



Thematic role tracking difficulties across multiple visual events influences role use in language production

Andrew Jessop ^{a,b} and Franklin Chang ^c

^aSchool of Psychology, The University of Liverpool, Liverpool, UK; ^bMax Planck Institute for Psycholinguistics, Nijmegen, Netherlands;

^cDepartment of English Studies, Kobe City University for Foreign Studies, Kobe, Japan

ABSTRACT

Language sometimes requires tracking the same participant in different thematic roles across multiple visual events (e.g., *The girl that another girl pushed chased a third girl*). To better understand how vision and language interact in role tracking, participants described videos of multiple randomly moving circles where two push events were presented. A circle might have the same role in both push events (e.g., agent) or different roles (e.g., agent of one push and patient of other push). The first three studies found higher production accuracy for the same role conditions compared to the different role conditions across different linguistic structure manipulations. The last three studies compared a featural account, where role information was associated with particular circles, or a relational account, where role information was encoded with particular push events. These studies found no interference between different roles, contrary to the predictions of the featural account. The foil was manipulated in these studies to increase the saliency of the second push and it was found that this changed the accuracy in describing the first push. The results suggest that language-related thematic role processing uses a relational representation that can encode multiple events.

ARTICLE HISTORY

Received 29 May 2021

Accepted 26 November 2021

KEYWORDS

Language; multiple object tracking; causality; thematic roles; relative clauses

In language, speakers can describe visual events they have seen using a sentence like *the girl chased the dog*. This sentence is sufficient when there is only one girl and one dog in the visual context. However, if the event took place on a busy playground where there were several girls, then this sentence does not clearly identify which girl is doing the chasing. One way for the speaker to resolve this ambiguity is to use a relative clause (Fox & Thompson, 1990), such as *the girl that pushed the boy chased the dog*. This relative clause refers to a previous pushing event with a boy, which allows the listener to identify which girl did the chasing. But using a relative clause is only possible if both the speaker and listener are tracking possible referents in the visual scene and recording thematic roles for these referents across different events.

This ability involves the interface between visual systems and language. It is common to think of vision and language as being separate modules (Fodor, 1983). In this approach, visual processing encodes the scene and passes that information to

the language system. The language system might then assign thematic roles like agent and patient (e.g., *the girl pushed the boy* could be encoded with a propositional representation like PUSH(GIRL,BOY), where the first argument is the agent and the second argument is the patient). If there are multiple girls involved in multiple push events, then it is possible to use indexes to distinguish which girl is involved in each event (e.g., *the girl₁ that the other girl₂ pushed pushed a third girl₃* might have two propositions PUSH(GIRL₁,GIRL₃) + PUSH(GIRL₂,GIRL₁)). Since this modular account encodes thematic roles within the language system, we will refer to it as the *Linguistic* account of thematic role processing.

In contrast to the Linguistic account, there is a growing body of work using visual world eye-tracking experiments which argue against a strictly modular separation between vision and language (Altmann & Kamide, 2007; Knoeferle et al., 2005; Tanenhaus et al., 1995). In these studies, incremental processing of linguistic input causes moment-by-moment changes in eye movements in the visual system.

CONTACT Andrew Jessop  andrew.jessop@liverpool.ac.uk

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

These eye-movement studies often use stimuli where linguistic information triggers looks to particular entities in the visual scene, based on their thematic role in the scene (Griffin & Bock, 2000; Knoeferle & Crocker, 2006; Knoeferle & Kreysa, 2012). Furthermore, these eye-movements take place even when the screen has been blanked such that the referents are no longer visible (Altmann, 2004; Knoeferle & Crocker, 2007). This means that these shifts involve an internal memory which tracks both the location and roles of the objects in the visual scene. In the above studies, participants searched for an object where it was previously seen. But it has also been shown that people will look for an object at a new goal location if they were given linguistic input that told them that it was moved there (Altmann & Kamide, 2009). These studies suggest that linguistic and visual representations are bidirectionally linked and changing thematic roles with language can immediately influence visual scene processing.

Although these studies appear to require a tight integration of vision and language, it is still not clear how this is done. Computational models of parsing in the visual world have trouble tracking distinct roles across multiple events with entities of the same kind (e.g., different girls in multiple push events, Mayberry et al., 2009). But outside of language work, there is relevant research on this issue in studies of purely visual processing within the multiple object tracking paradigm (MOT, Pylyshyn & Storm, 1988). In this paradigm, participants are presented with a display of objects (e.g., 9 crosses) that move in random patterns and are instructed to track a subgroup of these objects (the targets). Since the objects are visually identical, participants must track the targets until the test phase at the end of the trial. By varying the number of targets, it has been found that viewers can track a small set of objects in parallel (Alvarez & Franconeri, 2007; Oksama & Hyönä, 2016; Pylyshyn & Storm, 1988). Since this tracking can be done even when visual input is removed (Alvarez et al., 2001; Keane & Pylyshyn, 2006), it is assumed to involve an internal collection of pointers or indexes that can track the identity of objects in parallel by updating their positional information as the objects move around in the visual scene (Pylyshyn, 2000).

Studies of visual processing have found that thematic-role-related information can be extracted from

the motion patterns of simple shapes or dots (Barrett et al., 2005; Heider & Simmel, 1944; Michotte, 1946; Scholl & Tremoulet, 2000; Twomey et al., 2016). For example, Gao et al. (2009) showed participants scenes where multiple circles moved randomly, except for a particular *wolf* circle that was chasing a *sheep* circle around the screen (the wolf and sheep circles were visually identical). The observers were readily able to detect the wolf, as it tended to move in a fairly direct manner towards the sheep. This demonstrated that the agent of a chasing event can be identified from movement patterns alone (Dittrich & Lea, 1994; Frankenhuus et al., 2013; Galazka & Nyström, 2016; Gao & Scholl, 2011; Meyerhoff et al., 2013; van Buren & Scholl, 2017). By using identical shapes that move randomly, these visual paradigms remove semantic features that can influence thematic role processing and provide a framework for examining purely motion-based processing in the visual system and the extent to which this supports role-referent tracking.

Most research examining thematic roles in MOT has focused on single events. To better examine thematic role processing in multiple events, Jessop and Chang (2020) developed the Push-MOT task, which used actions that depicted an agent causally pushing a patient. Causal pushing events were first studied by Michotte (1946) and are one of the most widely investigated instances of causality in the perception literature (Leslie & Keeble, 1987; Mayrhofer & Waldmann, 2016; Saxe & Carey, 2006; Schlottmann et al., 2006; 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.

The Push-MOT task places these push events within an MOT paradigm and requires participants to identify the thematic roles of the objects through sentence production. Jessop and Chang (2020) presented adult participants with a display of nine visually identical circles that were initially moving in random patterns. Occasionally, the movement stopped and a push event took place between two circles. Then random motion resumed with the possibility of depicting up to two additional push events.

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 green*. To provide accurate descriptions, the participants had to track the roles of the circles in the multiple push events as they moved randomly until the colour information was provided at test, which allowed them to map the colour names to sentence positions. Across three studies, the participants could track the roles for two push events at above chance levels suggesting that there was the capacity to track the roles of about four objects.

Since the link between visual and linguistic processing is still not well understood, it is useful to lay out explicitly the processes that are involved in the Push-MOT task. Figure 1 depicts a hypothesized set of processes during a push event (time 7, 9, 11) and the test frame at the end of the trial after a period of random motion.

Since the circles have an identical shape and colour, the identity of circles must be tracked with pointers as they move and this is shown with the letters R, B, and G in Figure 1 (Process 1). We assume that basic motion information like velocity is also tracked with each circle (red arrows, Process 2; Iordanescu et al., 2009). To recognize actions like pushing, it is necessary to identify that the agent circle is moving directly towards the patient circle.

One heuristic that can do this is the chasing subtlety heuristic (Gao et al., 2009), which records the angle between the velocity of the object from Process 2 and the most direct path towards another object. But since this heuristic is more general and can be used for recognizing the components of a push event in this work, we will call it the *angle of approach*. This relational motion heuristic is computed in Process 3 and is depicted by the purple curve between the red velocity vector and the dashed line for the direct path. It can be used to recognize the two components of pushing where circle G is moving towards circle R at time 7 (angle of approach = 5 degrees) and the fact that circle R is moving away from circle G at time 11 (angle of approach = 170 degrees). But to recognize a complete action, an extra action identification process is needed that combines these components together (Process 4). For pushing, the agent first moves towards the patient (Time 7), they make contact (Time 9), and then the patient moves away (Time 11). These three components need to be in this order and without substantial delays between them in order to be seen as a pushing action (Leslie & Keeble, 1987). Once a pushing action has been recognized, then role information can be stored as features on circle tracking pointers (circle G has the AGENT feature, circle R has the PATIENT feature at time 11, Process 5). At test, colours are assigned to the circles. The speaker can

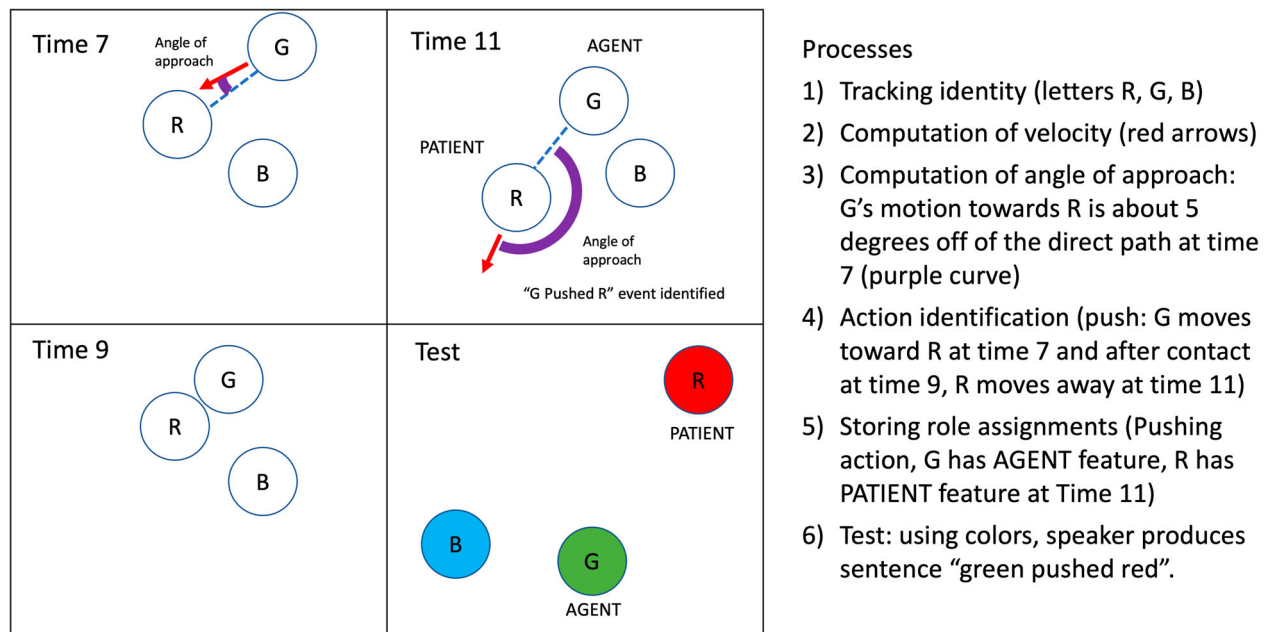


Figure 1. Processes in the Push-MOT task.

describe the scene as *green pushed red*, by searching the pointers for the AGENT feature and producing its colour word in the subject syntactic position. Then it searches for the PATIENT feature and produces the colour in the object syntactic position.

The above theory is sufficient when describing scenes where each circle is only involved in one pushing event. But it is not clear what will happen when the same circle is involved in two pushing events. If circle G pushed circle R and circle B pushed circle G, then the pointer for circle G should have both agent and patient features. One linguistic phenomenon that is dependent on the role overlap between two events is the difference in the processing of subject and object relative clauses. The above example where circle G had different roles in each push event (agent and patient) could be described with an object relative clause (ORC) sentence like *the green that blue pushed pushed red*. If instead the circle G pushed the circle B (same role of agent in both push events), then a subject relative clause (SRC) like *the green that pushed blue pushed red* would be appropriate. There is a large body of work in language comprehension which finds a preference for SRCs over ORCs in processing and this is called the relative clause asymmetry (Gordon & Lowder, 2012; Just & Carpenter, 1992; Sheldon, 1974). Traditionally, this difference has been explained by linguistic memory (Gibson, 1998) or input regularities (MacDonald & Christiansen, 2002; Reali & Christiansen, 2007). The present work examined whether role overlap in a visual scene can influence this asymmetry in language.

The first issue is to determine if there is any influence of overlapping roles on sentence production and whether this creates a bias for SRCs over ORCs within the Push-MOT task. If there is a bias, then we will examine whether this bias is due to the linguistic features of the task. In the modular Linguistic approach that we proposed earlier, visual processing extracts role information and it is passed to a separate linguistic module. We assume that role tracking in the linguistic module is not limited in the same way as the visual system and it can encode a larger number of events and roles. Without these limitations, role overlap itself does not create a processing bias and instead the behavioural differences in the task are due to the structures that are used. The first three studies will test the

Linguistic approach using different sentence structures. If these studies reveal a role for the visual component of this task, then additional studies will be done to examine the nature of that system.

Experiment 1: subject and object relative clause descriptions in the Push-MOT task

The first study examined how people described scenes involving two push actions that differed in how roles overlapped across the push actions (Figure 2). In the Same Role condition, the agent of the first push event was also the agent of the second push event. In the Different Role condition, the patient of the first push was the agent of the second push. At test, these three circles were given colours and participants had to describe their interaction in a sentence. They were asked to use a SRC sentence like *the red that pushed blue pushed green* for the Same Role condition and an ORC sentence like *the red that blue pushed pushed green* for the Different Role condition.

Previous work has found that participants can track about four objects in this task using internal pointers (Jessop & Chang, 2020). To represent these tracking pointers, we will use capital letters corresponding to the colours that are used at test for linguistically distinguishing the circles (R for RED, B for BLUE, G for GREEN, Figure 2), although these colours are not available when the push events are initially seen. In the first push in the Same Role condition, the circle R pushes the circle B, so the R pointer is marked with the feature AGENT1 and the B pointer is marked with the feature PATIENT1. Then when the second

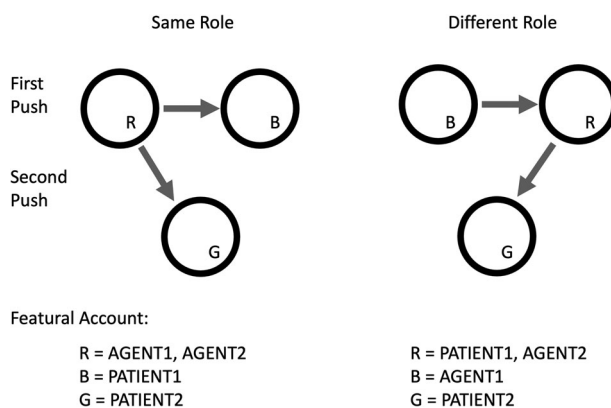


Figure 2. Same and Different Role conditions in Experiment 1 with corresponding Featural representations.

push is seen, the R pointer is given an additional AGENT2 feature and the G pointer is given the PATIENT2 feature. At test, the participant searches the pointers to find an agent, and the double agent features on R make it easy to select that as an agent. In the first push in the Different Role condition, the circle B is given AGENT1, circle R is given PATIENT1 and AGENT2, and the circle G is given PATIENT2. At test, searching for an agent is more difficult, since two different objects have been agents and the system must identify which one is the main clause agent and which is the embedded clause agent. In this account, roles are encoded as features on object-tracking pointers, so therefore this account will be called the Featural account. Since the bias for same roles arises from the fact that the same features are on a single pointer, this Featural account depends on the object tracking mechanisms in the visual system.

On the other hand, the *Linguistic* account argues that visual processing passes role information to the linguistic module. Linguistic role tracking is done with a different formalism which is not limited in the same way as visual role tracking. For example, a propositional representation like $PUSH(CIRCLE_1, CIRCLE_3) + PUSH(CIRCLE_2, CIRCLE_1)$ has two separate copies of the $CIRCLE_1$ code, so accessing its agent role in the first push does not require checking the patient role in the second push. In this work, we assume that role match or mismatch does not play a strong role within the linguistic module itself. Instead, any behavioural differences that are found should arise from the linguistic structures that are used (Gibson, 1998). This study contrasts the predictions of the Linguistic account with the Visual Featural account.

Methods

Participants

All of the experiments in this work were conducted under the approval of the Health and Life Sciences Committee on Research Ethics at the University of Liverpool (reference: 0450). 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. This sample size was

determined via a power analysis conducted using the *simr* 1.0.5 package (Green & MacLeod, 2016) in R version 4.0.2 (R Core Team, 2021). Data from study 1 of Jessop and Chang (2020) were entered into mixed-effects models matching the maximal specification described in the *Analysis* section. 1000 Monte Carlo simulations were performed as the sample size was increased from 10 to 30 in increments of 2. The results showed that a Push-MOT study with 20 participants completing 60 trials would provide 85% power (95% CI [82.63, 87.16]) to detect a log odds ratio of 0.25 with a two-sided alpha level of 0.05. On this basis, each of the studies presented in this work used samples of at least 20 participants. The supplementary materials contain data and scripts for this power analysis as well as other statistical analyses and example videos for this and the following studies (<https://doi.org/10.17605/OSF.IO/PKXZH>).

Stimuli and apparatus

The study consisted of 60 trials: 30 with each role overlap condition (Same/Different Role) with the same condition 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.

Each trial lasted 25 seconds and involved animated sequences where nine identical white circles (0.8° in diameter) moved randomly on a black background (Figure 3). These sequences were shown in full-screen on an LCD (2880×1800 ; $36.5^\circ \times 23.2^\circ$ visual angle), and they were designed and presented using the Processing 3 programming language (<https://processing.org/>).

Throughout each 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° /second (Figure 3A). This was controlled by an algorithm that changed their direction within a 120° window approximately every 250 milliseconds. Their direction was also changed whenever they moved closer than 4.2° to other circles (centre to centre), thereby forcing them to move away from each other. To prevent the circles from moving outside the display window, the circles were redirected towards the centre of the display whenever they were within 1.6° visual angle of the display edge.

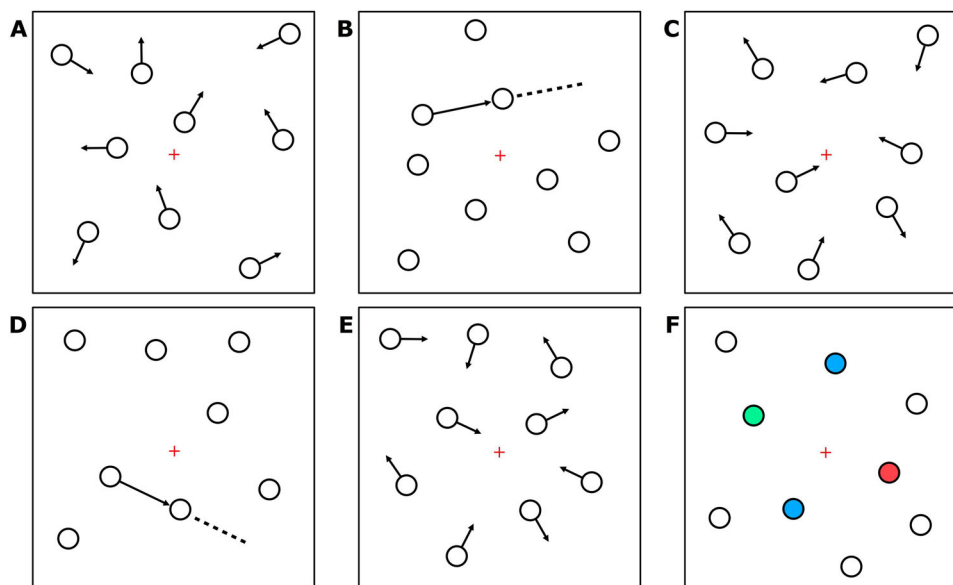


Figure 3. Push-MOT Task: (A) random motion, (B) the first push event, (C) random motion, (D) the second push event, (E) random motion, and (F) the test display. Motion arrows are not present in the actual display.

After 3 seconds of random movement, the first push event occurred (Figure 3B). Two of the circles were selected at random and assigned the roles of agent and patient. Then, all of the circles stopped moving as the agent and patient circles 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 3C) before a second push event occurred (Figure 3D). A third circle was randomly selected to be the patient, while one of the circles from the first push served as the agent, as determined by the event type condition. In the Same Role condition, the agent from the first push also carried out the second. However, in the Different Role condition, the patient from the first push became the agent in the second. Once both pushes were completed, all nine circles continued to move in random patterns for the remainder of the trial (approximately 15 seconds; Figure 3E).

After 25 seconds, all movement was terminated and four of the nine circles were highlighted in red, blue and green (Figure 3F). Since the aim of this study was to test whether the participants could produce a relative clause describing both push events, three of the highlighted circles were the circles from the two push events and the other was

a foil randomly selected from the circles that had not been involved in any push event. To encourage participants to use a relative clause, we created ambiguity by using the same colour (e.g., blue) to highlight both the foil circle and the circle that appeared in both of the push events; specifically, as either an agent in both (Same Role condition) or as a patient then an agent (Different Role condition). Since there were two circles with the same colour, a relative clause would be needed to identify which circle was being referred to. Participants were also instructed to start their sentences with the colour that appeared twice, which ensured consistency in the descriptions.

Procedure

The participants were guided through example trials for both the Same and Different Role conditions. They were instructed to track all the circles 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 circles interacted in a sentence such as *the red that pushed blue pushed green* (Same Role) or *the red that blue pushed pushed green* (Different Role), using the appropriate colour words for the given trial. The participants were instructed to start their utterance with a head noun phrase (NP) with the colour that appeared twice at test. They were also asked to fixate their gaze on the marker in the centre of the screen during the video. 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 aloud before clicking the circles 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 programme recorded the participants' selections and advanced to the next trial.

Coding

The participants' verbal descriptions were audio-recorded and later transcribed. They were coded for whether the target sentence matched the push event that was depicted (Response accuracy). The Same Role events should have been described with an SRC structure and the Different Role events with an ORC structure. Color words had to be in sentence positions within the structures that matched the visual scene. This response accuracy coding scheme allowed for the fact that Same Role events could be described in two different ways, namely by switching the push events in the relative and main clause (e.g., *the red that pushed blue pushed green* vs. *the red that pushed green pushed blue*). However, there was only one correct way to describe the events in the Different Role trials as the event in each clause could not be switched without changing the meaning of the sentence.

Analysis

All of the analyses in the present work used logistic mixed-effects models, as implemented in the *lme4* 1.1-26 package (Bates, Mächler, et al., 2015) via *lmerTest* 3.1-3 (Kuznetsova et al., 2017) in *R* version 4.0.2 (Core Team, 2021). The model parameters were computed using maximum likelihood estimation with the NLOpt optimization algorithm. For each model, the maximal random slopes were initially entered for the random intercept of subject (Barr et al., 2013). Convergence issues were addressed by sequentially removing random slope terms, starting with those that accounted for the least variance. The models were then checked for overparameterization using principal components analysis of the random effects with further simplifications being performed where

necessary (Bates, Kliegl, et al., 2015). Marginal and conditional R^2 statistics are reported as effect sizes, which denote the proportion of the variance explained by the model both with (conditional R^2) and without (marginal R^2) the random effects structure included (Johnson, 2014; Nakagawa et al., 2017; Nakagawa & Schielzeth, 2013). Visualizations of the data are provided in pirate plots with shaded bars to show the mean accuracy of the entire sample, jittered points for the individual participant performance levels averaged across all trials, and violin lines showing the probability density of the data.

Results

A logistic mixed-effects model was fit to the data with role overlap condition (Same/Different Role) as an effect-coded fixed factor. The analysis tested whether there was a reliable difference in response accuracy, which was coded as a binomial variable (match=1, mismatch=0) based on whether the participants' response matched the event in the trial. The maximal model supported by the data included subject as a random intercept without random slopes ($R_m^2 = .07$, $R_c^2 = .29$). As illustrated in Figure 4, participants were more likely to produce an appropriate relative clause sentence in the Same Role ($M = .74$, $SD = .44$) than the Different Role ($M = .53$, $SD = .5$) condition ($\beta = -0.58$, $SE = 0.07$, $t = -8.84$, $p < .001$).

This study demonstrated that the visual scenes in the Push-MOT task have an impact on the accuracy of the sentences used in language processing. The results support the predictions of the Featural account, where the mismatching role features in the Different Role condition create interference relative to matching role features in the Same Role condition. Since the Push-MOT task involves circles that are identical in shape and colour, the semantic plausibility of these structures is unlikely to explain this difference. However, the observed asymmetry could be affected by linguistic difficulties in using an ORC structure and the next study tried to address that possibility.

Experiment 2: Subject and passive relative clause descriptions in the Push-MOT task

The Push-MOT task uses self-moving objects that are potentially animate and previous work has shown that ORCs are used less often when the head NP is

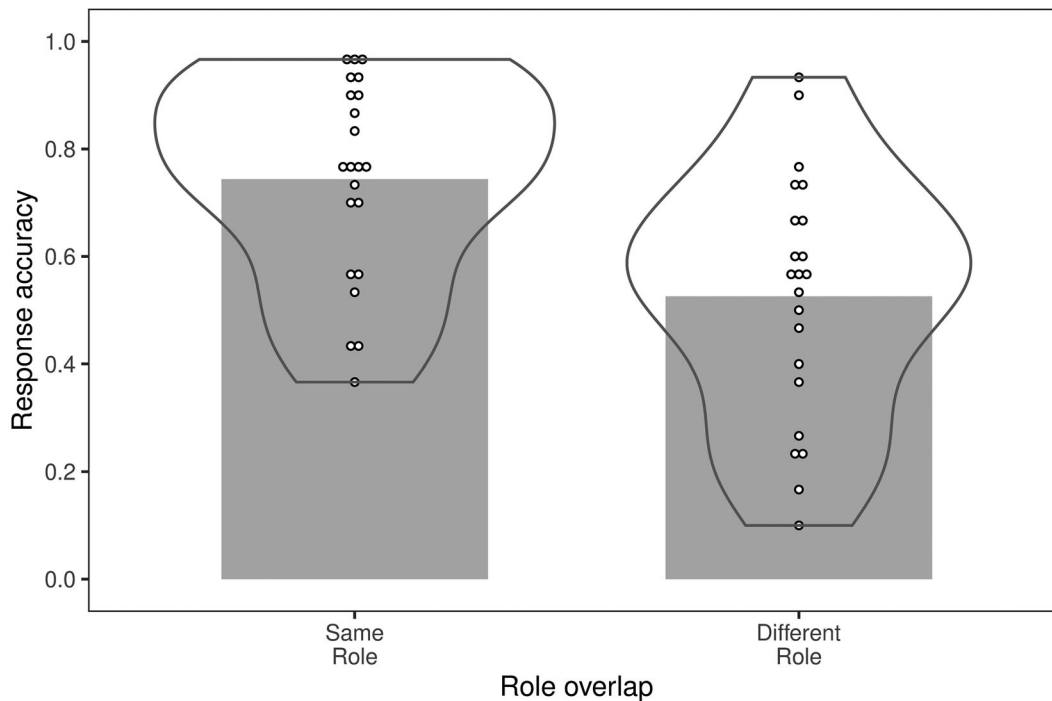


Figure 4. Response accuracy by Role Overlap conditions in Experiment 1. Circles depict the individual subject means, the violin lines show the distribution, and the bars show the condition means.

animate (Gennari & MacDonald, 2008; Mak et al., 2006, 2002; Traxler et al., 2002, 2005). To reduce this argument mismatch, it is possible to use passive relative clauses such as *The blue that was pushed by red pushed green* (Gennari et al., 2012; Gennari & MacDonald, 2008, 2009). 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 et al., 2016) and this is relevant for this study, since the circles are all conceptually identical. Thus, passive relatives may be favoured for the stimuli used in the Push-MOT task.

The visual component of Experiment 2 was identical to the first study (Figure 2). But the language component was changed, such that the Different Role condition was described with a passive relative. If the use of this structure removes the asymmetry, then that would suggest that the results in Experiment 1 arose from the mismatch between animacy and ORCs. However, if there is still a preference for the Same Role condition in the present study, then it would suggest that the basis of the asymmetry is not strongly dependent on the structure used.

Method

The design, stimuli, apparatus, and analysis were identical to Experiment 1. 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 first 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*). Response accuracy was binomially coded with the requirement that participants used an SRC or passive RC (PRC) to correctly identify the roles of the highlighted referents. The participants also produced reduced relatives (e.g., *the red pushed by blue pushed green*), which were treated as PRCs. Although PRCs could have been described with ORCs, this did not occur.

Results and summary

A logistic mixed-effects model was fitted with role overlap (Same/Different Role) as an effect-coded fixed factor (Figure 5). The maximal model supported by the data contained subject as a random intercept and the random slope of event type ($R_m^2 = .08$, $R_c^2 = .39$). It was

found that accuracy in tracking the pushes and describing them in an appropriate relative clause was higher for the Same Role ($M = .78, SD = .42$) condition compared to the Different Role ($M = .55, SD = .5$) condition ($\beta = -0.65, SE = 0.16, t = -4.02, p < .001$).

Experiment 2 found a bias for Same Role conditions similar to Experiment 1, even though the Different Role condition was described with passive relatives that are more common than ORCs structures in English (Roland et al., 2007) and are often produced with animate referents (Gennari & MacDonald, 2009). But the use of different structures still means that there may be some linguistic difference driving the bias for SRC structures. To remove the effects of sentence structure and isolate the role of visual processing in this task, Experiment 3 instructed participants to describe the same stimuli with simple active transitive sentences in both conditions.

Experiment 3: Active transitive descriptions for same/Different Role events

The two previous studies observed a Same Role preference using relative clause structures. The Featural account argues that this asymmetry is due to the tracking of roles in visual processing, so the Same

Role bias should appear even when linguistic differences are neutralized. To test this prediction, Experiment 3 presented participants with the same push events as the previous two studies, but instructed them to describe the scenes using simple active transitive sentences such as *red pushed green* in both conditions. When describing these scenes with transitive sentences, participants can ignore whether roles match or mismatch across the two push events, because they only need to linguistically encode one push event. But if a bias for Same Role events is found, then it suggests that visual tracking of the second push still influences language processing even though that push is not expressed linguistically.

Method

Participants

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

Design and stimuli

The design and stimuli involved the same Push-MOT paradigm as Experiment 1 and 2 with identical Same/Different Role conditions (Figure 2). However, the participants were tested on only one of the

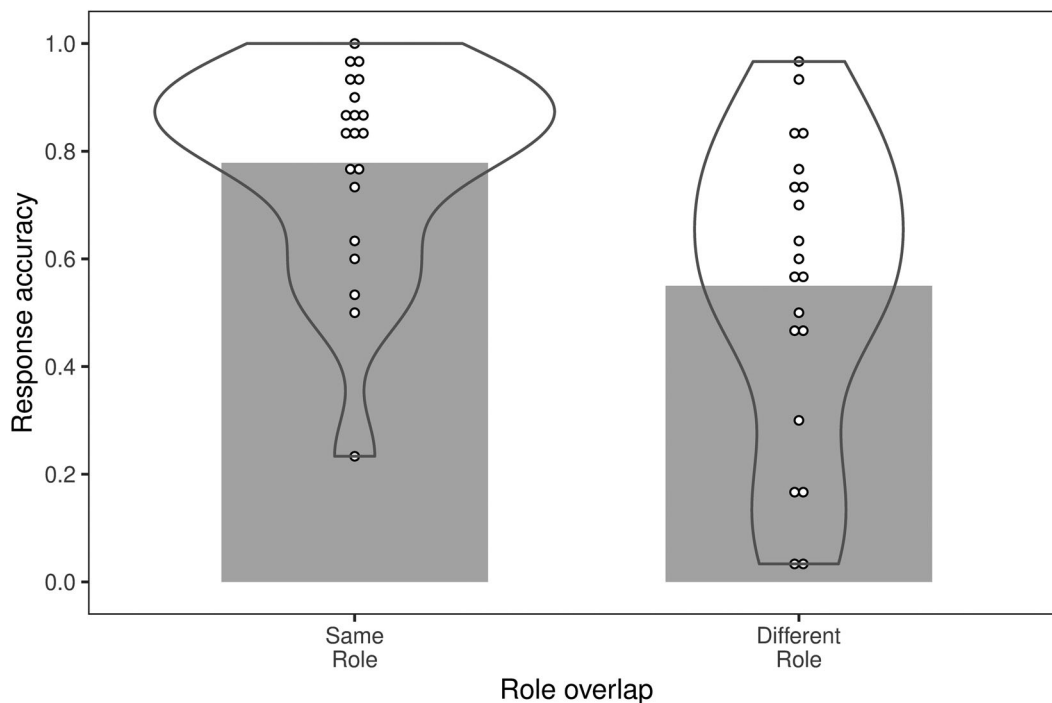


Figure 5. Response accuracy by Role Overlap conditions in Experiment 2. Circles depict the individual subject means, the violin lines show the distribution, and the bars show the condition means.

push events that occurred during the trial, rather than having to describe both interactions. At test, four circles were highlighted in four different colours (red/blue/green/pink). Two of these circles were the agent and patient of one of the push events, while the other two were unrelated foil circles that did not feature in any of the pushes. The study consisted of 60 trials: 30 with each condition (Same/Different Role), each of which was subdivided into 15 trials testing the first push and 15 trials testing the second push. 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 from appearing more than twice in a row.

Procedure

The procedure was the same as the previous studies, except the participants described the events at test using an active transitive sentence such as *red pushed green*. They were instructed to track all of the circles from both push events and were informed that either of them could be tested. Response accuracy was coded as a binomial outcome measure (match=1, mismatch=0), where participants were required to correctly identify the agent and patient highlighted at test.

Results and summary

Experiment 3 used a logistic mixed-effects model with role overlap (Same/Different Role) as an effect-coded fixed factor. The model supported the maximal random effects structure, which included event type as a random slope for subject ($R_m^2 = .01$, $R_c^2 = .19$). The results revealed significantly higher response accuracy for the trials Same Role ($M = .77$, $SD = .42$) than Different Role ($M = .68$, $SD = .47$) conditions ($\beta = -0.24$, $SE = 0.07$, $t = -3.29$, $p < .001$), as illustrated in [Figure 6](#). This demonstrates that the visual scene influences linguistic processing even when there is no difference in the linguistic structure being produced and no issue in selecting which push event to report.

To better understand the role of linguistic structures in this task, an additional model was fitted with the data from the three studies. The model included role overlap (Same/Different Role) as an effect-coded factor with the addition of experiment (1/2/3) as a fixed factor with two Helmert contrasts.

The first contrast compared the first two studies (Experiment 1: active ORC; Experiment 2: passive relative), while the second contrast compared Experiment 3 with the combined accuracy in Experiments 1 and 2. The random-effects supported by the data included role overlap as a random slope for subject ($R_m^2 = .06$, $R_c^2 = .30$). This analysis confirmed that accuracy was higher for the Same Role condition across the three studies ($\beta = -0.49$, $SE = 0.06$, $t = -8.54$, $p < .001$). The magnitude of this asymmetry did not significantly differ between Experiments 1 and 2 ($p = .730$), suggesting that switching from an ORC to a passive relative sentence structure did not affect the difficulty of the task. However, the size of the asymmetry was significantly smaller in the third study compared to the first two studies combined ($\beta = 0.12$, $SE = 0.04$, $t = 3.03$, $p = .002$), which appears to have been driven by higher accuracy levels for the Different Role trials in Experiment 3 ([Figure 6](#)).

These three studies addressed several questions about the link between role tracking in vision and language production. The Linguistic account assumes a strongly modular view of the relationship between vision and language and if referential information is redundantly encoded on separate propositions, then there may be no difference between Same/Different Role events. But in contrast to this prediction, a bias for Same Role events was found in sentence production across the three studies. There was no difference in the Same Role bias in Experiments 1 and 2, and this is inconsistent with the view that passive relatives might be a better match for animate circles than ORCs (Gennari & MacDonald, 2009). It is also at odds the possibility that extra linguistic operations (e.g., movement) in passives can change production accuracy. Finally, Experiment 3 used the same structure for both role overlap conditions, but still a difference was found. Thus, the results in the above studies suggest that there is a component of visual role tracking, that is independent of linguistic operations and can influence language production in different ways depending on the role overlap in the visual scene.

Experiment 4: Comparing visual accounts of overlapping roles in multiple push events

Although there appears to be an influence of visual role tracking on language, the exact nature of the

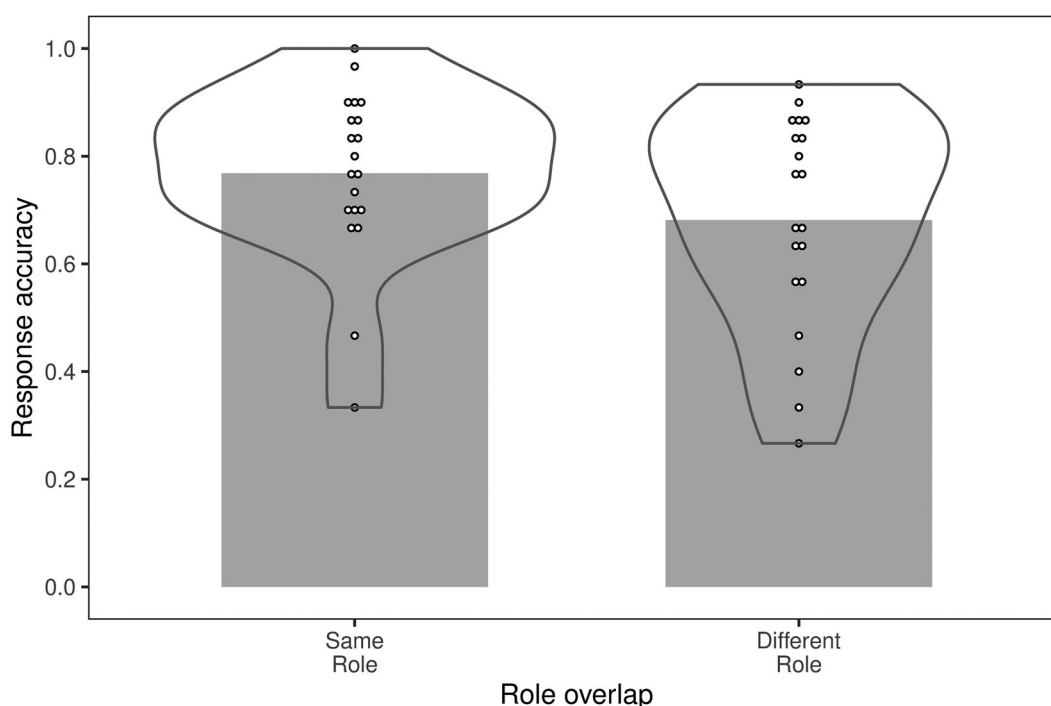


Figure 6. Response accuracy by Role Overlap conditions in Experiment 3. Circles depict the individual subject means, the violin lines show the distribution, and the bars show the condition means.

representation is not well understood. The Featural account argues that roles are properties of particular individuals and this can explain why agents can sometimes exist without patients (e.g., “She is eating”) or patients without agents (“the vase broke”). In the Same Role condition (left side of Figure 7), the consistent feature on the circle R makes it easy to select that referent for the subject of an active transitive. But the Different Role condition (right side of Figure 7) will be more difficult, because the circle R has different role features and more care is required to ensure that the appropriate referent is selected.

The Featural account predicts that mismatching roles will negatively influence performance. One way to test that prediction is to use a No Overlap condition, where there are different circles in each of the two push events (middle column in Figure 7). In the No Overlap condition, there is no referent with mismatching roles, so there should be no interference between roles in this condition (the role features in the Featural account are shown in the middle row in Figure 7). Thus, the Featural theory predicts that the Different Role condition will be less accurate than No Overlap and the Same Role condition should be more accurate than No Overlap.

An alternative approach for encoding roles is to treat them as relations between the agent and

patient (Hafri & Firestone, 2021). In this *Relational* account, patientless sentences like “she is eating” imply an implicit patient argument like “she is eating some food”. In Figure 1, it was argued that role tracking involves both non-relational motion information like velocity (Process 2) as well as relational heuristics like angle of approach (Process 3). But within the Relational account, the claim is that the storage of agent and patient information in Process 5 is also done in a relational manner instead of just adding role features to object-tracking pointers.

In the Relational account, the push event is recognized in Process 4, then in Process 5, role information is stored in a relational representation that is depicted in the bottom row in Figure 7. In the Same Role condition, the circle R occupies the agent position in both push relations (AGENT → PATIENT). It is this similarity across the two push events that makes it easier to select R as an agent compared to the Different Role condition, where R is in agent and patient positions in the relations. But the critical comparison is between the Different Role and No Overlap condition. In both of these conditions, there is no overlap of a referent in the same role position. This predicts no difference between these conditions, which is different from the Featural account’s prediction of greater interference in the Different Role condition.

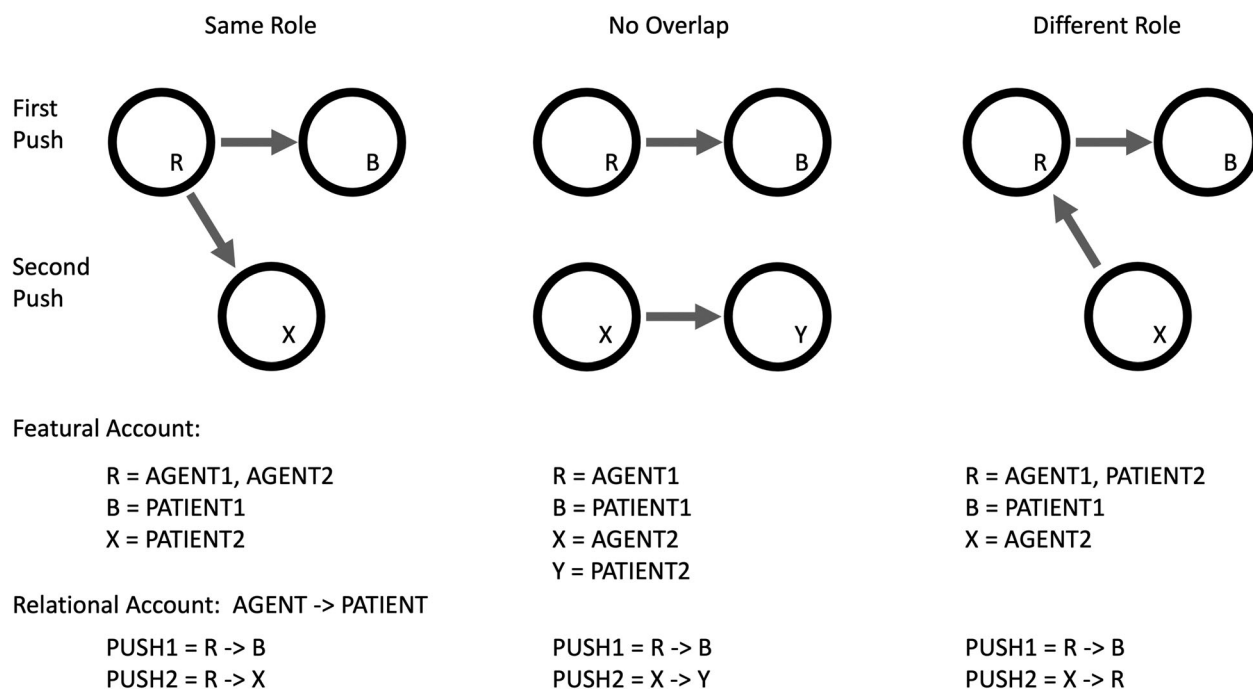


Figure 7. Same Role, Different Role, and No Overlap conditions in Experiment 4 with corresponding Featural and Relational representations.

Thus the Relational account only predicts an advantage when a referent occupies the same role across different relations.

In addition to testing whether a Featural or Relational representation is used to track thematic roles, it is also important to determine how dependent processing is on visual representations. One important difference between the language and visual components of the task in Experiment 3 is that only one push must be encoded linguistically in a transitive sentence. In the Linguistic account, the second push should not have an effect on the description of the first push, since it would not be economical to devote linguistic resources for processing unused thematic roles which can interfere with language production. On the other hand, there is extensive evidence that visual object tracking can track multiple objects in parallel and this was confirmed in the earlier work with the Push-MOT task (Jessop & Chang, 2020). Furthermore, Hafri et al. (2018) found that people track thematic roles in a visual judgement task even though roles were not relevant for the judgement and the task had no linguistic component. They found that these visual judgements were slowed by thematic roles switches from patient to agent. This study suggests that if visual systems are used for tracking roles in the Push-MOT task, it is

possible that task-irrelevant second push roles may influence the description of the first push.

To explicitly test this possibility, the foil circle presented at test was manipulated. In the first three studies, the foils were randomly selected from the circles that were not involved in a push event. In the fourth experiment, a Random Foil condition was created that followed this random selection procedure. This was compared with a Causal Foil condition, where the foil was selected from the second push event. Figure 8 depicts some of the differences between these conditions for the Same Role condition. In the Relational theory, the causal foil will highlight the second push (bold) and this could increase the same role advantage for the circle R to be selected as the subject of the sentence relative to the Random Foil condition. This interaction of role overlap and foil condition is not predicted by the Featural account, because the causal foil is not connected to the other role in its push event. Instead, the behaviour should be similar for both Random and Causal Foil conditions, because the R pointer is attached to two agent features in both conditions.

Overall, the Featural and Relational accounts differ in two predictions. Firstly, the Featural theory predicts a difference between Different Role and No Overlap conditions, while the Relational account predicts no

difference. In addition, the Relational account predicts an interaction of Foil and Role Overlap, while the Featural theory does not. Finally, the Linguistic account would predict no difference in the Role Overlap conditions and no effect of Foil on the assumption that the second push is not linguistically encoded.

Method

Participants

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.

Stimuli

Experiment 4 used the Push-MOT paradigm but with changes in the push events and the test event. In the previous studies, either the agent (Same Role) or patient (Different Role) from the first push event appeared as the agent of the second. This was reversed in this study, as the overlapping circle was always the agent from the first event and its role in the second push was manipulated (Figure 7). The study also included No Overlap trials in which the two events were independent and featured 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 studies. Two of these were the agent and patient targets from the first push, while the third circle was determined by the Foil Type

variable. In trials with a Random Foil, the third object was randomly selected from the distractor objects that did not appear in either push event. For Causal Foil trials, the foil was a circle from the second push event. Thus, for the Same Role events, the foil was always the second patient, and in the Different Role trials, it was always the second agent. In the No Overlap condition, the foil was selected at random from the second event and could be the agent or patient. The participants completed 60 trials with an equal number of trials for each level of Role Overlap and Foil Type, which were ordered using two counterbalance lists to ensure that the same conditions did not appear more than twice consecutively.

Procedure

The procedure was identical to the previous studies 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 3, 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*.

Results and summary

Experiment 4 followed a similar data analysis procedure as the previous studies, using generalized linear mixed-effects models with response accuracy as a binomial outcome variable, where a response is considered accurate when both the correct agent and patient were provided in a transitive sentence. The effect of Role Overlap was tested using centred contrasts where accuracy in both the Same Role and Different Role conditions were separately contrasted against the No Overlap trials, providing a test of whether thematic role consistency influenced accuracy compared to a neutral baseline. Additionally, Foil Type (Random/Causal) was included as an effect-coded fixed factor and was crossed with Role Overlap to assess whether the same effects are observed under different test conditions. The maximal model supported by the data contained subject as a random intercept with random slopes for Role Overlap but not Foil Type ($R_m^2 = .08$, $R_c^2 = .34$). Compared to the trials with No Overlap, the results showed that accuracy was boosted in the Same Role

Random Foil	Causal Foil
Featural Account:	
R = AGENT1, AGENT2	R = AGENT1, AGENT2
B = PATIENT1	B = PATIENT1
G =	G = PATIENT2
Relational Account:	
PUSH1 = R -> B	PUSH1 = R -> B
PUSH2 = R -> G	PUSH2 = R -> G
Linguistic Account:	
PUSH(RED, BLUE)	PUSH(RED, BLUE)

Figure 8. Featural, Relational, and Linguistic representations for Foil manipulation.

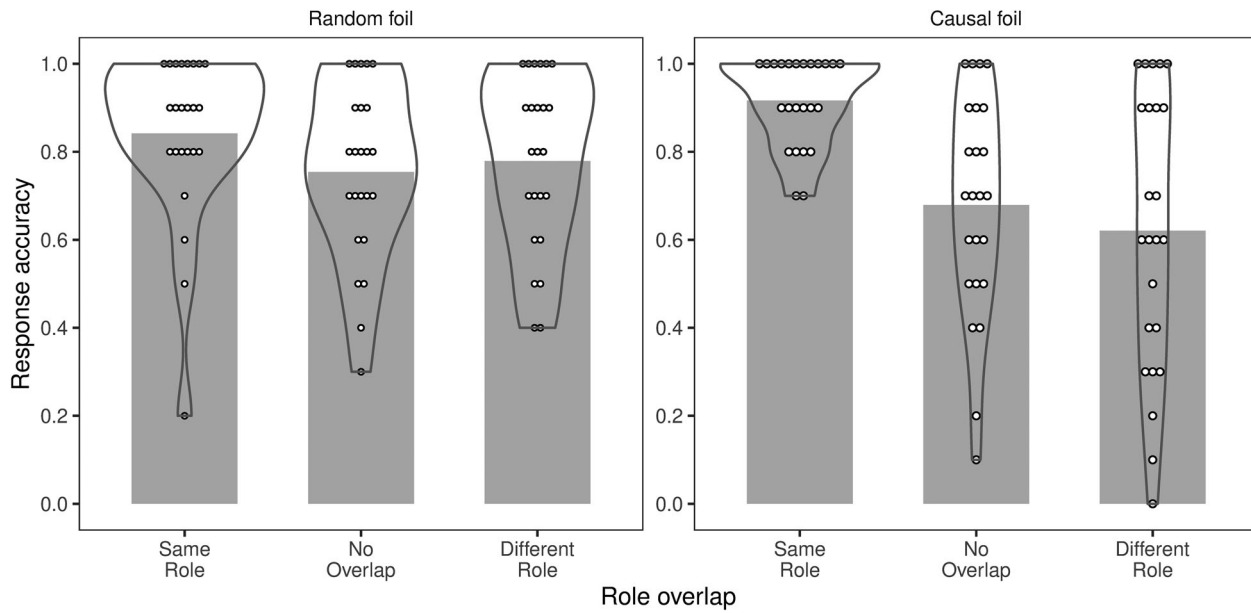


Figure 9. Response accuracy by Role Overlap and Foil Type in Experiment 4. Circles depict the individual subject means, the violin lines show the distribution, and the bars show the condition means.

condition, where the overlapping circle played the agent in both push events ($\beta = 1.11$, $SE = 0.25$, $t = 4.43$, $p < .001$). However, accuracy in the Different Role trials was not significantly different to the No Overlap conditions ($p = .989$). While there was no reliable main effect of Foil Type ($p = .152$), the difference between the Same Role and No Overlap trials was significantly larger in the Causal Foil condition ($\beta = -0.62$, $SE = 0.19$, $t = -3.27$, $p = .001$), but there was no interaction between Foil Type and the Different Role and No Overlap contrast ($\beta = 0.26$, $SE = 0.16$, $t = 1.66$, $p = .096$). These results are illustrated in Figure 9.

The findings of Experiment 4 suggest that accuracy is enhanced when the targets have the same thematic role across the events, while different roles do not appear to have a significant impact on performance compared to independent non-overlapping push events. These findings conflict with the predictions of the Featural account, which predicts greater difficulty in tracking Different Role conditions due to mismatched roles being stored with a single pointer (e.g., $A = \text{AGENT1, PATIENT2}$). The Relational account does not predict interference from mismatched roles, because interactions between relations depend on overlap in the same role on the same referent (Figure 7).

The foil manipulation offers additional insight into the way that visual role information was used in the Push-MOT task. Critically, the difference between

the Same Role and No Overlap conditions was larger when a Causal Foil was presented. This cannot be explained in the Featural account, because the same agent roles (AGENT1 and AGENT2) are activated in both conditions. In contrast, under the Relational account, the Causal Foil activates the second push relation, thereby increasing the same agent boost (Figure 8).

This study provided two kinds of support for the Relational account. First, there was no difference in accuracy between the Different Role and No Overlap conditions, which is at odds with the prediction of the Featural account that interference is created by mismatched roles. Second, accuracy levels were highest in the Same Role condition with Causal Foils, which suggests that the second push can influence the processing of the target even though participants have been told to ignore it. These results are at odds with the predictions of the Linguistic and Featural accounts and appear to support the Relational account. To replicate and expand these results, the next study tested these theories by manipulating patient overlap between the push events.

Experiment 5: Patient consistency in active transitive descriptions in the Push-MOT task

The four previous studies have provided evidence that thematic role consistency facilitates sentence

production accuracy in a visual task where the same referent is the agent in two causal actions. However, the aforementioned role tracking accounts predict a similar enhancement when the same referent is the patient in two push events. In the Featural account, it should be easier to identify a referent as a patient if it has two patient concept features. Similarly, the Relational account predicts a performance enhancement when the same referent occupies the same patient role across different relations. Linguistic studies have not found a strong bias for cases where the same referent is the patient in main and embedded clause (matrix-object ORC bias, de Villiers et al., 1979; Hakuta, 1981; MacWhinney & Pleh, 1988; Traxler et al., 2002), so linguistic mechanisms should not predict a strong bias for patients. The predictions of these accounts were tested in Experiment 5 by varying the role consistency of an overlapping patient (Figure 10).

Method

The design, stimuli, apparatus, and analysis were identical to Experiment 4 with one key change; whereas the first agent was always the overlapping circle in Experiment 4 (except in the No Overlap trials), it was the first patient that always appeared

in the second event in Experiment 5. For Causal Foil trials, the foil was a circle from the second push event. For the Same Role events, the foil was always the second agent, and in the Different Role trials, it was always the second patient. In the No Overlap condition, the foil was selected at random from the second event and could be the agent or patient. Following the same criteria as the previous study, an additional sample of 24 undergraduate participants were recruited at the University of Liverpool.

Results and summary

The results are depicted in Figure 11. The maximal model that was supported by the data contained subject as a random intercept with Foil Type as a random slope ($R_m^2 = .07$, $R_c^2 = .29$). This model found no reliable differences in accuracy between the Different Role and No Overlap conditions ($p = .291$), but accuracy was higher in the Same Role condition compared to the No Overlap condition ($\beta = 0.69$, $SE = 0.19$, $t = 3.55$, $p < .001$). There was no main effect of Foil Type ($p = .242$), nor were there any interactions between Foil Type and the Same Role ($p = .533$) or Different Role contrasts ($p = .881$).

Experiment 5 demonstrated that thematic role consistency in patients also enhances production accuracy

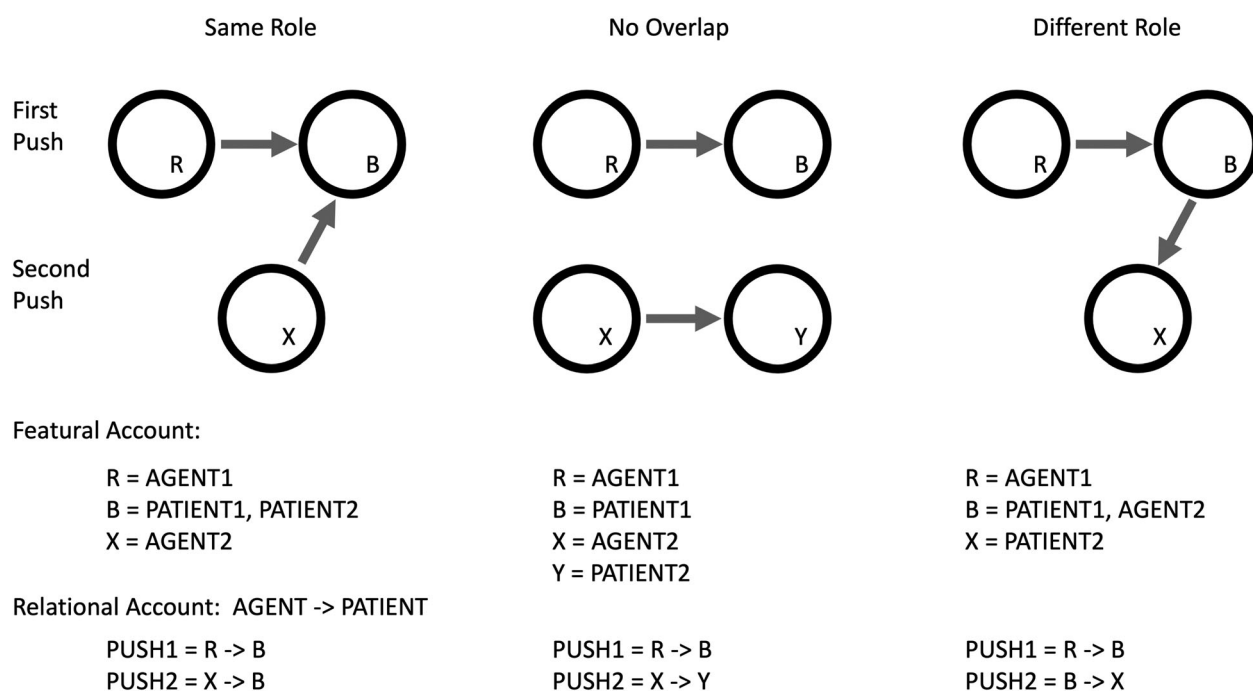


Figure 10. Same Role, Different Role, and No Overlap conditions in Experiment 5 with corresponding Featural and Relational representations.

with no significant difference between the Different Role and No Overlap conditions. These findings were consistent with agent overlap results in Experiment 4 and offer further support to the Relational account, which predicts that scenes with mismatching roles would not create interference relative to the No Overlap condition. They are incompatible with the predictions of the Linguistic and Featural accounts.

In contrast with the fourth study, Experiment 5 observed no reliable effects of Foil Type. The Relational theory predicts that a Causal Foil would raise the activation of the second push relation, increasing the likelihood that this overlapping circle would be selected as the patient. The absence of this effect could be because this Causal Foil enhancement occurred at the object position of the sentence, which may have weakened its overall effect. To test whether this lack of an effect of foil on the Same Role condition is due to the change in sentence position (subject to object) or change in role (agent to patient) between Experiments 4 and 5, this study was replicated with a passive transitive structure that placed the patient into the subject position.

Experiment 6: Patient consistency in passive descriptions in the Push-MOT task

Experiment 6 used the role overlap conditions as Experiment 5, where the overlapping target was the

patient from the first push (Figure 10). However, the participants were instructed to describe the scenes using passive transitive structures, such as *red was pushed by blue*. This places the overlapping role element in the subject position of the sentence, providing a closer comparison with the results of Experiment 4, where the overlapping target was produced in the first position. We predict that when a scene contains a *double patient*, it will 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). If the lack of an effect of Foil Type in Experiment 5 was due to the overlapping target appearing in the object position, then there should be an effect of Foil Type when the overlap is in the subject position.

Method

Experiment 6 used the same design, stimuli, and analysis as Experiment 5. 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

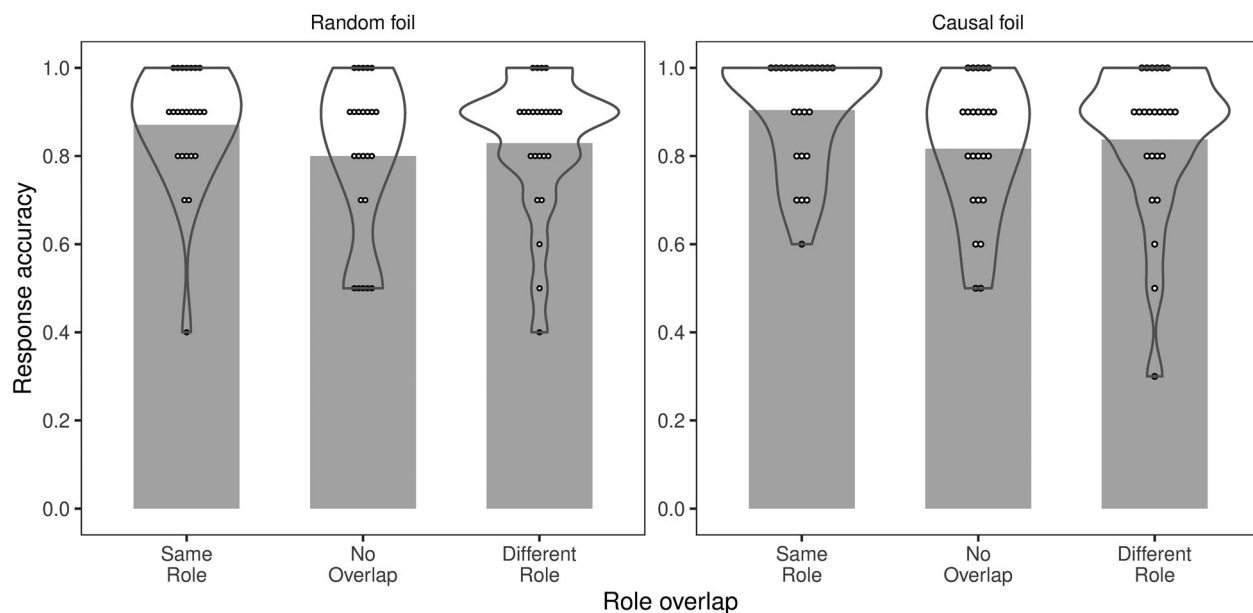


Figure 11. Response accuracy by Role Overlap and Foil Type in Experiment 5. Circles depict the individual subject means, the violin lines show the distribution, and the bars show the condition means.

undergraduate participants were recruited at the University of Liverpool.

Results and summary

The results are depicted in Figure 12. The maximal model that was supported by the data contained subject as a random intercept with the Same Role contrast as a random slope ($R_m^2 = .04$, $R_c^2 = .29$). The model found no significant difference in accuracy between the Different Role and the No Overlap conditions ($p = .433$), but this contrast showed an interaction with Foil Type ($\beta = 0.35$, $SE = 0.16$, $t = 2.16$, $p = .031$), as accuracy in the Different Role condition was lower in trials with Causal Foils but higher in trials with Random Foils. Consistent with the previous studies, accuracy was greater in the Same Role condition ($\beta = 0.76$, $SE = 0.17$, $t = 4.34$, $p < .001$) than the No Overlap condition. However, this did not interact with Foil Type ($p = .283$). There was also no overall main effect of Foil type ($p = .334$).

Experiment 6 replicated the advantage for double patients, which is at odds with the Linguistic account. The lack of interference from mismatched roles compared to non-overlapping events is hard to explain in the Featural account. The results supported the Relational account, where different push events can enhance each other only when they overlap on the same role. Also, as in Experiment 4, there was an interaction between Foil Type and the contrast between the Different Role and No Overlap trials. Although the direction of the effect is not predicted by the Relational theory, this significant difference shows that the second push is influencing the description of the first push, even though participants were told that they could ignore the second push.

To better understand the results from the final three studies, an additional analysis was conducted using the accuracy scores in the combined data from Experiments 4, 5 and 6. Identical to the separate analyses used for these datasets, a mixed-effects model was fit to the data with Role Overlap (Same Role/Different Role/No Overlap) and Foil Type (Random/Causal) as fixed factors. Additionally, experiment (4/5/6) was entered as a Helmert coded fixed factor with two contrasts. The first contrast examined the linguistic difference in Experiments 5 and 6, which both involved an overlapping patient but required the participants to describe the events in either an

active or passive structure. The second contrast tested the role overlap difference between Experiment 4 (agent overlap) and the combined accuracy of Experiments 5 and 6 (patient overlap). The model that converged included subject as a random intercept with Role Overlap and Foil Type as random slopes ($R_m^2 = .05$, $R_c^2 = .27$).

Consistent with the individual results for all three studies, there were no significant differences between the Different Role and No Overlap conditions ($p = .505$), whereas accuracy was significantly higher for the Same Role conditions than the No Overlap trials ($\beta = 0.72$, $SE = 0.12$, $t = 5.76$, $p < .001$). The magnitude of the Same Role advantage was not significantly different between the agent overlap of Experiment 4 and the patient overlap of Experiments 5 and 6 ($\beta = 0.16$, $SE = 0.09$, $t = 1.82$, $p = .069$). Furthermore, the contrast between the active structures in Experiment 5 and the passive structures in Experiment 6 did not yield a significant difference in accuracy ($\beta = 0.25$, $SE = 0.15$, $t = 1.69$, $p = .092$) nor were there interactions with other variables. Since participants were told that they only needed to describe the first push, the second push should not be encoded linguistically and hence the differences found here are hard to explain on the Linguistic account. In addition, the results were similar for agents and patient overlap, which is not consistent with studies using relative clause structures. The Featural theory predicts a difference between No Overlap and Different Role conditions, but that was not found.

While the effects of Role Overlap may emerge from role tracking throughout the entire sequence, the effect of the Foil Type could only occur at the end of the trial when the test circles were given colours. The effect of the Foil Type manipulation was more variable across the three individual studies and the main effects and interactions with role (agent/patient) or structure (active/passive) were not strong. Nevertheless, the combined analysis found an interaction between Foil Type and Role Overlap; specifically, there appeared to be a boost in accuracy for Same Role conditions ($\beta = -0.3$, $SE = 0.11$, $t = -2.81$, $p = .005$) and a reduction in accuracy for the Different Role conditions ($\beta = 0.21$, $SE = 0.1$, $t = 2.18$, $p = .030$) when participants were presented with a Causal Foil at test. A three-way interaction was also observed, as the increase in Same Role accuracy associated with the Causal Foil was greater in

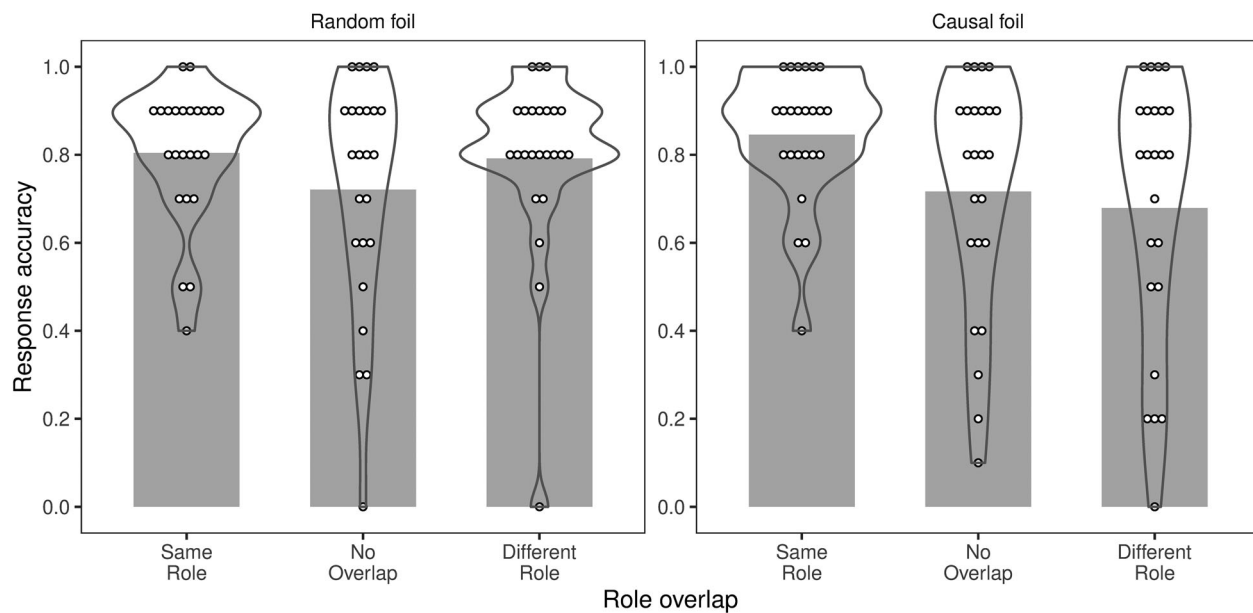


Figure 12. Response accuracy by Role Overlap and Foil Type in Experiment 6. Circles depict the individual subject means, the violin lines show the distribution, and the bars show the condition means.

Experiment 4, which involved an overlapping agent, compared to Experiments 5 and 6, which involved a patient overlap ($\beta = -0.15$, $SE = 0.08$, $t = -2.04$, $p = .041$). These interactions of foil with role overlap conditions suggest that the Causal Foil activated the role information associated with the other member of the second push as predicted by the Relational account and this modulated the description of the first push.

General discussion

The present work examines role tracking in multiple push events with an overlapping participant. The first three studies examined the predictions of the Linguistic account, which assumes that thematic role processing within the linguistic system uses a formalism which is not restricted to a limited number of visual object tracking pointers (e.g., each referent could have a unique code and different copies of the code could represent its role in different propositions). This account predicts that overlap in linguistic thematic role processing itself does not change production accuracy, and instead structural planning is the main way to explain behavioural differences in the task. These predictions were not supported in the first three studies. In Experiments 1 and 2, accuracy for Same Role events was higher than Different Role events, and this Same Role bias was insensitive

to the linguistic properties of the structures used (ORC vs. passive relatives). In Experiment 3, both the Same Role and Different Role conditions were described with a transitive structure, but there was still a Same Role bias even without any linguistic differentiation. The Push-MOT task also rules out other linguistic factors such as biases for particular concepts, because all of the referents were visually identical circles that could only be distinguished by their movements in the scenes. In addition, the participants described the scenes with colour words as names which further minimized the similarity to the input that participants had previously experienced. Together, these results provide support for a visual component of thematic role processing.

Two accounts of how the visual system encodes thematic roles were proposed and their predictions compared in Experiments 4-6. In these studies, the Same/Different Role conditions were combined with a No Overlap condition where there were two push events without any common participants. The Featural account argues that role match/mismatch is encoded on each object referent/pointer, so the Same Role condition will be more accurate than the No Overlap condition, and that condition will be more accurate than the Different Role condition. The Relational account on the other hand argues that enhancement only occurs when the same referent occupies the same role across different events,

and this predicts that the Different Role and No Overlap conditions should be similar. The last three studies all found Same Role was higher than Different Role and No Overlap conditions, which were not different from each other, and this supports the Relational account over the Featural account.

To further probe the Relational account, the foil circle was manipulated in Experiments 4–6 such that it could come from the second causal push or another random circle that was not involved in a push event. In the Featural account, the foil should not matter, because the object trackers already have the relevant information about the thematic roles for each object. But in the Relational account, the causal foil will enhance the thematic roles for the second push and that can cause enhancement or interference with the first push depending on the role overlap condition. It was found that foil condition interacted with role overlap condition in Experiments 4 and 6. In the combined analysis, the Causal foil boosted the Same Role condition and reduced the Different Role condition relative to the No Overlap condition, which is consistent with the Relational account.

The present role tracking results are most consistent with a visual locus. But this claim is controversial, because thematic roles are traditionally viewed as being part of the uniquely human language faculty. To address this issue, it is worth reviewing the processes in the Push-MOT task (Figure 1) and examine evidence that these abilities are available in primates that have not evolved special representations for language. There is evidence that primates can track multiple objects (Process 1; Albiach-Serrano et al., 2010; Barth & Call, 2006), encode velocity (Process 2; Bradley & Goyal, 2008), and goal-directed relational actions (Process 3; Kano & Call, 2014). In addition, they can distinguish causal pushes (pushing a button causes juice to appear) from non-causal pushes (juice appears followed by a button press, Process 4; Tennie et al., 2019). They were also able to generalize this knowledge to their own actions suggesting that they stored an abstract encoding of the event with an agent role that could be changed (Process 5). Although some language-trained primates can generate subject-verb-object sentences using a lexigram keyboard (Process 6; Savage-Rumbaugh et al., 1986), this process does not appear naturally, so this appears to require linguistic processing that is specific to human enculturation. If humans

and primates share some of these abilities, then it is possible that thematic role processing in humans involves some purely visual mechanisms.

Although the animal evidence suggests that some of the processes in the Push-MOT task could be done without language, it does not rule out the possibility that the linguistic ability to use thematic roles might involve systems that were co-opted from systems that originally supported vision. Or it is possible that a separate linguistic thematic role system evolved in humans. In either case, work with the Push-MOT task helps to characterize the nature of that system. A linguistic role tracking system would need to be able to track relational information for multiple push events for up to 4 objects (Jessop & Chang, 2020). It would need to allow the description of events to be influenced by role information from events that are never described. Thus, the results in the present study are consistent with a linguistic locus, if the linguistic system had some of the same features that are known to characterize role tracking in vision.

In addition to providing evidence about the nature of role tracking, the present results suggest a new factor that could be influencing the relative clause asymmetry in language comprehension. Here higher accuracy was found for same role events in production. Higher same role accuracy in production could increase the frequency of SRC sentences in the distribution that is used for language acquisition (MacDonald, 2013). Learners could use this distribution to acquire their language representations and that could create a bias for SRC in comprehension, even without any visual input. This visual role-overlap bias could be one factor that helps to explain why SRC structures are preferred over ORC structures in languages of the world, even when the linguistic properties of the language appear to bias against it (Jäger et al., 2015; Kwon et al., 2010, 2013; Lin & Bever, 2006; Miyamoto & Nakamura, 2003; Ueno & Garnsey, 2008; Vasishth et al., 2013). One problem with this idea is that Experiments 5 and 6 also demonstrated an advantage for patient overlap, although there is no matrix-object ORC bias in language studies (de Villiers et al., 1979; Hakuta, 1981; MacWhinney & Pleh, 1988; Traxler et al., 2002). One possible explanation is that patients are often inanimate and do not act on others, so this might reduce the frequency of multiple events with overlapping patients and therefore the patient overlap bias would not influence the input

distribution as strongly as the agent overlap bias. This idea that visual biases could help explain the SRC asymmetry in language comprehension is speculative, but it suggests a direction for future research.

The present study is one of the few studies that attempts to link research on purely visual motion perception to work on language processing. There are still many open questions and areas that require further study. This work converges with findings from visual world eye-tracking studies, where eye movements to static pictures were immediately changed by language-induced shifts in thematic roles. Here using dynamic motion in videos, people integrated role information from two push events, that were separated in time by random motion, and this internal representation influenced a later description of this event. That suggests that there is some role-overlap sensitive representation that mediates between vision and language.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by a grant from the Economic and Social Research Council (ES/J500094/1) awarded to Andrew Jessop as part of a PhD studentship at the University of Liverpool. Andrew Jessop and Franklin Chang are members of the International Centre for Language and Communicative Development (LuCiD). Franklin Chang was also supported by Grants-in-Aid for Scientific Research grant 19K12733 from the Japan Society for the Promotion of Science. All supplementary materials relating to this research are available in an Open Science Framework repository (<https://doi.org/10.17605/OSF.IO/PKXZH>).

ORCID

Andrew Jessop  <http://orcid.org/0000-0002-2207-4663>
Franklin Chang  <http://orcid.org/0000-0003-1142-1911>

References

- Albiach-Serrano, A., Call, J., & Barth, J. (2010). Great apes track hidden objects after changes in the objects' position and in subject's orientation. *American Journal of Primatology*, n/a–n/a. <https://doi.org/10.1002/ajp.20790>
- Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm". *Cognition*, 93(2), B79–B87. <https://doi.org/10/cv72rg>
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4), 502–518. <https://doi.org/10/dmbxd5>
- Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, 111(1), 55–71. <https://doi.org/10/cf5mrp>
- 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., Wolfe, J., Horowitz, T., & Arsenio, H. (2001). Limits on multielement tracking. *Journal of Vision*, 1(3), 347–347. <https://doi.org/bf4b4f>
- 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>
- Barth, J., & Call, J. (2006). Tracking the displacement of objects: A series of tasks with great apes (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, and *Pongo pygmaeus*) and young children (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, 32(3), 239–252. <https://doi.org/10.1037/0097-7403.32.3.239>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv Preprint arXiv:1506.04967*. <https://doi.org/arXiv:1506.04967>
- 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/gcrnkW>
- Bradley, D. C., & Goyal, M. S. (2008). Velocity computation in the primate visual system. *Nature Reviews Neuroscience*, 9(9), 686–695. <https://doi.org/10.1038/nrn2472>
- Core Team, R. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- de Villiers, J. G., Tager Flusberg, H. B., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, 8(5), 499–518. <https://doi.org/10/gf85s9>
- Dittrich, W. H., & Lea, S. E. (1994). Visual perception of intentional motion. *Perception*, 23(3), 253–268. <https://doi.org/10/b8f2gw>
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33(6), 715. <https://doi.org/10.1006/jmla.1994.1034>
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.
- 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>

- Frankenhuis, W. E., House, B., Barrett, H. C., & Johnson, S. P. (2013). Infants' perception of chasing. *Cognition*, 126(2), 224–233. <https://doi.org/10/f4kjxx>
- 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., 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>
- 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>
- 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>
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76. [https://doi.org/10.1016/S0010-0277\(98\)00034-1](https://doi.org/10.1016/S0010-0277(98)00034-1)
- 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>
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10/f8hm9v>
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279. <https://doi.org/10/ck2rpp>
- Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*. <https://doi.org/f6mq>
- Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, 175, 336–352. <https://doi.org/10.1016/j.cognition.2018.02.011>
- Hakuta, K. (1981). Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese. *Cognition*, 9(3), 197–236. <https://doi.org/10/bshx7h>
- 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>
- 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>
- 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/10.1167/9.4.1>
- 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. (2020). Thematic role information is maintained in the visual object-tracking system. *Quarterly Journal of Experimental Psychology*, 73(1), 146–163. <https://doi.org/10/dckk>
- Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R^2_{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), 122–149. <https://doi.org/10/cfx7xr>
- Kano, F., & Call, J. (2014). Great apes generate goal-based action predictions: An eye-tracking study. *Psychological Science*, 25(9), 1691–1698. <https://doi.org/10.1177/0956797614536402>
- 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>
- Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, 30(3), 481–529. <https://doi.org/10/cjzqdb>
- Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, 57(4), 519–543. <https://doi.org/10/fpdxq9>
- 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. <https://doi.org/10/fmz4fr>
- Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*, 3, 1–15. <https://doi.org/10.3389/fpsyg.2012.00538>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- 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>
- Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. (2013). Subject/object processing asymmetries in Korean relative clauses:

- Evidence from ERP data. *Language*, 89(3), 537–585. <https://doi.org/10/gdthms>
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3), 265–288. <https://doi.org/10/dc4dh2>
- Lin, C.-J. C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. In D. Baumer, D. Montero, and M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 254–260). Cascadilla Proceedings Project.
- 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>
- MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29(2), 95–141. [https://doi.org/10.1016/0010-0277\(88\)90034-0](https://doi.org/10.1016/0010-0277(88)90034-0)
- 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/fc7hbz>
- Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, 33(3), 449–496. <https://doi.org/10/bj7583>
- Mayrhofer, R., & Waldmann, M. R. (2016). Causal agency and the perception of force. *Psychonomic Bulletin & Review*, 23(3), 789–796. <https://doi.org/10/gcrnnc>
- 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*. (T. R. Miles & E. Miles, Trans). Basic Books.
- Miyamoto, E. T., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In G. Garding and M. Tsujimura (Eds.) *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 342–355). Cascadilla Press.
- 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>
- 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>
- 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>
- Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences*, 4(5), 197–207. [https://doi.org/10.1016/S1364-6613\(00\)01477-7](https://doi.org/10.1016/S1364-6613(00)01477-7)
- 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>
- Real, 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>
- 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>
- Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General*, 115(3), 211. <https://doi.org/10.1037/0096-3445.115.3.211>
- Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, 123(1–2), 144–165. <https://doi.org/10/d6cbfq>
- 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., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299–309. <https://doi.org/10/dvnhhr2>
- 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. [https://doi.org/10.1016/S0022-5371\(74\)80064-2](https://doi.org/10.1016/S0022-5371(74)80064-2)
- 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>
- Tennie, C., Völter, C. J., Vonau, V., Hanus, D., Call, J., & Tomasello, M. (2019). Chimpanzees use observed temporal directionality to learn novel causal relations. *Primates*, 60(6), 517–524. <https://doi.org/10.1007/s10329-019-00754-9>
- 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>
- Twomey, K. E., Chang, F., & Ambridge, B. (2016). Lexical distributional cues, but not situational cues, are readily used to learn abstract locative verb-structure associations. *Cognition*, 153, 124–139. <https://doi.org/10/f8t3fq>

- Ueno, M., & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes, 23*(5), 646–688. <https://doi.org/10/dkgvr9>
- van Buren, B., & Scholl, B. J. (2017). Minds in motion in memory: Enhanced spatial memory driven by the perceived animacy of simple shapes. *Cognition, 163*, 87–92. <https://doi.org/10.1016/j.cognition.2017.02.006>
- 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>