi.org/10/fwxc7w>

- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71–99.  
<https://doi.org/10/c6xzzk>
- Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, 108(1), 201–209.  
<https://doi.org/10/cg43vv>
- Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, 10(4), 19–19.  
<https://doi.org/10/c2tk7x>
- Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics*, 69(4), 567–577. <https://doi.org/10/d8npg8>
- Feria, C. S. (2013). Speed has an effect on multiple-object tracking independently of the number of close encounters between targets and distractors. *Attention, Perception, & Psychophysics*, 75(1), 53–67.  
<https://doi.org/10/f4h3jc>
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33(6), 715.
- Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying “that” is not saying “that” at all. *Journal of Memory and Language*, 48(2), 379–398.
- Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 453–469). Oxford, UK: Oxford University Press.

- Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44(4), 516–547. <https://doi.org/10/fsz5m8>
- Fillmore, C. J. (1967). The case for case. In E. Bach & R. J. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). New York: Holt, Rinehart & Winston.
- Fitz, H., & Chang, F. (2017). Meaningful questions: The acquisition of auxiliary inversion in a connectionist model of sentence production. *Cognition*, 166, 225–250. <https://doi.org/10/gd59wh>
- Fitz, H., & Chang, F. (2018). Sentence-level ERP effects as error propagation: A neurocomputational model. *PsyArXiv*. <https://doi.org/10/gdvmd9>
- Fitz, H., Chang, F., & Christiansen, M. H. (2011). A connectionist account of the acquisition and processing of relative clauses. In E. Kidd (Ed.), *The acquisition of relative clauses: Processing, typology and function* (pp. 39–60). Amsterdam: John Benjamins Publishing Company.  
<https://doi.org/10.1075/tilar.8.04fit>
- Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition*, 107(3), 904–931. <https://doi.org/10/dtj5nh>
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1), 3–71.

- Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 297–316.  
<https://doi.org/10/cpk82m>
- Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, 21(7), 920–925. <https://doi.org/10/d66qbw>
- Franconeri, S. L., Lin, J. Y., Enns, J. T., Pylyshyn, Z. W., & Fisher, B. (2008). Evidence against a speed limit in multiple-object tracking. *Psychonomic Bulletin & Review*, 15(4), 802–808. <https://doi.org/10/d9sw3j>
- Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, & Psychophysics*, 74(4), 691–702. <https://doi.org/10/fzvmmk>
- Frankenhuis, W. E., House, B., Clark Barrett, H., & Johnson, S. P. (2013). Infants' perception of chasing. *Cognition*, 126(2), 224–233.  
<https://doi.org/10/f4kjxx>
- Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., & Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive Brain Research*, 24(1), 41–47. <https://doi.org/10/b6dcgg>
- Galazka, M., & Nyström, P. (2016). Infants' preference for individual agents within chasing interactions. *Journal of Experimental Child Psychology*, 147, 53–70. <https://doi.org/10/f3r3zr>
- Gao, T., & Scholl, B. J. (2011). Chasing vs. stalking: Interrupting the perception of animacy. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 669–684. <https://doi.org/10/fhf32z>

- Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, *59*(2), 154–179. <https://doi.org/10/b25zm6>
- Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*(2), 161–187. <https://doi.org/10/fq7g5r>
- Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23. <https://doi.org/10/bf3wkb>
- Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, *65*(2), 141–176. <https://doi.org/10/f364j5>
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193. <https://doi.org/10/fvjkg5>
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.
- Gibson, E., & Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, *28*(1-2), 125–155. <https://doi.org/10/b2x7gg>
- Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements: Cognitive neuroscience. *Nature Reviews Neuroscience*, *4*(3), 179–192. <https://doi.org/10/c25gzb>
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

- Gordon, P. C., & Hendrick, R. (2005). Relativization, ergativity, and corpus frequency. *Linguistic Inquiry*, 36(3), 456–463. <https://doi.org/10/d3p89g>
- Gordon, P. C., & Lowder, M. W. (2012). Complex sentence processing: A review of theoretical perspectives on the comprehension of relative clauses: Review of complex sentence processing. *Language and Linguistics Compass*, 6(7), 403–415. <https://doi.org/10/gcrnmp>
- Gordon, P. C., Hendrick, R., & Johnson, M. L. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1411. <https://doi.org/10/fttx89>
- Gordon, P. C., Hendrick, R., & Johnson, M. L. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51(1), 97–114. <https://doi.org/10/fwbmwx>
- Gordon, P. C., Hendrick, R., Johnson, M. L., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 32(6), 1304–1321. <https://doi.org/10/c4bf77>
- Green, C., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, 101(1), 217–245. <https://doi.org/10/b5mw6c>
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29(2), 261–290. <https://doi.org/10/cxzhr>
- Gruber, J. S. (1965). *Studies in lexical relations*. (PhD thesis). Massachusetts Institute of Technology.

- Guilfoyle, E., Hung, H., & Travis, L. (1992). Spec of IP and spec of VP: Two subjects in austronesian languages. *Natural Language & Linguistic Theory*, 10(3), 375–414. <https://doi.org/10/c8z4kh>
- Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10/b2t3v6>
- Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, 111(2), 151–167. <https://doi.org/10/bmk9c5>
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243–259. <https://doi.org/10/ftcck7>
- Hidaka, S., Teramoto, W., & Sugita, Y. (2015). Spatiotemporal processing in crossmodal interactions for perception of the external world: A review. *Frontiers in Integrative Neuroscience*, 9. <https://doi.org/10/gccxjf>
- Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, 203(3), 88–97. <https://doi.org/10/ftqv2r>
- Holcombe, A. O., & Chen, W.-Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition*, 123(2), 218–228. <https://doi.org/10/dtfx7p>
- Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review*, 13(3), 516–523. <https://doi.org/10/d98nfd>

- Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, *69*(2), 172–184. <https://doi.org/10/dxkrp2>
- Howe, P. D. L., & Holcombe, A. O. (2012). The effect of visual distinctiveness on multiple object tracking performance. *Frontiers in Psychology*, *3*. <https://doi.org/10/gcrnmw>
- Howe, P. D. L., Horowitz, T. S., Akos Morocz, I., Wolfe, J. M., & Livingstone, M. S. (2009). Using fMRI to distinguish components of the multiple object tracking task. *Journal of Vision*, *9*(4), 10–10. <https://doi.org/10/d3pfqx>
- Howe, P. D. L., Incledon, N. C., & Little, D. R. (2012). Can attention be confined to just part of a moving object? Revisiting target-distractor merging in multiple object tracking. *PLoS ONE*, *7*(7), e41491. <https://doi.org/10/f35j7c>
- Howe, P. D. L., Pinto, Y., & Horowitz, T. S. (2010). The coordinate systems used in visual tracking. *Vision Research*, *50*(23), 2375–2380. <https://doi.org/10/bqvj2d>
- Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*(1), 3–27. <https://doi.org/10/b5sk2h>
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. <https://doi.org/10/cx7k4d>
- Huff, M., Meyerhoff, H. S., Papenmeier, F., & Jahn, G. (2010). Spatial updating of dynamic scenes: Tracking multiple invisible objects across viewpoint changes. *Attention, Perception, & Psychophysics*, *72*(3), 628–636. <https://doi.org/10/bvf58p>

- Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. W. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition*, 18(9), 1368–1391. <https://doi.org/10/c4md8n>
- Humphreys, G. F., Mirković, J., & Gennari, S. P. (2016). Similarity-based competition in relative clause production and comprehension. *Journal of Memory and Language*, 89, 200–221. <https://doi.org/10/gctpzx>
- Hyman, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E., & Caggiano, J. M. (2009). Did you see the unicycling clown? Inattentional blindness while walking and talking on a cell phone. *Applied Cognitive Psychology*, 24(5), 597–607. <https://doi.org/10/bfzk76>
- Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, 9(4), 1–1. <https://doi.org/10/d4cz9d>
- Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Jackendoff, R. S. (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, 18(3), 369–411. <https://doi.org/10/dkhv4k>
- Janciauskas, M., & Chang, F. (2017). Input and age-dependent variation in second language learning: A connectionist account. *Cognitive Science*. <https://doi.org/10/gcrnm4>
- Jäger, L., Chen, Z., Li, Q., Lin, C.-J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, 79-80, 97–120. <https://doi.org/10/gctpzv>

- Jessop, A., & Chang, F. (2018). Thematic role information is maintained in the visual object tracking system. *Accepted Manuscript*.
- Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's  $R^2_{\text{GLMM}}$  to random slopes models. *Methods in Ecology and Evolution*, 5(9), 944–946. <https://doi.org/10/f6j4dj>
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Human Perception*, 265.
- Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, 101(1), 1–42. <https://doi.org/10/fbrvrq>
- Karnath, H.-O. (2001). New insights into the functions of the superior temporal cortex. *Nature Reviews Neuroscience*, 2(8), 568–576. <https://doi.org/10/d9r3rc>
- Karnath, H.-O., Rüter, J., Mandler, A., & Himmelbach, M. (2009). The anatomy of object recognition: Visual form agnosia caused by medial occipitotemporal stroke. *Journal of Neuroscience*, 29(18), 5854–5862.
- Keane, B., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology*, 52(4), 346–368. <https://doi.org/10/ff342b>
- Kemmerer, D. (2012). The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's Area. *Language and Linguistics Compass*, 6(1), 50–66.

- Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171–184.  
<https://doi.org/10/c9ncqh>
- Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children’s comprehension of syntax. *Child Development*, *87*(1), 184–193.  
<https://doi.org/10/f79wz7>
- Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children’s processing of relative clauses. *Language and Cognitive Processes*, *22*(6), 860–897. <https://doi.org/10/cdgbzq>
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. <https://doi.org/10/dq5z6b>
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. <https://doi.org/10/fqnf5c>
- Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559–1571.  
<https://doi.org/10/dd859w>
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

- Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, 72(1), 5–18.
- Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin & Review*, 15(6), 1135–1140.  
<https://doi.org/10/fwfmtc>
- Kwon, N., Gordon, P. C., Lee, Y., Kluender, R., & Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of prenominal relative clauses in Korean. *Language*, 86(3), 546–582. <https://doi.org/10/dpnqzk>
- Lalanne, C., & Lorenceau, J. (2004). Crossmodal integration for perception and action. *Journal of Physiology-Paris*, 98(1-3), 265–279.  
<https://doi.org/10/c2thgg>
- Lappin, S., & Fox, C. (2015). *The handbook of contemporary semantic theory* (2nd ed.). Sussex, UK: John Wiley & Sons Ltd.
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3), 265–288. <https://doi.org/10/dc4dh2>
- Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, 2(1), 10–18. <https://doi.org/10/dxd9p7>
- Levelt, W. J. (1989). *Speaking: From intention to articulation*. ACL. MIT Press Series in Natural-Language Processing. MIT Press, Cambridge, Massachusetts.

- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10/dtk7j>
- Levy, R., & Gibson, E. (2013). Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. *Frontiers in Psychology*, 4. <https://doi.org/10/gctmns>
- Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language and Cognition*, 8(3), 346–368. <https://doi.org/10/gd579s>
- Lin, C. J. C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. *Proceedings of the 25th west coast conference on formal linguistics* (pp. 254-260). Somerville, MA: Cascadilla Proceedings Project.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494–1502. <https://doi.org/10/gbsd4m>
- 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. <https://doi.org/10/cwspvm>
- Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, 12(1), 142–149. <https://doi.org/10/fp976z>
- Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, 58(4), 441–486. <https://doi.org/10/fpcswd>

- Luu, T., & Howe, P. D. L. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, & Psychophysics*, 77(6), 1919–1929. <https://doi.org/10/f7mh82>
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4. <https://doi.org/10/gbfpt3>
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54. <https://doi.org/10/fs62r9>
- Mack, A., & Rock, I. (1998). *Inattention blindness*. MIT press Cambridge, MA.
- MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29(2), 95–141.
- Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, 91, 5–27. <https://doi.org/10/gd4q4q>
- Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47(1), 50–68. <https://doi.org/10/d944mc>
- Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, 54(4), 466–490. <https://doi.org/10/fc7hbx>
- Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, 37(3), 243–282.
- McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, 3(6), 1417–1429. <https://doi.org/10/c934h6>

- McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic Roles as Verb-specific Concepts. *Language and Cognitive Processes*, 12(2-3), 137–176.  
<https://doi.org/10/dp8v4q>
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283–312.
- Meyerhoff, H. S., Huff, M., & Schwan, S. (2013). Linking perceptual animacy to attention: Evidence from the chasing detection paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 39(4), 1003. <https://doi.org/10/f46kmz>
- Michotte, A. (1946). *The perception of causality*. Basic Books.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3), 774–785.
- Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62(1), 302–331.  
<https://doi.org/10/dfhrmx>
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1. <https://doi.org/10/fm6g7b>
- Miyamoto, E. T., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In *Proceedings of the 22nd West Coast conference on formal linguistics* (Vol. 22, pp. 342–355). Cascadilla Press Somerville, MA.

- Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, 67(6), 949.
- Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattention blindness. *Psychological Science*, 12(1), 9–17. <https://doi.org/10/fmbz72>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining  $R^2$  from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10/f4pkjx>
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination  $R^2$  and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134), 20170213. <https://doi.org/10/gddpnq>
- Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, 30(6), 869. <https://doi.org/10/cs7q44>
- Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, 5(2), 193–207. <https://doi.org/10/dqgznw>
- Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, 11(5), 631–671. <https://doi.org/10/ctzjnn>

- Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology*, 56(4), 237–283. <https://doi.org/10/ccphg3>
- Oksama, L., & Hyönä, J. (2016). Position tracking and identity tracking are separate systems: Evidence from eye movements. *Cognition*, 146, 393–409. <https://doi.org/10/f74g45>
- Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, 49(3), 197–237. <https://doi.org/10/dj32n7>
- Osgood, C. E., & Bock, J. K. (1977). Salience and sentencing: Some production principles. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory* (pp. 89–140). Hillsdale, NJ: Lawrence Erlbaum.
- Papafragou, A. (2015). The representation of events in language and cognition. *The Conceptual Mind: New Directions in the Study of Concepts*, 327.
- Pearson, M. (1998). Predicate raising and “VOS” order in malagasy. In I. Paul (Ed.), *The structure of Malagasy* (Vol. 2, pp. 94–110). Los Angeles: UCLA Department of Linguistics.
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39, 633–651.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329–347. <https://doi.org/10/f45frn>
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32(1), 65–97. <https://doi.org/10/dsztgv>

- Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, 11(7), 801–822. <https://doi.org/10/ddgdjd>
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3), 179–197. <https://doi.org/10/bt3rft>
- Pylyshyn, Z. W., Burkell, J., Fisher, B., Sears, C., Schmidt, W., & Trick, L. M. (1994). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology*, 48(2), 260–282. <https://doi.org/10/bb63sj>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23. <https://doi.org/10/c7kj9f>
- Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science*, 6(1), 77–97. <https://doi.org/10/dz4qx9>
- Rochat, P., Morgan, R., & Carpenter, M. (1997). Young infants' sensitivity to movement information specifying social causality. *Cognitive Development*, 12(4), 537–561. <https://doi.org/10/b2h2dc>
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348–379. <https://doi.org/10/ctrs5f>
- Ryokai, K., Farzin, F., Kaltman, E., & Niemeyer, G. (2013). Assessing multiple object tracking in young children using a game. *Educational Technology Research and Development*, 61(2), 153–170. <https://doi.org/10/f4sn4n>

- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.  
<https://doi.org/10/fcqz9d>
- Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, 3(1), 2–2. <https://doi.org/10/bs4gdr>
- Schlottmann, A., Ray, E. D., Mitchell, A., & Demetriou, N. (2006). Perceived physical and social causality in animated motions: Spontaneous reports and ratings. *Acta Psychologica*, 123(1-2), 112–143.  
<https://doi.org/10/c6smbn>
- Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment. *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*, 4629.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38(2), 259–290. <https://doi.org/10/c6ccs9>
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299–309. <https://doi.org/10/dvnhr2>
- Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, 80(12), 159–177. <https://doi.org/10/ch9wn3>
- Schriefers, H., Friederici, A. D., & Kuhn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 34(4), 499–520. <https://doi.org/10/b7f6v6>

- Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Experimentale*, 54(1), 1.
- Sekuler, R., McLaughlin, C., & Yotsumoto, Y. (2008). Age-related changes in attentional tracking of multiple moving objects. *Perception*, 37(6), 867–876. <https://doi.org/10/cr7vgv>
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13(3), 272–281.
- Simons, D. J. (2010). Monkeying around with the gorillas in our midst: Familiarity with an inattentional-blindness task does not improve the detection of unexpected events. *I-Perception*, 1(1), 3–6. <https://doi.org/10/dxwwh8>
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28(9), 1059–1074. <https://doi.org/10/gdh8td>
- Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, 13(2), 113–142. <https://doi.org/10/d6wq4q>
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116(1), 71–86. <https://doi.org/10/dm3wgs>
- Staub, A., Dillon, B., & Clifton, C. (2017). The matrix verb as a source of comprehension difficulty in object relative sentences. *Cognitive Science*, 41, 1353–1376. <https://doi.org/10/gbgkqq>

- Straube, B., Wolk, D., & Chatterjee, A. (2011). The role of the right parietal lobe in the perception of causality: A tDCS study. *Experimental Brain Research*, 215(3-4), 315–325. <https://doi.org/10/fcfjzd>
- Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, 12(1), 49–100. <https://doi.org/10/dwn9j3>
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. <https://doi.org/10/b6msqx>
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3), 209–253. <https://doi.org/10/fxgmt4>
- Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, 108(1), 1–25. <https://doi.org/10/b3qqr3>
- Tombu, M., & Seiffert, A. E. (2011). Tracking planets and moons: Mechanisms of object tracking revealed with a new paradigm. *Attention, Perception, & Psychophysics*, 73(3), 738–750. <https://doi.org/10/d5d9m4>
- Tran, A., & Hoffman, J. E. (2016). Visual attention is required for multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 42(12), 2103. <https://doi.org/10/gdh8tc>
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47(1), 69–90. <https://doi.org/10/bdqtdt>
- Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses.

*Journal of Memory and Language*, 53(2), 204–224.

<https://doi.org/10/bnxtjf>

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention.

*Cognitive Psychology*, 12(1), 97–136. <https://doi.org/10/bgz2qm>

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion

of a single object. *Perception*, 29(8), 943–951. <https://doi.org/10/dbqtgx>

Trick, L. M., Guindon, J., & Vallis, L. A. (2006). Sequential tapping interferes

selectively with multiple-object tracking: Do finger-tapping and tracking

share a common resource? *The Quarterly Journal of Experimental*

*Psychology*, 59(7), 1188–1195. <https://doi.org/10/bmc65j>

Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in

children: The “Catch the Spies” task. *Cognitive Development*, 20(3), 373–

387. <https://doi.org/10/d3n49z>

Twomey, K. E., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A

lexical distributional account of English locative verb class acquisition.

*Cognitive Psychology*, 73, 41–71. <https://doi.org/10/f6f78k>

van Buren, B., Gao, T., & Scholl, B. J. (2017). What are the underlying units of

perceived animacy? Chasing detection is intrinsically object-based.

*Psychonomic Bulletin & Review*. <https://doi.org/10/gch7d7>

Van Dyke, J. A. (2003). Distinguishing effects of structure and decay on

attachment and repair: A cue-based parsing account of recovery from

misanalyzed ambiguities. *Journal of Memory and Language*, 49(3), 285–

316. <https://doi.org/10/fbr6g4>

- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.  
<https://doi.org/10/frdwtb>
- van Gompel, R. P., & Pickering, M. J. (2007). Syntactic parsing. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 289–307). Oxford, UK: Oxford University Press.
- Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS ONE*, *8*(10), e77006. <https://doi.org/10/gdqbfh>
- Wagers, M. W., Borja, M. F., & Chung, S. (2018). Grammatical licensing and relative clause parsing in a flexible word-order language. *Cognition*, *178*, 207–221. <https://doi.org/10/gdthfm>
- Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–61). Cambridge, MA: Cambridge, MA: MIT Press.
- Ward, E. J., & Scholl, B. J. (2015). Inattention blindness reflects limitations on perception, not memory: Evidence from repeated failures of awareness. *Psychonomic Bulletin & Review*, *22*(3), 722–727. <https://doi.org/10/f7br4f>
- Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.
- Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*, *133*(1), 51–56. <https://doi.org/10/b4v2km>
- Wells, J., Christiansen, M. H., Race, D., Acheson, D., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and

- relative clause comprehension. *Cognitive Psychology*, 58(2), 250–271.  
<https://doi.org/10/fnvq23>
- White, P. A. (2014). Perceived causality and perceived force: Same or different? *Visual Cognition*, 22(5), 672–703. <https://doi.org/10/gcrnph>
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419.  
<https://doi.org/10/c9qkrd>
- Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, 136(1), 82–111. <https://doi.org/10/dxzc3x>
- Woods, A. J., Hamilton, R. H., Kranjec, A., Minhaus, P., Bikson, M., Yu, J., & Chatterjee, A. (2014). Space, time, and causality in the human brain. *NeuroImage*, 92, 285–297. <https://doi.org/10/f53cft>
- Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24(3), 295–340.  
<https://doi.org/10/dmz7g4>
- Young, M. E., & Sutherland, S. (2009). The spatiotemporal distinctiveness of direct causation. *Psychonomic Bulletin & Review*, 16(4), 729–735.  
<https://doi.org/10/c99fvd>
- Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, 16(5), 553–566. <https://doi.org/10/c3nmbt>