Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–18
© Experimental Psychology Society 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1747021819882842
qjep.sagepub.com
SAGE

Thematic role information is maintained in the visual object-tracking system

Andrew Jessop[1] and Franklin Chang[2]

Abstract

Thematic roles characterise the functions of participants in events, but there is no agreement on how these roles are identified in the real world. In three experiments, we examined how role identification in push events is supported by the visual object-tracking system. Participants saw one to three push events in visual scenes with nine identical randomly moving circles. After a period of random movement, two circles from one of the push events and a foil object were given different colours and the participants had to identify their roles in the push with an active sentence, such as *red pushed blue*. It was found that the participants could track the agent and patient targets and generate descriptions that identified their roles at above chance levels, even 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 results were consistent with previous findings of an average tracking capacity limit of four objects, individual differences in this capacity, and the use of attentional strategies. The studies demonstrated that thematic role information can be maintained when tracking the identity of visually identical objects, then used to map role fillers (e.g., the agent of a push event) into their appropriate sentence positions. This suggests that thematic role features are stored temporarily in the visual object-tracking system.

Keywords

Language; multiple object tracking; causality; thematic roles; agent; patient; object pointers

Received: 22 June 2018; revised: 19 May 2019; accepted: 15 July 2019

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, 1987), which 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 using 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, Ferretti, & Amyote, 1997). To address this issue, Dowty (1991) hypothesised that nouns are mapped into sentence arguments using two conceptual prototypes: the proto-agent and the 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, whereas the *window* is only event-independent, hence John should be the subject. Thus, Dowty's proto-role theory argues that conceptual

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Kobe City University of Foreign Studies, Kobe, Japan

Corresponding author:
Andrew Jessop, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.
Email: andrew.jessop@mpi.nl

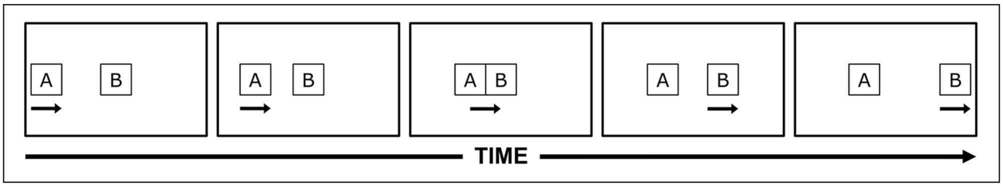

Figure 1. An illustration of the launch effect (Michotte, 1946).

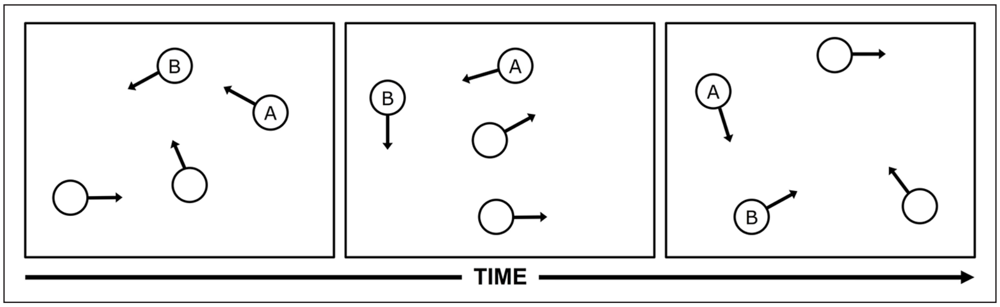

Figure 2. A diagram of the wolf chasing the sheep task (Gao et al., 2009).

features can be used to directly determine the prominence of different arguments without explicitly identifying thematic roles.

Although conceptual feature-based linguistic accounts of thematic roles are popular (e.g., Dowty, 1991; McRae et al., 1997), the proposed features such as *sentience* and *cause* are themselves difficult to define. With respect to causality, Hume (1748/2000) famously argued that when one billiard ball hits another, there is nothing in the scene that necessarily ensures that the movement of the second ball was caused by the first, rather than an accidental coincidence. In response to this, Kant (1781/1997) proposed that our understanding of the visual world includes a priori innate concepts that provide the basis for the subjective impression of causality. Such proposals helped drive the pioneering experimental work by Michotte (1946), who used a launching display to examine causal pushing between two moving shapes (e.g., square A and square B in Figure 1; the letters were not present in the actual stimuli). 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, even though the squares are not moving at the point of contact (centre frame of Figure 1; Hume, 1748/2000). Subsequent research has shown that this effect is strongest when there is physical contact and an immediate reaction (e.g., Schlottmann, Ray, Mitchell, & Demetriou, 2006; Scholl & Tremoulet, 2000; 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 spatiotemporal properties (Leslie & Keeble, 1987; Oakes, 1994; Oakes & Cohen, 1990). The existence of such abilities in infants suggests that thematic role features like causality are not learned associations that develop gradually with extensive experience, but instead stem from perceptual abilities in the human visual system. This can help to explain why adults and infants will perceive animacy and intentionality in the movement of simple shapes bearing minimal resemblance to real-world activities (Barrett, Todd, Miller, & Blythe, 2005; Csibra, Gergely, Bıŕó, Koós, & Brockbank, 1999; Gergely, Nádasdy, Csibra, & Bíró, 1995; Heider & Simmel, 1944).

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 (see Figure 2). 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. In Figure 2, circle A is the only circle moving in a direct path to circle B in these frames (letters were not present in the actual stimuli). Critically, this identification cannot be done based on a single frame because, occasionally, other circles are accidentally moving towards another circle during random motion. Only by aggregating across the whole period of tracking can the true wolf be identified. 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 3–4 months (Frankenhuis, House, Clark Barrett, & Johnson, 2013; Galazka & Nyström, 2016; Rochat, Morgan, & Carpenter, 1997). Interestingly, chase detection appears to be negatively affected by the number of potential wolves and sheep in the display (Meyerhoff, Papenmeier, Jahn, & Huff, 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 & Scholl, 2017), mirroring the findings of object-tracking studies that show that target-merging significantly reduces accuracy in identifying the target objects (Howe & Holcombe, 2012; Scholl, Pylyshyn, & Feldman, 2001). As an object's angle of approach over time (or multiple frames) can help to identify the agent of the chasing action, this work suggests that relational features between object pointers are being maintained during tracking and can be used to identify thematic roles.

Research on chasing interactions shows that thematic roles can be computed by tracking the movement of the various objects in a 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 and Storm, 1988). Neuroimaging research has suggested that regions of the dorsal visual pathway are the primary cortical areas responsible for both MOT (Battelli et al., 2001; Howe, Horowitz, Akos Morocz, Wolfe, & Livingstone, 2009) and the perception of causality in launch events (Blakemore & Decety, 2001; Fugelsang, Roser, Corballis, Gazzaniga, & Dunbar, 2005; Straube, Wolk, & Chatterjee, 2011; Woods et al., 2014). Furthermore, several non-human species can successfully track moving targets, retain their location when they become occluded, and discriminate between causal and non-causal launching events based on movement features (Flombaum, Kundey, Santos, & Scholl, 2004; Hoffmann, Rüttler, & Nieder, 2011; O'Connell & Dunbar, 2005), which suggests that evolutionary pressures may have shaped specialised mechanisms for tracking these features. Thus, there is behavioural and neuroimaging evidence for the link between object tracking 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 is some flexibility in the tracking capacity that is largely determined by both the attentional demands of the task (e.g., Alvarez & Franconeri, 2007; Bettencourt & Somers, 2009; Franconeri, Jonathan, & Scimeca, 2010; Tombu & Seiffert, 2008) and individual capabilities (e.g., Green & Bavelier, 2006; Oksama & Hyönä, 2004; Sekuler, McLaughlin, & Yotsumoto, 2008; Trick, Perl, & Sethi, 2005). Tracking accuracy falls linearly as the number of targets is increased (e.g., Oksama & Hyönä, 2004; Pylyshyn & Storm, 1988) and viewers sometimes appear to use serial attention-switching strategies instead of parallel tracking (e.g., Oksama & Hyönä, 2004). However, participants can separate randomly moving targets from distractors without using conscious eye movements (e.g., 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, Meyerhoff, Papenmeier, & Jahn, 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, Pinto, & Horowitz, 2010; Zelinsky & Neider, 2008). Therefore, many researchers have concluded that a small number of objects can be monitored in parallel (e.g., Alvarez & Scholl, 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, we use the general term *object pointers* to refer to the parts of the visual system that track objects in the scene. Recent models have characterised these pointers as a form of multifocal attention, in which each target simultaneously receives an independent focus within the limits of our available resources (Alvarez & Scholl, 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. In this work, we examine whether these limits also apply to the tracking of thematic role features.

While 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 visual 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 (e.g., 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). Here, we examine whether thematic role-related features require focused serial attention, whether they are tracked automatically with location, or some combination of the two.

Despite an extensive literature examining thematic role-related motion feature processing in visual perception, object-tracking mechanisms are not an explicit component of linguistic or psycholinguistic theories (e.g., Ferreira & Slevc, 2007; Lappin & Fox, 2015; van Gompel & Pickering, 2007; Chang, Bock, & Goldberg, 2003). Psycholinguistic approaches to thematic roles have examined how long-term conceptual knowledge can support thematic role assignment (e.g., doctors are typical agents of verbs like *operate*; Ferretti, McRae, & Hatherell, 2001; Hare, Jones, Thomson, Kelly, & McRae, 2009; McRae et al., 1997; McRae & Matsuki, 2009) and how thematic role information enhances online eye movements in the visual world during sentence comprehension (e.g., Altmann, 2004; Altmann & Kamide, 1999; Knoeferle & Crocker, 2006, 2007; Knoeferle, Crocker, Scheepers, & Pickering, 2005). These theories have not posited any fixed limitation on the number of roles that can be processed, although their use in online processing depends on working memory (e.g., Gibson, 1998; Just & Carpenter, 1992; Kintsch, 1988). For example, Fletcher and Bloom (1988) found that participants could recall stories made up of 19 propositions with multiple thematic roles for each proposition and more than 5 entities that had to be tracked in the story. Another feature of the psycholinguistic approaches is that thematic roles are bound to concepts (Chang, 2002; Mayberry, Crocker, & Knoeferle, 2009; St John & McClelland, 1990), so these models cannot distinguish between two different tokens of the same conceptual type (e.g., dog A and B). Thus, whereas the visual system can only track a small number of individuals, language-based approaches assume that a relatively large number of bindings are available for linking roles and concepts.

In this work, we will assume a pipeline through which visual information is passed to the conceptual system and then used to control language processing. Our question is whether the visual component constrains thematic role processing. As most psycholinguistic theories do not explicitly consider the contribution of visual processing, these approaches implicitly assume that role-related features are extracted from the scene and passed directly to the conceptual system to assign thematic roles to concepts. In this *conceptual account*, individuals are distinguished with additional conceptual features (e.g., AGENT = DOG, LOUDBARK, BLACKTAIL). We consider an alternative *visual account*, in which object tracking in the visual/spatial system is responsible for storing thematic role information during the short period of time that tracking is maintained (e.g., 30 s) before being passed on to the conceptual system. This system tracks individuals based on location and motion information, so it does not require them to have distinctive conceptual features. To compare these two accounts, we used an adapted MOT task in which the target objects were visually identical. The participants were instructed to track the agent and patient from the pushes for a short period, with up to three events occurring in each trial. Afterwards, they were asked to describe one of the pushes in an active transitive sentence to mark the agent and patient of the interaction (e.g., *green pushed red*). As the dependent measure is the accuracy of their linguistic description of a visual event, the task examines the whole process between perception and language. The visual account assumes that people can track several identical objects as they move in random patterns after the pushes until the test event. The account proposes that thematic role features from the push events are linked to the pointers that support tracking, which means that participants can use them to link thematic roles to the colours at test. Furthermore, if tracking the individual targets is capacity limited, the task should become more difficult when the number of agents and patients exceed the tracking capacity. In contrast, the conceptual account assumes that thematic role information is passed to the conceptual system after each push. However, as this system relies on conceptual features to distinguish individuals, it predicts that accuracy will be at chance levels when describing scenes with identical individuals. This research aimed to establish whether the constraints of visual object tracking influence thematic role use in language production.

Experiment 1: Tracking roles without eye movements

To examine whether the limits of the visual system influence the ability to track thematic roles and encode them in descriptions, we developed a challenging *Push-MOT* task in which the participants saw one to three push events and had to describe one of the events in a sentence. In this task, participants saw displays in which nine identical circles moved around randomly (Figure 3a; arrows were not present in the actual display). 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 three different colours randomly (Figure 3d) and the participants were asked to describe the action that occurred between these objects in an active sentence like *green pushed red*. In trials with multiple push interactions, only one of these interactions would be presented at test. As the objects were identical and the test phase was

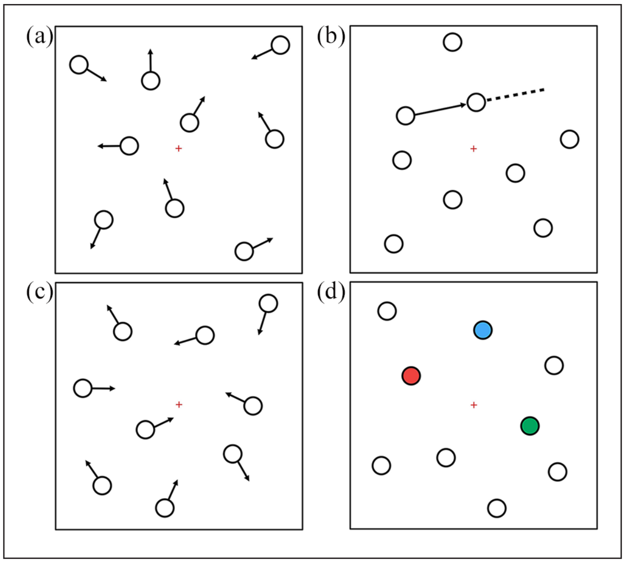

Figure 3. A diagram of the Push-MOT task showing (a) random movement, (b) the push event, (c) random movement, and (d) the test display.

separated from the push events by periods of random motion, the only way to correctly describe the push relationship was to track the multiple agents and patients in the push events.

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 an eye-tracker. Previous MOT studies have found that participants do not need to serially switch their gaze from target-to-target but can successfully track objects while fixating on a marker in the centre of the display (Luu & Howe, 2015; Pylyshyn & Storm, 1988). Therefore, if role-related features are maintained in the object-tracking system, then their sentence descriptions should be accurate even when fixating their overt gaze in a central position.

It is possible that the capacity limitations of the object-tracking system might also 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 and three patients), which is beyond the calculated tracking capacity and should be more difficult. Therefore, we 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). These predictions depend on the assumption of the visual account that thematic role information can be temporarily stored with the pointers that are used for tracking objects for the short duration of the trial. On the conceptual account, thematic role information is directly passed to the conceptual system after each push event. As the objects are identical, there are no conceptual features that can be used to link thematic role features with the colours at test, so this approach predicts low or chance level accuracy and no relationship between accuracy and the number of push events in the scene.

Methods

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 a 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 in which the participants did not fixate their gaze.

Design. The study followed a within-subject design with the number of push events (one/two/three) that occurred during the trial as the independent variable. Each participant completed 60 trials in total, 20 for each push event frequency. In trials with more than one push, only one of the events would be highlighted at test. This was controlled by a *test event* variable (first/second/third), which determined the push event that was tested. In trials with two push events, the participants were tested on the first push 5 times and the second push 15 times. For the three push trials, they were tested on the first push 5 times, the second push 5 times, and the last push 10 times. These levels were selected to avoid over-testing the earlier events in the trial as only the first push event could be tested in trials in which only one push event occurred. The trials were carefully randomised using two counterbalancing lists, such that the same number of pushes did not occur twice in a row, nor were they repeatedly tested on the same event for multiple trials. Critically, there were no overlapping objects between the pushes; each push event involved circles that did not appear in any of the previous pushes.

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 programming language (https://processing.org/) and were presented on a 17 in LED monitor with a screen resolution of 1280×1024 pixels and a 60 Hz refresh rate. The participants were positioned approximately 57 cm 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, which were viewed at a distance of approximately 57 cm (all visual angles reported were calculated based on this distance). These objects were white unfilled circles 0.8° in diameter. A red fixation cross ($0.4^{\circ} \times 0.4^{\circ}$) was positioned in the centre of the tracking field, which occupied approximately $37.6^{\circ} \times 30.2^{\circ}$ visual angle.

Each trial lasted 25 s. During the first 3 s, 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° vector window approximately every 250 ms. The circles moved at a constant speed of $6^{\circ}/\text{s}$. If the objects were closer than 4.2° (centre-to-centre), their direction was changed so that they moved away. At these levels, the expected tracking capacity is approximately four objects (Alvarez & Franconeri, 2007).

After 3 s 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 pseudorandomly, with the algorithm only selecting objects that had not featured in previous pushes. Thus, none of the circles ever appeared in more than one interaction. The push event was an implementation of Michotte's (1946) launch effect; the agent travelled along a direct vector towards the patient, wherein, 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 s, 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 s of random movement between each push.

After 25 s, 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. In trials in which multiple push events occurred, only one of these pushes would be tested. Video examples of the stimuli are available in an Open Science Framework repository (https://osf.io/k7t83/).

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 s. When the agent, patient, and foil objects changed colour at the end of the trial (Figure 3d), 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, 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. Logistic mixed-effects models were fit to the data using the *lme4* 1.1.21 package with the *bobyqa* optimizer algorithm (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.6.1 (R Core Team, 2019). The dependent measure of the analysis was sentence accuracy, which reflected whether the participants' active transitive description of the trial (e.g., *red pushed blue*) correctly identified both the agent and patient of the event being tested (1 = *correct*, 0 = *incorrect*). If either of these roles were incorrect, then the entire utterance was considered inaccurate. As the participants were given three referents at test (red/blue/green) which they could produce in one of two sentence slots (agent/patient), there were six possible responses for any given trial. Therefore, the likelihood of producing a correct description by chance was computed as 0.1666. The fixed effects structure of the model consisted of the number of push events (one/two/three) as a centred continuous predictor. The random effects structure of the model represented the maximal model supported by the data (Barr, Levy, Scheepers, & Tily, 2013). Subject and test event (the push event highlighted at test; first/second/third) were entered as random intercepts with the number of pushes predictor as a random slope for each intercept. If necessary, the random effects structure was simplified until model convergence was achieved, starting with the random slopes that accounted for the least amount of variance. The hypothesised effect of the number of pushes predictor was tested via likelihood-ratio (χ^2) comparisons through the sequential decomposition of the model. The marginal and conditional statistics are also reported as effect sizes (Johnson, 2014; Nakagawa, Johnson, & Schielzeth, 2017; Nakagawa & Schielzeth, 2013). 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 effects structure included (conditional

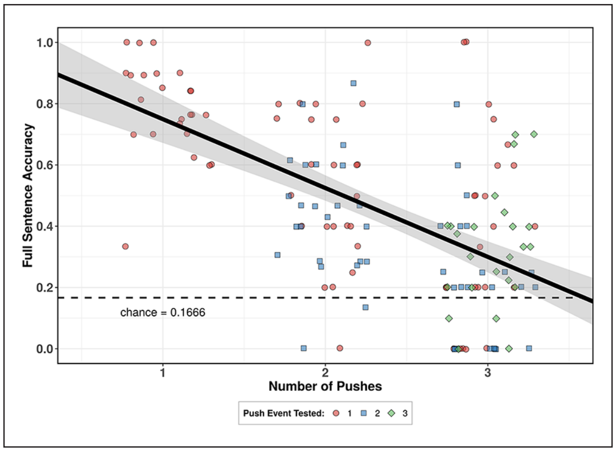

Figure 4. The mean proportion of correct descriptions by the number of pushes, the push event tested, and participant for Experiment 1.

R^2) or excluded (marginal R^2) from the calculation. All of the statistics reported were bootstrapped ($R=1{,}000$) to obtain 95% confidence intervals (CIs) and accurate p values (Luke, 2017). The figures used to illustrate these models use points to show the mean proportion of correct descriptions by the number of pushes, the push event tested, and participant. These points are jittered along the x-axis for clarity. Regression lines on the figure are fitted to the aggregate data by participant, so these illustrations are slightly different from the logistic model beta estimates, which are based on the individual trial data.

Results

Consistent with the criteria applied in previous MOT work (Pylyshyn & Storm, 1988), trials were rejected when the participant fixated on an area more than 2° 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.

The maximal model that converged contained subject as a random intercept with the number of pushes as a random slope, as well as test event as a random intercept. As illustrated in Figure 4, the mixed-effects model found a significant negative effect of the number of pushes in the scene, $\beta = 1.06\ [-1.27,\ -0.85]$, $SE = 0.11$, $\chi^2(1) = 112.71$, $p < .001$, suggesting that sentence accuracy decreased as additional push events needed to be tracked. The maximal model with the number of pushes as a fixed predictor accounted for 17.78% of the variance in the data without the random effect structure and 21.44% of the variance when it was included ($R_m^2 = 0.1778,\ R_c^2 = 0.2144$).

Following this analysis, bootstrapped exact binomial tests were performed to examine whether the participants' description accuracy was above chance (calculated as 0.1666) in the different conditions. The first set of tests looked at performance with different push event frequencies. The binomial tests showed that description accuracy remained consistently above chance, even in trials in which three push events occurred (one push: $M=0.79$ [0.75, 0.83]; two pushes: $M=0.47$ [0.42, 0.51]; three pushes: $M=0.3$ [0.26, 0.35]; all p values $< .001$). This was consistent with the fact that the regression line reaches chance levels at 3.62 [3.02, 4.52] push events.

The previous analyses show that the participants were able to identify the agent and patient circles in displays with three push events while fixating their gaze in a central position. However, it is possible that the participants were strategically tracking particular objects rather than monitoring all three events (see Yantis, 1992). Therefore, in a second set of binomial tests, we analysed whether performance in trials with three push events varied depending on which of the pushes was highlighted at test (the test push). We focused on this set size because the number of objects that need to be tracked exceeds the predicted capacity of four (Alvarez & Franconeri, 2007). These binomial tests found that description accuracy was above chance when the participants were tested on the first and last push events in the trial (one push: $M=0.33$ [0.24, 0.42], $p=.003$; three pushes: $M=0.33$ [0.27, 0.4], $p<.001$), but not when they were tested on the second push event ($M=0.23$ [0.15, 0.3], $p=.139$). This suggests that they were not always able to track all six objects in parallel, but instead favoured the objects in the first and last push events. Although they were unable to make overt gaze shifts, the participants could covertly shift their focus of attention to follow the objects from particular push events. This would mean that the above chance performance with three pushes was due to the combination of a parallel object-tracking system with a capacity of around four items (for the present stimuli) combined with attention strategies to support the more difficult trials.

We found that participants were able to produce accurate descriptions that identified the correct agent and patient at above chance levels in this task. As all the circles were identical, they could not use object properties like colour or shape to track the identity of each circle. Accuracy decreased with additional pushes, which suggests that attention was taxed as more events needed to be tracked. For the difficult three-push trials, the participants were able to provide accurate descriptions using strategies in combination with parallel tracking. Thus, the results were more consistent with the visual account, in which the properties of the object-tracking system support and limit the ability to encode thematic role information. As the participants appeared to use covert shifts of attention to support their behaviour, we decided to tax attention with a concurrent distraction task to better understand the role of attention strategies in tracking thematic roles.

Experiment 2: Distraction task

Experiment 1 demonstrated that observers can track the agents and patients of multiple push events while fixating their gaze on the centre of the display. However, as accuracy decreased with additional push events, it remains unclear whether the participants were monitoring all of the pushes in parallel or using covert shifts of attention to support their tracking. To attenuate potential covert switching, we tested a new group of participants with the Push-MOT task while they also performed 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, Horowitz, & Vogel, 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, Mcgeorge, Pearson, & Milne, 2004, 2006; 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 & Scholl, 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 (Pylyshyn & Storm, 1988), so we selected a distraction task in which success depends on focal attention. The participants provided a speeded response (via keypress) 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 description accuracy remains above chance, then it would support the visual account in which the object-tracking system is maintaining thematic 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. In all, 24 undergraduate participants were recruited from the same population as in Experiment 1.

Apparatus. The study used animated display sequences that were designed and presented using the Processing programming language (https://processing.org/) and shown in full-screen on a widescreen monitor ($2880 \times 1800;$ $\sim 36.5^{\circ} \times 23.2^{\circ}$ visual angle).

Stimuli. The stimuli involved the same Push-MOT task used in Experiment 1 (see Figure 3), with the addition of a simple distraction task (see next section).

Procedure. The study followed the same overall procedure as in Experiment 1, 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 s ($M = 11 \pm 2$ 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 keypress.

Results

Performance on the colour change task was considered accurate for a given trial if the average response time to the changes was less than 1 s, 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 contained the random intercept of test event (without random slopes) and the random intercept of subject with the number of pushes as a random slope. As illustrated in Figure 5, the mixed-effects model found a significant negative effect of the number of pushes in the scene, $\beta = -0.73\ [-0.97,\ -0.49]$, $SE = 0.12,\ \chi^2(1) = 46.08,\ p < .001$, as sentence accuracy decreased as additional push events needed to be tracked. This model, with the number of pushes as a fixed predictor, accounted for 8.72% of the variance in the data without the random effect structure and 17.68% of the variance when it was included ($R_m^2 = 0.0872, R_c^2 = 0.1768$).

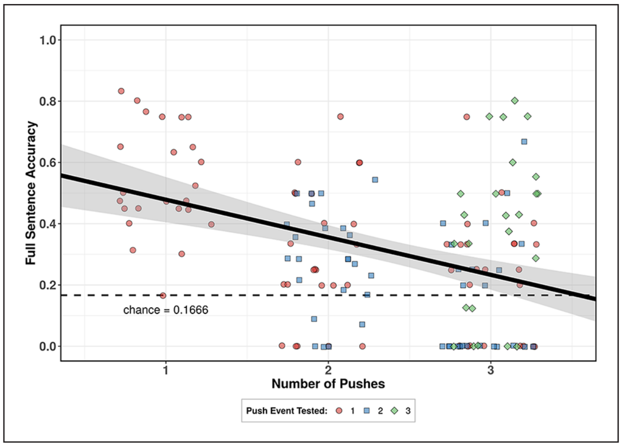

Figure 5. The mean proportion of correct descriptions by the number of pushes, the push event tested, and participant for Experiment 2.

Following the same procedure as in Experiment 1, exact binomial tests with bootstrapping were used to examine whether the participants' description accuracy remained above chance in the different conditions. The first set of tests showed that description accuracy was above chance in trials with one push (M=0.54 [0.5, 0.59], p < .001), two pushes (M=0.29 [0.24, 0.33], p < .001), and three pushes (M=0.28 [0.23, 0.33], p < .001). This was consistent with the regression line of the mixed-effects model, which predicts that that performance will reach chance levels at 3.52 [2.65, 5.24] push events.

To examine whether strategies were used, binomial tests were conducted for each of the test events in trials with three pushes. These analyses found that accuracy was above chance when the participants were tested on the last push event in the trial (M=0.36 [0.29, 0.43], p < .001), but not when they were tested on the first (M=0.23 [0.15, 0.32], p=.161) or second push event (M=0.17 [0.09, 0.25], p=.501).

Experiment 2 showed that participants can accurately track push events at above chance levels when simultaneously responding to a distraction task. While accuracy degraded with additional pushes, it remained above chance in trials with three push events. However, the participants do not appear to have tracked all three pushes in parallel, but instead favoured the most recent push in these trials and performed at chance levels for the earlier events. The findings were largely consistent with the first study, implying that the ability to identify agents and patients was not blocked by the introduction of a distraction task. It is likely that some attention processing was utilised to complete the secondary task as overall accuracy appeared to be lower than in the first study. However, performance does not appear to have been strongly impacted, consistent with the visual account that object pointers track agent and patient features in parallel so that these features can then be used in language production.

Experiment 3: Temporarily invisible objects task

A key assumption of the visual account 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 & Scholl, 2005; Flombaum, Scholl, & Pylyshyn, 2008; Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006; Scholl & Pylyshyn, 1999). For example, Scholl and 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. When an object disappears and reappears, the impression that the two instances represent the same entity requires some internal pointer that is linked to the moving object. 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, Incledon, & Little, 2012; Iordanescu, Grabowecky, & Suzuki, 2009; Luu & Howe, 2015). Whereas these cues may be similar to those used for role identification, one important difference is that the 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). Thus, it is possible that participants can track motion features across occlusion but lose track of thematic role features. We test this in the Push-MOT task 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 (<500 ms) vanish. This provides a way of testing whether thematic role-related features are also maintained by pointers during occlusion.

Method

Participants. A total of 18 undergraduate participants were recruited from the same population as in Experiments 1 and 2.

Stimuli. The stimuli used the same core Push-MOT task as in Experiment 1 (see Figure 1), with three important changes. First, the randomness of the objects' movement patterns was greatly reduced, with direction changes occurring much less frequently (approximately every 1,000 ms).

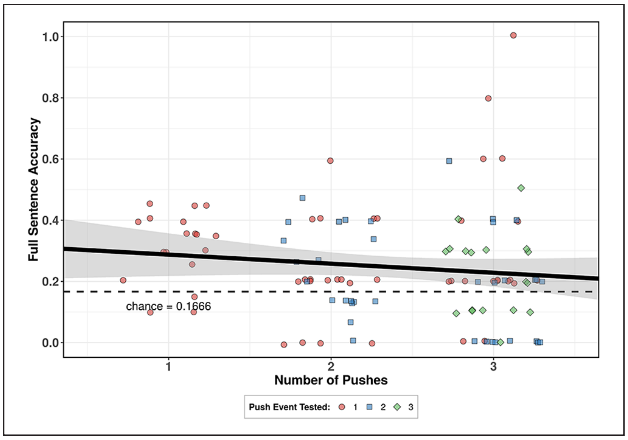

Figure 6. The mean proportion of correct descriptions by the number of pushes, the push event tested, and participant for Experiment 3.

Second, there was no restriction on the distance between the objects and they could pass through each other. When the distance between two (or more) objects was less than 2.5° (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 ms at a time. Although the movement patterns were randomly generated by the experimental programme, each target object was occluded for approximately 0.8 s per trial on average ($M = 837\,\text{ms} \pm 477\,\text{ms}$).

Procedure. The procedure was identical to Experiment 2, 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. Also, no task-specific exclusion criteria were applied in Experiment 3.

Results

The maximal mixed-effects model that converged contained test event and subject as random intercepts without random slopes. The model observed a negative slope of the number of pushes in the scene (see Figure 6), but this did not reach statistical significance, $\beta = -0.2\ [-0.38,\ -0.02]$, $SE = 0.09$, $\chi^2(1) = 3.08$, $p = .109$, $R^2_m = 0.008$, $R^2_c = 0.0221$. Whereas this suggests that accuracy remained consistent as the number of push events in the scene increased, it does not indicate whether this accuracy level was above chance. As in the previous studies, performance was compared against chance levels (calculated at 0.1666) using exact binomial tests with bootstrapping ($R = 1{,}000$). These tests found that the overall description accuracy was above chance in trials with one push ($M = 0.31$ [0.26, 0.36], $p < .001$), two pushes ($M = 0.24$ [0.19, 0.28], $p = .009$), and three pushes ($M = 0.23$ [0.19, 0.28], $p = .015$), showing that participants were able to track agents and patients even for three push events.

To see whether the high accuracy levels observed in trials with three pushes were supported by strategies, we tested whether the description accuracy in these trials varied depending on which of the push events was highlighted at test. These analyses found that accuracy was above chance when the participants were tested on the first push event in the trial ($M = 0.32$ [0.23, 0.42], $p = .010$), but not when they were tested on the second ($M = 0.17$ [0.09, 0.25], $p = .542$) or last push event ($M = 0.22$ [0.16, 0.28], $p = .093$). This suggests that participants may have strategically focused on the first push event.

Although this task was difficult the participants were still able to identify agents and patients at above chance levels overall. In contrast with the previous experiments, increases in the number of push events did not reduce accuracy. One possible explanation is that the trials with one and two push events had longer periods of random motion, meaning there was a greater chance for occlusion to occur and reduce performance in these trials. The impact of occlusion on overall description accuracy is consistent with the visual account, which assumes that thematic role information is associated with the pointers that support tracking during periods of occlusion.

Combined analysis

Task differences

To compare the performance across the three tasks, we combined the data from all the three studies into one analysis (Figure 7). The same logistic mixed-effects model structure that was used in the previous analyses was also fit to the combined data, with the addition of experiment (1/2/3) as a Helmert coded factor. The maximal model that converged included the random intercept of test push (with no random slopes), plus the random intercept of subject with random slopes for the number of the pushes. This model showed that there were differences in the overall accuracy levels between the three studies. The number of correct descriptions produced at test was higher in the first experiment that used eye-tracking, compared with the second study which involved a distraction task, $\beta = -0.73[-0.99, -0.47]$, $SE = 0.13$, $\chi^2(1) = 8.01$, $p = .009$. The participants were even less likely to produce an accurate description in the third experiment, in which the objects became temporarily invisible, compared with the first two studies combined, $\beta = -0.58[-0.74,\ -0.43]$, $SE = 0.08$, $\chi^2(1) = 24.59$, $p < .001$. The model also confirmed the negative main effect of the number of pushes in the scene, $\beta = -0.64[-0.79, -0.48]$, $SE = 0.08$, $\chi^2(1) = 48.31$, $p < .001$, and this had a significant interaction with experiment; a steeper negative slope was observed for the number of

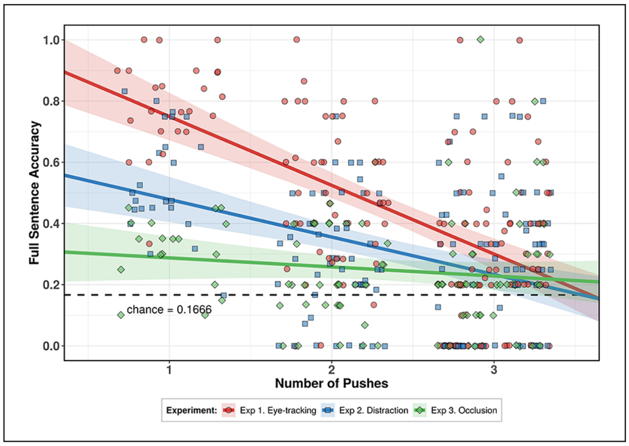

Figure 7. The mean proportion of correct descriptions by the number of pushes, the push event tested, and participant for all the three studies.

pushes variable in Experiment 1 than in Experiment 2, $\beta=0.52\ [0.25, 0.79]$, $SE=0.14$, $\chi^2(1)=9.92$, $p=.005$, whereas the negative slope observed in Experiment 3 was significantly flatter than in the first two experiments, $\beta=0.44\ [0.28, 0.6]$, $SE=0.08$, $\chi^2(1)=22.85$, $p<.001$. This model, with number of pushes and experiment as fully crossed fixed predictors, accounted for 15.06% of the variance in the data without the random effects, and 20.23% when they were included ($R_m^2=0.1506$, $R_c^2=0.2023$). The three experiments used the same methods for testing participants (labelling circles with colours and eliciting active sentence descriptions of the push event) and the time between seeing the push events and the test events were also similar. Therefore, as the only differences between the experiments were related to the visual features of the task, the interaction between experiment and the number of pushes in the trial means that these features affected the slope of this effect, which is consistent with a visual locus for the processing of thematic role information.

Individual differences

As earlier work has observed large individual differences in MOT capacity (Drew & Vogel, 2008; Huang, Mo, & Li, 2012; Oksama & Hyönä, 2004), an additional analysis was performed to assess whether the participants with the highest description accuracy levels in trials with, for example, one push event, were the same participants that produced the most accurate descriptions in the trials with two or three push events. A series of Pearson's correlations were fit to the combined dataset, which were bootstrapped to obtain 95% CIs and accurate p values ($R=1{,}000$). These correlations showed that description accuracy in the trials with one push event had a strong positive relationship with accuracy in trials with two push events, Pearson's $r(64)=0.56$ [0.41, 0.72], $p<.001$, and this is what we would expect if these were due to the same underlying parallel object-tracking mechanism. A smaller positive correlation was observed between trials with three events and those with one push (Pearson's $r(64)=0.26$ [0.04, 0.48], $p=.120$), and two pushes, Pearson's $r(64)=0.25$ [0.04, 0.45], $p=.107$, but these relationships did not reach statistical significance. The lack of a correlation with the three event trials is also predicted if participants are adopting strategies to support behaviour in these trials, as these strategies would be different from the parallel tracking abilities used for one and two push trials, and the same participant might also vary their strategy on different trials.

Role assignment accuracy

Throughout this research, full sentence accuracy served as the dependent variable of all the analyses, which required the correct agent and patient to be identified in the description. It is possible that there were 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), which would suggest similar accuracy levels between agents and patients, as these roles are defined relative to each other. Differences in accuracy between the agent and the patient would provide additional information about how the viewers completed the Push-MOT task.

To examine whether the participants were more accurate in identifying agents or patients, a mixed-effects model was fit to the accuracy of each individual argument in the sentence. In order to meet the assumption of independence necessary for applying linear models, it was not possible to include all of the trials for the combined data in this analysis. During the test phase of the tasks, the participants were provided with three targets highlighted in different colours and were required to produce an active transitive sentence to identify the agent and the patient of the push event (e.g., *red pushed blue*). Consequently, identifying the correct agent (e.g., red) meant that the participants could not produce this target in the patient slot, and vice versa, making accuracy in each position dependent on the other. To mitigate this dependency, the accuracy of the alternative sentence argument was held constant. The data included in the analysis were the agent accuracy scores only from the trials in which the correct patient was provided (49.92% of the total dataset), and patient accuracy scores in trials in which the correct agent was provided (55.91% of the total dataset). This made it possible to gain some insight into whether role assignment errors were more likely to occur for the agent or patient of the sentence.

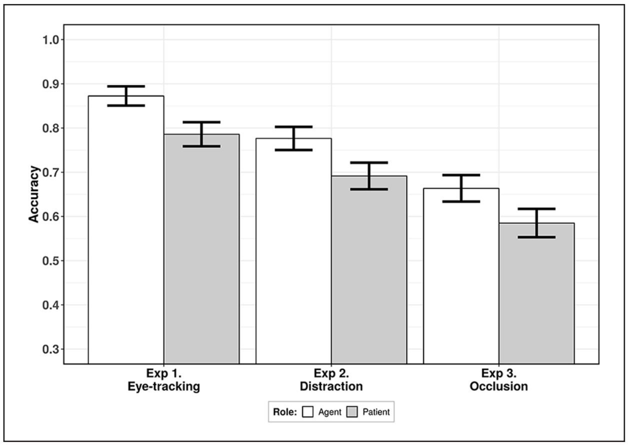

Figure 8. The accuracy in identifying the agent and patient objects when the alternative role has been produced correctly. The error bars represent the standard error after removing the random effects of the mixed-effects model using the remef package for R (Hohenstein & Kliegl, 2019).

The logistic mixed-effects model fit to these data included the number of pushes as a predictor, experiment (1/2/3) as a Helmert coded factor, and role (agent/patient) as an additional effect coded factor. The model that converged contained test push as a random intercept only, with subject as a random intercept with a random slope for role. The model showed that the odds of providing a correct response were higher for agents than patients, $\beta = 0.54\,[0.35, 0.72]$, $SE = 0.09$, $\chi^2(1) = 27.46$, $p < .001$, with accuracy for agents being 8.47% [5.66%, 11.25%] higher than patients. Critically, this agent advantage did not interact with the number of pushes in the trial, $\beta = -0.2\,[-0.44, 0.05]$, $SE = 0.12$, $\chi^2(1) = 1.04$, $p = .326$, nor did it vary between Experiments 1 and 2, $\beta = -0.12$ $[-0.32, 0.08]$, $SE = 0.1$, $\chi^2(1) = 0.46$, $p = .488$, or one/two and three, $\beta = -0.35\,[-0.79, 0.07]$, $SE = 0.22$, $\chi^2(1) = 0.95$, $p = .317$, suggesting that this effect was not related to other aspects of the task (see Figure 8). The model explained 17.69% of the variance in the data without the random effects, and 20.68% when they were included ($R_m^2 = 0.1769$, $R_c^2 = 0.2068$). This agent advantage, and the lack of an interaction with other variables, may be due to the fact that agents precede patients in the sentences produced at test. Alternatively, it could also be that the features for tracking roles are biased for agents (e.g., agents in push events tend to move towards the patient). Once these role-related features are assigned to a particular pointer, there is no extra burden in maintaining this information in tracking, which could explain why there is no interaction with other visual variables.

General discussion

Linguistic accounts of thematic roles focus on the identification of roles when dealing with entities that differ in sentience, volition, cause, and other semantic features (e.g., Dowty, 1991; Fillmore, 1967; Jackendoff, 1987). However, sentience and volition are internal mental states of others that cannot be directly observed, only inferred based on the available data such as the physical appearance or movement patterns of the entity (Scholl & Gao, 2013). It is also unclear how features that can be inferred from visual information, such as *cause*, are stored, and used to assign thematic roles. To examine this issue, we developed a Push-MOT paradigm based on MOT studies, which have previously shown that viewers can monitor a small number of objects in parallel (e.g., Pylyshyn & Storm, 1988) and track a single agent and patient interaction (e.g., Gao et al., 2009). In the three experiments presented here, the participants watched displays in which nine identical circles moved in random patterns, before some of these circles engaged in push events. During each push, the uninvolved objects were stationary, so the agent and patient of each event was clear. These pushes would occur up to three times per trial, with different objects in each event. Afterwards, all nine circles would resume their random movement patterns for approximately 10 s. According to a conceptual account, thematic role information is passed immediately to the conceptual system and stored there as role-concept links. As the circles were identical and the conceptual system did not have continuous access to features that could uniquely identify each circle, the participants' descriptions of the events should be random. Alternatively, from the perspective of a visual account, the thematic role features are stored with the object pointers that track each circle. As the pointer's position was updated to follow the circle, the role features were also updated. If tracking has been maintained until the test event, then thematic role information was available and this could be used to map the correct colour referent into the agent and patient sentence position. If tracking was lost, then description accuracy for each target should be random.

The results of all three experiments showed that viewers 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 in which 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 (e.g., Tombu & Seiffert, 2008). However, in both the cases, participants were consistently above chance in tracking and describing two push events in parallel, matching the capacity estimates of object-tracking studies using similar display settings (Alvarez & Franconeri, 2007). 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. Overall accuracy in the third experiment was lower than in the first two studies, which may be due to the difficulty in performing the motion extrapolation needed to bind the disappearing and reappearing objects (Fencsik et al., 2007; Howe et al., 2012; Luu & Howe, 2015). Across the three experiments, the participants' role tracking capacity appeared to be consistent with the reports of other object-tracking studies. The above chance accuracy as well as the fact that visual manipulations influenced the results suggests that thematic roles were maintained in the visual system during tracking.

There was a strong negative effect of the number of pushes in the first two 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 object) tracking capacity estimated for the present stimuli (Alvarez & Franconeri, 2007). They often shifted their attention to the final event, which is a logical strategy as this push needs to be retained for the least amount of time before test and is, therefore, less likely to be lost during tracking. A similar negative effect was observed in Experiment 3, but it did not reach statistical significance, which was partly because occlusion occurred during the periods of random motion and the earlier pushes were more impacted by this than later events. Overall, this effect of set size on response accuracy is consistent with previous object-tracking research that reports a linear decrease in performance as the number of targets increases (e.g., Oksama & Hyönä, 2004; Pylyshyn & Storm, 1988). In addition to these group-level effects, there was also evidence of individual differences in performance; some participants showed high description accuracy across all of the conditions, whereas others produced fewer correct sentences overall (Oksama & Hyönä, 2004). The fact that description accuracy decreased with additional push events and that participants appeared to adopt strategies to deal with these limitations provides evidence for a limited capacity system for storing role-related information during tracking.

Furthermore, accuracy in identifying agents was superior to that of patients across the three experiments. One possible explanation is that agents were produced before patients at test, creating a temporal advantage for agents. It is also possible that the agent advantage originated in the role-related features being tracked with the targets. Several features, like chasing subtlety (Gao et al., 2009), are more diagnostic for agents than patients. For instance, the wolf moves directly towards the sheep, but a sheep can run in multiple directions to escape the wolf (see also Cohn & Paczynski, 2013). Similarly, the push actions in our task involved a self-propelled agent that moves 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, Kaufman, & Baillargeon, 2009). Therefore, the availability and salience of relational and non-relational thematic role cues might produce biases for particular roles, which could explain the consistent preference for agents in this work. Further research is needed to determine the exact features that create this linguistic agent advantage.

One of the motivations for the present line of research arose out of the limitations of neural network models of language. Language users have productive syntactic rules; if you can say *Jerry loves Zenon*, then you can also say *Zenon loves Jerry*, even without previous experience with this sentence. However, neural network models find this kind of generalisation to be difficult as they encode language regularities with slow biologically based neural learning mechanisms (Fodor & Pylyshyn, 1988; Marcus, 1998). To address this problem, Chang (2002) developed a neural network model of language production called the Dual-path model, which used fast-changing variables to support syntactic generalisations. This model has successfully explained a range of different language phenomena from adult sentence production (e.g., structural priming: Chang, Dell, & Bock, 2006; heavy noun phrase shift/accessibility in English and Japanese: Chang, 2009), sentence comprehension (Fitz & Chang, 2019), and language acquisition (verb semantics: Twomey, Chang, & Ambridge, 2014; auxiliary inversion: Fitz & Chang, 2017; critical periods in second-language acquisition: Janciauskas & Chang, 2017). However, the fast-changing variables in this model cannot be supported by slow biological changes in neural circuitry. Therefore, it was argued that these variable binding abilities may have originally evolved for other functions (Chang, 2002), such as the fast-binding object pointers used for location tracking, before being later adapted for thematic roles in language meaning. This predicts that thematic role processing will be influenced by the limitations of the object-tracking system, and the present research provides support for this hypothesis. Furthermore, the participants could easily describe the push events at test, which is consistent with the model's claim that object pointers can be activated by the language system to support sentence generation. More generally, this work argues that humans create abstract syntactic representations that are compositional and productive (Fodor & Pylyshyn, 1988) by adapting variables that evolved in the visual system to support object tracking (Pylyshyn & Storm, 1988).

Limitations on thematic role bindings in object tracking may also help to explain apparent limits on the number of arguments that can be attached to a verb (i.e., verb subcategorization or valency). For example, to understand a *giving* event, we need three arguments: the giver, the object given, and the recipient of the object. However, there are other elements that are not required (e.g., the location where the event took place), which are expressed with adjunct phrases instead. As speakers can produce complex sentences with multiple arguments from multiple verbs, it is not clear why individual verbs are restricted in the number of arguments. Evidence for universal limitations on verb arguments comes from the ValPaL database (Hartmann, Haspelmath, & Taylor, 2013), which has argument structures for a sample of 80 verbs in 36 typologically different languages. Out of 574 basic frames in the corpus, no verb has more than 4 arguments, and 97% of the verbs have 3 or fewer arguments. Additional evidence comes from languages with serial verbs or complex verbs, where two verbs are combined. Although the combination of these verbs could create situations in which six or more arguments are possible, in actuality, the arguments are fused to limit the maximum number in the meaning of the complex verb (Bisang, 2009). One explanation for these typological observations is that visual object tracking may exert limits on verb subcategorization. To understand a visual action, the entities involved in the action must be tracked during their interaction. If tracking role-related features has a limited capacity, as we have found in this study, then it is natural that verb arguments will be restricted to four or fewer items. Thus, these limits on verb subcategorization in linguistic typology provide independent evidence for limits on action understanding.

The question of how we understand the interactions between objects in the world is still not well understood. Our minds will find causal relationships in even impossible events (e.g., a magic show in which tapping a hat causes a bird to appear). The mismatch between our visual input and our subjective experience creates a fundamental problem in explaining how we learn about the world (Hume, 1748/2000). One answer to this issue was to assume that we cannot know the world in itself (Kant, 1781/1997, 1783/2004), but rather experience it filtered through a priori biases such as those in the visual system. Across three experiments, we found evidence that the encoding of causality in multiple push events was filtered by biases in the visual object-tracking system, which in turn influenced the accuracy of thematic role information in language production. This provides a processing account of how vision and language are linked (Bloom, Peterson, Nadel, & Garrett, 1999; Jackendoff, 1983; Langacker, 1987; Miller & Johnson-Laird, 1976; Osgood, 1952; Osgood & Bock, 1977; Slobin, 1996; Talmy, 1988; Wolff, 2007) and a new paradigm for studying this system. By bridging between studies of dynamic visual processing and linguistic approaches to meaning, this work offers new ways to study how language evolved to fit the mechanisms provided by the visual brain.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a grant from the Economic and Social Research Council (number 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 ESRC International Centre for Language and Communicative Development (LuCiD). The support of the Economic and Social Research Council is gratefully acknowledged. Franklin Chang is also supported by KAKEN Grant 19K12733 from the Japanese Society for the Promotion of Science.

ORCID iDs

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

References

Allen, R., Mcgeorge, P., Pearson, D., & Milne, A. B. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology*, *18*, 337–347. Retrieved from https://doi.org/10/djhwpz

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*, 1101–1116. Retrieved from https://doi.org/10/fcns38

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm." *Cognition, 93*, B79–B87. Retrieved from https://doi.org/10/cv72rg

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

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), Article 14. Retrieved from 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*, 461–476. Retrieved from https://doi.org/10/ctcdg2

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, *10*, 949–963. Retrieved from 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*, 255–278. Retrieved from 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*, 313–331. Retrieved from 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. Retrieved from https://doi.org/10/gcrnkw

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*, 985–995. Retrieved from https://doi.org/10/cf6czw

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, *9*(7), Article 9. Retrieved from https://doi.org/10/fcd32p

Bisang, W. (2009). Serial verb constructions. *Language and Linguistics Compass*, *3*, 792–814. Retrieved from https://doi.org/10/bpcwnr

Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.

Bloom, P., Peterson, M. A., Nadel, L., & Garrett, M. F. (1999). *Language and space*. Cambridge: The MIT Press.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349–354. Retrieved from https://doi.org/10/dcxnpn

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*, 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*, 374–397. Retrieved from https://doi.org/10/ftwhs5

Chang, F., Bock, J. K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, *90*, 29–49. Retrieved from https://doi.org/10/d95nsb

Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234–272. Retrieved from https://doi.org/10/cxcrx7

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of agents: The processing of semantic roles in visual narrative. *Cognitive Psychology*, *67*, 73–97. Retrieved from https://doi.org/10/gdh8xw

Csibra, G., Gergely, G., Bıŕó, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition*, *72*, 237–267. Retrieved from https://doi.org/10/d6h395

Dittrich, W. H., & Lea, S. E. (1994). Visual perception of intentional motion. *Perception*, *23*, 253–268. Retrieved from https://doi.org/10/b8f2gw

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*, 547–619. Retrieved from https://doi.org/10/bg3ktc

Drew, T., Horowitz, T. S., & Vogel, E. K. (2013). Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition*, *126*, 213–223. Retrieved from https://doi.org/10/f4nx8c

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience*, *28*, 4183–4191. Retrieved from https://doi.org/10/dbq653

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, *108*, 201–209. Retrieved from 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*, 19–19. Retrieved from 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*, 567–577. Retrieved from https://doi.org/10/d8npg8

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*, 516–547. Retrieved from 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, NY: 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. Retrieved from https://doi.org/10/gd59wh

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52. Retrieved from https://doi.org/10/gf2hx2

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, *27*, 235–244. Retrieved from https://doi.org/10/bt3wkw

Flombaum, J. I., Kundey, S. M., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in Rhesus Macaques: A study of the tunnel effect. *Psychological Science*, *15*, 795–800. Retrieved from https://doi.org/10/dn839q

Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in visual tracking through occlusion: The high-beams effect. *Cognition*, *107*, 904–931. Retrieved from https://doi.org/10/dtj5nh

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

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*, 920–925. Retrieved from https://doi.org/10/d66qbw

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*, 691–702. Retrieved from https://doi.org/10/fzvmmk

Frankenhuis, W. E., House, B., Clark Barrett, H., & Johnson, S. P. (2013). Infants' perception of chasing. *Cognition*, *126*, 224–233. Retrieved from 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*, 41–47. Retrieved from 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. Retrieved from 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*, 154–179. Retrieved from 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*, 669–684. Retrieved from https://doi.org/10/fhf32z

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193. Retrieved from https://doi.org/10/fvjkg5

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Green, C., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*, 217–245. Retrieved from https://doi.org/10/b5mw6c

Gruber, J. S. (1965). *Studies in lexical relations* (Doctoral thesis). Massachusetts Institute of Technology, Cambridge.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*, 151–167. Retrieved from https://doi.org/10/bmk9c5

Hartmann, I., Haspelmath, M., & Taylor, B. (Eds.) (2013). *Valency patterns Leipzig*. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*, 243–259. Retrieved from https://doi.org/10/ftcck7

Hoffmann, A., Rüttler, V., & Nieder, A. (2011). Ontogeny of object permanence and object tracking in the carrion crow, Corvus corone. *Animal Behaviour*, *82*, 359–367. Retrieved from https://doi.org/10/dj386k

Hohenstein, S., & Kliegl, R. (2019). Remef: Remove partial effects (R package version 1.0.6.9000). Retrieved from https://github.com/hohenstein/remef/

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*, 516–523. Retrieved from 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*, 172–184. Retrieved from 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*, Article 307. Retrieved from 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), Article 10. Retrieved from 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. Retrieved from 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*, 2375–2380. Retrieved from https://doi.org/10/bqvj2d

Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 414–428. Retrieved from https://doi.org/10/fx5ct3

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*, 628–636. Retrieved from https://doi.org/10/bvf58p

Hume, D. (2000). *An enquiry concerning human understanding* (T. L. Beauchamp, Ed.). Oxford, UK: Clarendon Press. (Original work published 1748)

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*, 597–607. Retrieved from 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), Article 1. Retrieved from https://doi.org/10/d4cz9d

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge: The MIT Press.

Jackendoff, R. S. (1983). *Semantics and cognition*. Cambridge: The MIT Press.

Jackendoff, R. S. (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, *18*, 369–411. Retrieved from https://www.jstor.org/stable/4178548

Janciauskas, M., & Chang, F. (2017). Input and age-dependent variation in second language learning: A connectionist account. *Cognitive Science*, *42*, 519–554. Retrieved from https://doi.org/10/gcrnm4

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*, 944–946. Retrieved from 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*, 122–149. Retrieved from https://doi.org/10/cfx7xr

Kant, I. (1997). *The critique of pure reason* (P. Guyer & A. W. Wood, Eds.). Cambridge, UK: Cambridge University Press. (Original work published 1781)

Kant, I. (2004). *Prolegomena to any future metaphysics* (G. Hatfield, Ed.). Cambridge, UK: Cambridge University Press. (Original work published 1783).

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*, 346–368. Retrieved from https://doi.org/10/ff342b

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182. Retrieved from https://doi.org/10/cxhhsw

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*, 481–529. Retrieved from 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*, 519–543. Retrieved from 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*, 95–127. Retrieved from https://doi.org/10/fmz4fr

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*, 1135–1140. Retrieved from https://doi.org/10/fwfmtc

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Lappin, S., & Fox, C. (2015). *The handbook of contemporary semantic theory* (2nd ed.). Sussex, UK: John Wiley & Sons.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*, 265–288. Retrieved from https://doi.org/10/dc4dh2

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502. Retrieved from 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*, 601–608. Retrieved from https://doi.org/10/cwspvm

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, *58*, 441–486. Retrieved from 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*, 1919–1929. Retrieved from https://doi.org/10/f7mh82

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge: The MIT Press.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, *37*, 243–282.

Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, *33*, 449–496. Retrieved from https://doi.org/10/bj7583

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, *12*, 137–176. Retrieved from https://doi.org/10/dp8v4q

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*, 1417–1429. Retrieved from https://doi.org/10/c934h6

Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2013). A single unexpected change in target- but not distractor motion impairs multiple object tracking. *i-Perception*, *4*, 81–83. Retrieved from https://doi.org/10/gcdz36

Michotte, A. (1946). *The perception of causality*. New York, NY: Basic Books.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Belknap Press.

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 inattentional blindness. *Psychological Science*, *12*, 9–17. Retrieved from https://doi.org/10/fmbz72

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), Article 20170213. Retrieved from https://doi.org/10/gddpnq

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. Retrieved from https://doi.org/10/f4pkjx

Oakes, L. M. (1994). Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, *30*, 869–879. Retrieved from https://doi.org/10/cs7q44

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*, 193–207. Retrieved from https://doi.org/10/dqgznw

O'Connell, S., & Dunbar, R. I. M. (2005). The perception of causality in chimpanzees (Pan spp.). *Animal Cognition*, *8*, 60–66. Retrieved from https://doi.org/10/fbcjn8

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*, 631–671. Retrieved from 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*, 237–283. Retrieved from 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. Retrieved from https://doi.org/10/f74g45

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, *49*, 197–237. Retrieved from 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.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97. Retrieved from 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*, 801–822. Retrieved from 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*, 179–197. Retrieved from https://doi.org/10/bt3rft

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rochat, P., Morgan, R., & Carpenter, M. (1997). Young infants' sensitivity to movement information specifying social causality. *Cognitive Development*, *12*, 537–561. Retrieved from https://doi.org/10/b2h2dc

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, *3*, 6–21. Retrieved from https://doi.org/10.1167/3.1.2

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*, 112–143. Retrieved from https://doi.org/10/c6smbn

Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. Rutherford & V. A. Kuhlmeier (Eds.), *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 197–230). Cambridge: The MIT Press.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290. Retrieved from https://doi.org/10/c6ccs9

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*, 159–177. Retrieved from https://doi.org/10/ch9wn3

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*, 299–309. Retrieved from https://doi.org/10/dvnhr2

Sekuler, R., McLaughlin, C., & Yotsumoto, Y. (2008). Age-related changes in attentional tracking of multiple moving objects. *Perception*, *37*, 867–876. Retrieved from https://doi.org/10/cr7vgv

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*, 3–6. Retrieved from https://doi.org/10/dxwwh8

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074. Retrieved from https://doi.org/10/gdh8td

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.

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*, 113–142. Retrieved from https://doi.org/10/d6wq4q

St John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217–257. Retrieved from https://doi.org/10/c5z7jt

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*, 315–325. Retrieved from https://doi.org/10/fcfjzd

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*, 49–100. Retrieved from https://doi.org/10/dwn9j3

Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, *108*(1), 1–25. Retrieved from https://doi.org/10/b3qqr3

Tran, A., & Hoffman, J. E. (2016). Visual attention is required for multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 2103–2114. Retrieved from https://doi.org/10/gdh8tc

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. Retrieved from https://doi.org/10/bgz2qm

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*, 1188–1195. Retrieved from https://doi.org/10/bmc65j

Trick, L. M., Perl, T., & Sethi, N. (2005). Age-related differences in multiple-object tracking. *The Journals of Gerontology, Series B*, *60*(2), P102–P105.

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. Retrieved from https://doi.org/10/f6f78k

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.

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.

Ward, E. J., & Scholl, B. J. (2015). Inattentional blindness reflects limitations on perception, not memory: Evidence from repeated failures of awareness. *Psychonomic Bulletin & Review*, *22*, 722–727. Retrieved from https://doi.org/10/f7br4f

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*, 419–433. Retrieved from https://doi.org/10/c9qkrd

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, *136*, 82–111. Retrieved from 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. Retrieved from https://doi.org/10/f53cft

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*, 295–340. Retrieved from https://doi.org/10/dmz7g4

Young, M. E., & Sutherland, S. (2009). The spatiotemporal distinctiveness of direct causation. *Psychonomic Bulletin & Review*, *16*, 729–735. Retrieved from 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*, 553–566. Retrieved from https://doi.org/10/c3nmbt