Cognition 250 (2024) 105868

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Asymmetries in encoding event roles: Evidence from language and cognition

Ercenur Ünal [a,*], Frances Wilson [b], John Trueswell [c], Anna Papafragou [d]

[a] Department of Psychology, Ozyegin University, Istanbul, Turkey
[b] Department of Psychological and Brain Sciences, University of Delaware, Newark, DE, USA
[c] Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA
[d] Department of Linguistics, University of Pennsylvania, Philadelphia, PA, USA

ARTICLE INFO

Keywords:
Event
Event perception
Event cognition
Thematic roles
Event roles
Eye-tracking
Argument structure
Causatives

ABSTRACT

It has long been hypothesized that the linguistic structure of events, including event participants and their relative prominence, draws on the non-linguistic nature of events and the roles that these events license. However, the precise relation between the prominence of event participants in language and cognition has not been tested experimentally in a systematic way. Here we address this gap. In four experiments, we investigate the relative prominence of (animate) Agents, Patients, Goals and Instruments in the linguistic encoding of complex events and the prominence of these event roles in cognition as measured by visual search and change blindness tasks. The relative prominence of these event roles was largely similar—though not identical—across linguistic and non-linguistic measures. Across linguistic and non-linguistic tasks, Patients were more salient than Goals, which were more salient than Instruments. (Animate) Agents were more salient than Patients in linguistic descriptions and visual search; however, this asymmetrical pattern did not emerge in change detection. Overall, our results reveal homologies between the linguistic and non-linguistic prominence of individual event participants, thereby lending support to the claim that the linguistic structure of events builds on underlying conceptual event representations. We discuss implications of these findings for linguistic theory and theories of event cognition.

1. Introduction

Humans are incredibly adept at communicating their thoughts about the world. Often these thoughts involve objects and their properties, as well as a dynamic stream of activity including constantly unfolding units known as *events*. Events have been defined as “a segment of time at a given location that is conceived by an observer to have a beginning and an end” (Zacks & Tversky, 2001, p.3; see also Carlson, 1998; Elman, 2009; Shipley, 2008; Yates et al., 2023). In this view, conceptualizing an event involves recognizing changes along a set of properties (Radvansky & Zacks, 2014), including spatial or temporal elements, states of objects or interactions among the entities involved, as well as more abstract properties (Mathis & Papafragou, 2022). When communicating about events, people need to make choices about what information to include about these properties. How does the way people conceptualize events relate to the way they communicate about them?

Several linguists have suggested that linguistic structure reflects the underlying organization of human event cognition (Jackendoff, 1983, 1990; Pustejovsky, 1995; Talmy, 1985). Furthermore, this idea has characterized influential theories of language production (Levelt, 1989, see also Bock et al., 2004; Lashley, 1951; Paul, 1886/1970; Wundt, 1900/1970) and acquisition (Gleitman, 1990; Jackendoff, 1996; Miller & Johnson-Laird, 1976; Pinker, 1989; see Papafragou & Grigoroglou, 2019 for an overview). On this view, a conceptual representation of broad details of an event and its components in cognition is considered to be both the basis of ‘thinking for speaking’ (Slobin, 1996) and a prerequisite for the acquisition of linguistic meaning. In this paper, we focus on a key component of events, the representation of event participants (or *thematic roles* within linguistic theory), and explore the hypothesis that cognition offers structured representations of events that are strikingly similar to the structure proposed to characterize events in language.

1.1. *Thematic roles in language*

Thematic roles describe the semantic relation that each syntactic

---

* Corresponding author address at: Ozyegin University, Nişantepe Mahallesi Orman Sokak 34794 Çekmeköy, Istanbul, Turkey.
E-mail address: ercenur.unal@ozyegin.edu.tr (E. Ünal).

https://doi.org/10.1016/j.cognition.2024.105868
Received 18 September 2022; Received in revised form 20 May 2024; Accepted 28 June 2024

constituent (typically a noun phrase (NP) or a prepositional phrase (PP)) has with the category that projects it, typically a verb in a verb phrase (see e.g., Fillmore, 1968; Jackendoff, 1990). The following sentences describe a simple (1) or a more complex (2) event:

(1) A woman is hugging a cat.
(2) A man is hitting a ball into a basket with a tennis racquet.

In sentence (1) the NP "a woman" functions as the *Agent*, or the causer of the action, and the NP "a cat" functions as the *Patient*, or the entity directly affected by that action. Events with more participants include other roles and are relatively more complex. Consider caused motion events, as the one described in (2): in addition to a human Agent ("a man") and an inanimate Patient ("a ball"), this event also includes an *Instrument* (the PP "with a tennis racquet"), or the means with which the Agent makes the Patient undergo a change, as well as an end result of the event or the *Goal* (the PP "into a basket"). Thus, the underlying structure of the sentences in (1) and (2) can be represented as (1′) and (2′) respectively:

(1') [A woman] AGENT is hugging [a cat] PATIENT.
(2') [A man] AGENT is hitting [a ball] PATIENT [into a basket] GOAL [with a tennis racquet] INSTRUMENT.

The treatment of thematic roles in linguistics is famously thorny, especially if one moves beyond analysis of specific verbs (McRae et al., 1997) and attempts to define Agents, Patients and so on in more abstract ways (for a review, see Rappaport Hovav, 2017). On one influential proposal (Dowty, 1991), typical Agent and Patient roles correspond to stable and recurring clusters of meaning properties of verbs across the lexicon (e.g., sentience and volition for typical Agents, and causal affectedness or change of state for typical Patients). Other theories define roles in terms of a constituent's particular position within decomposed verb-semantic structures (Hale & Keysar, 1993; Jackendoff, 2002; Pinker, 1989). Despite these disagreements, thematic roles provide a useful framework for representing linguistic generalizations about verb-argument relations (Davis, 2011; Levin, 2014).

For present purposes, we focus on a proposal within linguistic theory for a relative ranking of thematic roles, referred to as the *Thematic Hierarchy* (Grimshaw, 1981; Pinker, 1989). In this view, roles that contribute to structured event representations to a greater extent are ranked higher in the hierarchy and are more salient (Levin & Rappaport Hovav, 2005). From a linguistic perspective, salient roles are mentioned more frequently, and encoded in specific syntactic positions, such as subject or object (Grosz et al., 1995), or at the beginning or end of an utterance (Gernsbacher, 1989; Meyer et al., 1998). According to a prominent characterization of the Thematic Hierarchy (Baker, 1997; Jackendoff, 1990), Agents are considered to be more salient than Patients; Patients are considered to be more salient than Goals; and Instruments would be considered the least salient among these roles. This characterization is largely based on the linguistic behavior of these roles. Specifically, Agents and Patients are more likely to be encoded in syntactic positions that are considered to be obligatory in sentences. In many languages that use active voice more frequently, Agents are encoded as subjects and Patients are encoded as direct objects of the verb. By contrast, Goals and Instruments are much less likely to be encoded in such positions. Instruments, especially, are considered to be the least salient role as they are encoded in language through a variety of structures and syntactic positions and are highly unlikely to be selected as verb arguments (Rissman et al., 2015). In fact, Baker (1997) refers to Instruments as secondary roles.

It seems reasonable to assume that linguistic Agents, Patients and related roles have cognitive counterparts that can be used to capture the internal structure ('who-did-what-to whom') of an event. Indeed, several researchers have proposed that linguistic roles such as Agents and Patients map relatively directly onto underlying conceptual event roles in cognition, at least in the most typical cases (Dowty, 1991; Levin & Pinker, 1991; Levin & Rappaport Hovav, 2005; Pinker, 1989). For instance, Jackendoff (1990) claimed that thematic roles are "relational notions defined structurally over conceptual structure" (p. 47) and "every putative thematic role assignment must be justified on the grounds of its place in conceptual structure" (p. 50). Furthermore, Baker (1997) takes as a point of departure "the common practice of assuming that thematic roles are part of the conceptual system" and goes on to propose a hypothesis on which "there must be a homomorphic, perhaps even an isomorphic relationship between this aspect of the conceptual system and the corresponding linguistic representation" (pp.38–39).

This line of reasoning has implications for the Thematic Hierarchy. Even though the hierarchy was mainly formulated to capture the linguistic behavior of sentential constituents, it has often been assumed that this hierarchy reflects the mapping between linguistic and cognitive representation of events (Dowty, 1991; Grimshaw, 1981; Pinker, 1989; Rissman & Majid, 2019; Strickland, 2017). If so, the relative ranking of thematic roles should be reflected in non-linguistic conceptualization of the same event participants. In the next section, we review the literature to date that suggests that the non-linguistic conception of events directly relates to linguistic organization (see also Rissman & Majid, 2019; Ünal, Ji, & Papafragou, 2021 for reviews) before turning to our own empirical investigation of event roles in language and cognition.

1.2. Representing event roles in language and cognition

Infant work inspired by the linguistic analysis of event structure supports the presence of a correspondence between the linguistic and conceptual organization of events (see Göksun et al., 2010; Wagner & Lakusta, 2009 for an overview). Pre-linguistic infants discriminate simple causative events that have Agents from non-causative versions of the same events that do not have Agents (Leslie & Keeble, 1987; Saxe et al., 2005). Furthermore, infants recognize changes to Agents and Patients of causative events (Cohen & Oakes, 1993; Golinkoff & Kerr, 1978; Leslie & Keeble, 1987). Thus, some understanding of event participants that fill the roles of Agents and Patients emerge before children learn to encode these roles in language (see also Baldwin et al., 2001; Cohen et al., 1999; Gergely et al., 1995; Gergely & Csibra, 2003; Göksun et al., 2013; Oakes & Cohen, 1990; Wolff, 2007, 2008; Woodward, 1998; Yin & Csibra, 2015).

A further piece of evidence suggesting a tight link between linguistic and conceptual event role structure comes from speech production studies that have examined participants' eye-movements. This work has shown that roles relevant for linguistic descriptions of events can be extracted rapidly. In an early demonstration, Griffin and Bock (2000) showed that when participants are asked to detect the entity directly undergoing the action (i.e., the Patient), they showed a preference for fixating on the Patients over Agents within 300 ms of inspecting line drawings of simple events involving two animate entities. In another study, participants were briefly presented with drawings of *shooting* events and were able to name both the action and the entity performing the action (i.e., the Agent) after viewing the event for 200 ms (Dobel et al., 2007). Nevertheless, participants were better at identifying both the action and the Agent in coherent scenes as opposed to incoherent scenes (see also Dobel et al., 2011; Zwitserlood et al., 2018). Finally, when viewing *reaching-to-grasp* events in preparation for identifying the person performing the reaching and the object being reached, participants fixated on the Agents before the Patients (Webb et al., 2010). Other work showed that the relation between linguistic message formulation and attention to event participants may be flexible. For instance, briefly presented attentional cues to an event participant has been shown to affect linguistic choices (Gleitman et al., 2007, see also

Kuchinsky et al., 2011). Furthermore, after describing an event using active sentences people were more likely to first fixate on the Agent of an unrelated event, whereas after describing an event using Passive sentences they were more likely to first fixate on the Patient of an unrelated event (Sauppe & Flecken, 2021), suggesting that recent linguistic experience may modulate attention allocated to event participants.

In a set of studies more relevant for present purposes, Hafri et al. (2013) tested if the conceptual roles of Agent and Patient and the events comprised by them could be rapidly extracted from brief presentations of visual stimuli. Participants saw naturalistic photographs of a wide range of two-participant events and were asked to identify the event category, the Agent, the Patient, or a combination of event category and Agent and Patient roles. The results revealed that participants successfully recognized event categories, the event roles and combinations of the two even from the briefest visual displays of 37 ms. A later study demonstrated that Agent and Patient roles of depicted events can be extracted even when participants' attention is occupied by a task that does not encourage the encoding of event roles (Hafri et al., 2018). Finally, in a recent demonstration, Rissman and Lupyan (2022) asked participants to categorize pictures of cartoons depicting two-participant (Agent-Patient) events and found that participants were able to categorize the pictures based on who was the Agent and who was the Patient. Nevertheless, there was some individual variation as some participants failed to categorize the pictures based on the Agent-Patient dimension even after feedback. Importantly, the properties of typical Agents and Patients proposed by linguistic theories (Dowty, 1991) predicted categorization accuracy and speed for those participants who successfully learned to categorize the pictures according to the Agent-Patient dimension, suggesting that Agent and Patient categories may have similar structures in language and cognition.

It has also been shown that the relative salience of key motion event components is similar across language and non-linguistic cognition. When describing motion events, both children and adults frequently encode the Goal or the endpoint of the motion but tend to omit the Source or the starting point of the motion (Lakusta & Landau, 2005; Papafragou, 2010; see also Do et al., 2020, 2022; Johanson et al., 2019; Landau & Zukowski, 2003; Chen et al., 2024). Crucially, the bias in expressing Goals over Sources extends to non-linguistic cognition. For instance, pre-linguistic infants encode Goals over Sources in motion events (Lakusta et al., 2007, 2017; Lakusta & Carey, 2015). Both children and adults are also less likely to recognize changes to a Source object compared to changes to a Goal object (Chen et al., 2024; Papafragou, 2010) although the Goal-bias in memory is stronger for events involving intentional animate agents (Lakusta & Landau, 2012) and is subject to task effects (Chen et al., 2024). These findings lend support to the idea that some event components might be prioritized over others in similar ways across language and cognition.

Finally, both older and recent findings suggest that the relative salience of event roles varies by showing that Instruments are not considered to be central for event structure in language. When retelling stories adult English speakers frequently omit instruments, especially when they are highly typical (e.g., *stabbing* <u>*with a knife*</u> vs. *stabbing* <u>*with an icepick*</u>; Brown & Dell, 1987; Lockridge & Brennan, 2002). More recent work shows that speakers omit instruments from their event descriptions even when the instruments are atypical and their interlocutor does not have visual access to the event (Grigoroglou & Papafragou, 2019a). However, linguistic judgment tasks reveal that the Instrument role is not considered to be equally peripheral across all events. For instance, adult English speakers were more likely to consider the Instrument to be a key event participant for verbs that required an instrument (e.g., *slice*) compared to verbs that merely allowed the presence of an instrument (e.g., *drink*; Rissman et al., 2015; Rissman & Rawlins, 2017). A cross-linguistic extension of this study revealed similar patterns in adult speakers of Spanish and Mandarin Chinese (Rissman et al., 2019).

1.3. The present study

The work just reviewed leaves open two issues that need to be addressed in order to fully evaluate the extent of the homology between event structure in language and cognition. First, prior work has focused on specific types of events (e.g., *shooting events*, Dobel et al., 2007; *reaching-to-grasp events*, Webb et al., 2010) that are relatively simple and thus have a limited number of event participants (mostly, Agents and Patients). Even though a limited number of studies have investigated roles beyond Agents and Patients (Goals and Sources: Lakusta & Landau, 2005, 2012; Papafragou, 2010; Instruments: Rissman et al., 2015, 2019, 2022), this work has typically studied those event components individually and not in relation to other event roles. Thus, the status of roles beyond Agents and Patients, their conceptual counterparts (Rissman & Majid, 2019), and their ranking with respect to each other in more complex events (Fillmore, 1968; Wolff, 2007) remain debated.

Second, as the previous overview shows, event cognition and language production have been typically studied independently, by different communities of scholars and through different methodologies. For instance, one line of work on event identification has used naming tasks measuring the accuracy and speed of single-word descriptions of the action or event participants (Dobel et al., 2007; Zwitserlood et al., 2018). However, this work did not consider non-linguistic measures of event cognition. Conversely, linguistically inspired work on event cognition did not include directly matched language production tasks (Hafri et al., 2013, 2018; Webb et al., 2010). Similar issues characterize the study of Instruments, with one line of work focusing on the frequency of mention of Instruments in language production (Brown & Dell, 1987; Grigoroglou & Papafragou, 2019a, 2019b; Lockridge & Brennan, 2002) and a separate line of work using linguistic judgments to investigate the status of Instruments (Rissman et al., 2015, 2019, 2022; Rissman & Rawlins, 2017). It is important to obtain more direct evidence connecting the organization of events in language to their conceptual structure.

In the present study, we pursue a direct test of the hypothesized homologies between event structure in language and cognition by comparing linguistic and non-linguistic measures of the salience of multiple event roles within the same event stimulus. We focus on complex multi-component causative events that involve a human Agent moving a Patient towards a Goal using an Instrument (e.g., a man hitting a ball into a bucket with a golf club). Causative events are considered to be complex as they involve a subordinate event (i.e., an agent performing an action on an object), a main event (i.e., an object changing location) and the relation between the two (Talmy, 2000; see also Bunger et al., 2016). Because of this perceived relation, these sequences of actions are interpreted as part of a complex causative event rather than unrelated actions. This contrasts with simple events involving fewer roles, such as two participant Agent-Patient events (e.g., a woman hugging a cat) that cannot be further divided into smaller units that would be construed as events. Causative events provide an excellent case for empirically testing the proposed correspondences between multiple event roles in language and cognition due to their relative complexity and internal structure (see Ünal, Richards, et al., 2021 for a similar approach).

In Experiment 1, we explore the frequency of mention of Agents, Patients, Goals and Instruments in a language production task. In Experiments 2–4, we use the very same events to explore the salience of the same event components in non-linguistic tasks involving visual search (Griffin & Bock, 2000) and change detection (e.g., Rensink et al., 1997). We ask whether there is an internal hierarchy of event components, as predicted by the Thematic Hierarchy (Baker, 1997; Jackendoff, 1990), and whether that hierarchy is shared between language and cognition. On this view, Agents should be more salient than Patients which should be more salient than Goals, and Instruments should be the less salient than Goals. Alternatively, the relative salience of the event components may follow a different hierarchy across both language and cognition, or completely different hierarchies may emerge across linguistic and non-linguistic measures of salience. This last possibility would indicate that event roles in language are not homologous with conceptual event structure.[1]

2. Experiment 1: Linguistic description of causative events

In Experiment 1 we used a picture description task to assess the relative salience of event components in language. We asked participants to describe relatively complex caused motion events. Of interest was whether the role prominence asymmetries postulated by the Thematic Hierarchy (Baker, 1997; Jackendoff, 1990) would be reflected in the frequency of mention of each event component in language production.

2.1. Method

Participants. Twenty adult native speakers of English (16 females, Mean age = 18.95, SD = 1.06 months) participated in the experiment. All participants were undergraduate students at the University of Delaware and participated in the experiment for course credit.

Materials. Twenty-four test images were created using clip art images in Adobe Illustrator. The images depicted caused motion events, in which an Agent was using a tool or body part (Instrument) to move an object affected by the action (a Patient) to a destination (Goal). For instance, in one of the events, a man (Agent) was using a broom (Instrument) to push some dirt (Patient) into a dustpan (Goal; see Fig. 1 for an example and the Appendix for a complete list of events). The Agent of each action was always an adult human, and the Patient, Goal and Instrument were always inanimate objects.

Because (a)typicality of certain roles, especially more peripheral ones, such as Goals or Instruments, might affect how often these roles are mentioned in production (cf. Brown & Dell, 1987; Grigoroglou & Papafragou, 2019a; Lockridge & Brennan, 2002), we obtained typicality ratings for these stimuli from a separate group of 20 adult native speakers of English. These participants saw each of the 24 test images four times in total, each time with a circle around a different event component (Agent, Patient, Goal or Instrument). They indicated how typical the circled person or object was for the event shown in the picture on a 7-point scale. The scale was labeled with numbers from 1 to 7. The extreme values were labeled as 1 = "very atypical" and 7 = "very typical", respectively. Participants saw the images in a single randomized order, with the constraints that the same image did not occur two times in a row and that the same event component did not occur more than two times in a row. We compared the mean typicality ratings across Agents (M = 5.16, SD = 1.03), Patients (M = 4.48, SD = 1.90), Goals (M = 4.81, SD = 1.56) and Instruments (M = 4.29, SD = 1.61) to each other. A one-way ANOVA revealed that mean typicality ratings did not differ across components ($F(3,92) = 1.431$, $p < .239$).

Procedure. Participants were tested individually in the lab. Stimuli were presented on a Dell laptop computer, using *E*-Prime (Schneider et al., 2002a, 2002b). Participants were told that they would view images depicting actions and they would have to describe what happened in the image using a single sentence as soon as they saw the image.

Fig. 1. Example event: [A man] AGENT pushing [dirt] PATIENT [into a dustpan] GOAL [with a broom] INSTRUMENT.

Participants viewed each image for as long as needed to describe it. Once they finished describing the image, they pressed a button to proceed to the next trial. Individual items were presented in a different randomized order for each participant. Participants' descriptions were audio recorded. The experiment lasted approximately 10 min.

Coding. The data were coded to determine how often each event component was mentioned in participants' descriptions. Simple mention of an entity was not considered mention of an event component unless the entity was construed appropriately as the event Agent, Patient, Goal or Instrument. For instance, in an event in which a man kicked a boot into a suitcase, the sentence *There is a man and a suitcase* were not coded as Agent and Goal mentions since this description only names the objects and does not capture the relations between these entities or the roles filled by these entities in the event. Such uses corresponded to 0.6% of Agent entities, 2.5% of all responses for Patient entities, 3% of all responses for Goal entities, and 9% of all responses for Instrument entities.

Agents always appeared as Subject Noun Phrases (*A man is hitting a golf ball towards a bucket*), Patients as Direct Object Noun Phrases (*A man is pulling a pine tree towards a house*), and Goals as Prepositional Phrases (*A man pulling a* log *to a campfire*). Information pertaining to Instruments showed more syntactic variation. Instruments appeared in their canonical positions as Prepositional Phrases (*A man is hitting an apple with an umbrella;* 11% of all responses). Additionally, however, Instrument information was incorporated into Verbs (*The man* is *raking the* leaves; 23% of responses) or encoded as Direct Objects of verbs such as *use* (*The man is using an umbrella to hit an apple*; 4% of all responses). For present purposes, all the above uses were treated as Instruments. (The overall pattern of results remains the same under a stricter coding scheme that only codes Prepositional PPs as Instruments.)

2.2. Analytical strategy

Data from this (and all subsequent experiments) were analyzed with linear mixed effects modelling with crossed random intercepts for Subjects and Items. The models were fit with the *lme4* package (version 1.1.17; Bates et al., 2015) in *R* (version 4.2.2; R Core Team, 2022). Figures were produced using ggplot2 package (version 3.2.1, Wickham, 2016). Data and analysis code (for this and all subsequent experiments) are available at https://osf.io/rtc6m/.

In order to test the relative ranking of roles predicted by the Thematic Hierarchy (Agents > Patients > Goals > Instruments), the fixed effect of Component (and Condition in Experiments 2–4) was tested with three planned comparisons using forward difference contrast coding (Schad et al., 2020). This coding system compares the mean of the

---

[1] Beyond salience, of course, there are multiple factors that influence how often people mention an event role: e.g., a speaker's pragmatic understanding of what is most relevant or informative in a given context, information structural demands, as well as language-specific linguistic factors. Such factors have been found to affect the frequency of mention of event components, such as Agents (Ünal, Richards, et al., 2021), Instruments (Grigoroglou & Papafragou, 2019a), and Goals or Sources of motion (Bunger et al., 2012; Do et al., 2020). Our point is simply that, other things being equal, we do expect cognitively salient roles to be mentioned more often. We raise the role of pragmatic factors in Experiment 1 and return to these issues at the end of the paper.

dependent measure for one level of the fixed event to the mean of the adjacent level of the fixed effect. The first contrasts compares Agents to Patients (Agents numerically coded $+3/4$, Patients numerically coded $-1/4$, Goals numerically coded $-1/4$, Instruments numerically coded $-1/4$), the second contrast compares Patients to Goals (Agents numerically coded $+1/2$, Patients numerically coded $+1/2$, Goals numerically coded $-1/2$, Instruments numerically coded $-1/2$) and the third contrast compares Goals to Instruments (Agents numerically coded $+1/4$, Patients numerically coded $+1/4$, Goals numerically coded $+1/4$, Instruments numerically coded $-3/4$). Please see the Supplementary Material for pairwise comparisons with corrections for multiple comparisons for the fixed effect of Component (and Condition in Experiments 2–4) that were tested with *emmeans* (version 1.5.5–1; Lenth, 2021) and *multcomp* (version 1.4–16; Hothorn et al., 2008) packages.

2.3. Results

We tested how frequently each component was mentioned as a proportion of all trials. Fig. 2 presents the proportion of mention of each component. The dependent measure was binary values for mention ($1 =$ yes, $0 =$ no) at the trial level. The model revealed that Agents were mentioned more than Patients ($\beta = 0.771$, $SE = 0.285$, $z = 2.709$, $p = .007$) which were mentioned more than Goals ($\beta = 1.210$, $SE = 0.213$, $z = 5.675$, $p < .001$) which were mentioned more than Instruments ($\beta = 2.319$, $SE = 0.174$, $z = 13.332$, $p < .001$).

Given that there were no differences among the four components in terms of typicality, it is unlikely that this factor can explain the observed asymmetries. Nevertheless, to take a closer look at the role of typicality, we conducted a control analysis by comparing a model that only includes the fixed effect of Typicality to a model that includes the fixed effects of both Typicality and Component. The second model fit the mention data better than the first (χ^2 (3) $= 566.11$, $p < .001$). The model revealed a fixed effect of the factor Typicality ($\beta = 0.134$, $SE = 0.064$, $z = 2.088$, $p = .037$): across all items, as typicality of an event component increased, frequency of mention also increased. However, since the original asymmetries remained, the typicality of the person or object filling these event roles cannot explain the role asymmetries in our data.

Lastly, we turn to physical properties of the entities fulfilling different event roles. Since we did not control for component size in order to preserve plausibility, it is possible that the event components that were mentioned more frequently were larger. To evaluate this possibility, we measured the size of each component in each event as a percentage of image area using the Tobii Studio AOI tool. First, we compared the overall sizes of Agents ($M = 4.73$, $SD = 1.98$), Patients ($M = 4.18$, $SD = 2.90$), Goals ($M = 11.18$, $SD = 7.24$) and Instruments ($M = 5.01$, $SD = 3.91$) to each other. A one-way ANOVA revealed a main effect of Component ($F(3,92) = 12.97$, $p < .001$). Pairwise comparisons with corrections for multiple comparisons revealed that Goals were larger compared to Agents ($p < .001$), Patients ($p < .001$), and Instruments ($p < .01$). No other size contrasts were significant. Next, we compared a model included Size of the event component (for that specific item) as a fixed factor to a model that included fixed effects of both Size and Component. The second model fit the frequency of mention data better compared to the model that included only Size as a fixed effect (χ^2 (3) $= 615.77$, $p < .001$). The model revealed that all of the previously reported differences in frequency of mention remained the same. However, the fixed effect of Size was not statistically significant ($\beta = -0.029$, $SE = 0.015$, $z = -1.929$, $p = .054$). Thus, the asymmetries in the frequency of mention of Agents, Patients, Goals and Instruments were independent of the physical attributes of the people or objects filling in those roles.

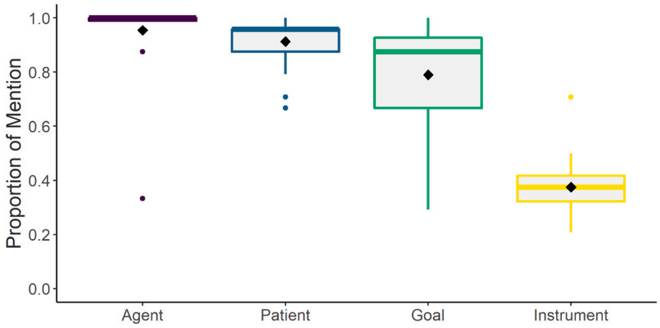


Fig. 2. Mean proportion of mention of event components (Experiment 1).
Note. Black diamonds represent the group mean. Horizontal colored bars represent the median, boxes represent the interquartile range (25th and 75th percentile), whiskers represent the range excluding the outliers, and colored dots represent outlier participants.

2.4. Discussion

Experiment 1 compared the frequency of linguistic encoding of Agents, Patients, Goals and Instruments, in a linguistic task. As expected, Agents were mentioned more frequently than Patients. Patients were followed by Goals which were followed by Instruments. Importantly, these asymmetries were not accounted by the typicality or physical features such as size of the people or objects fulfilling the event roles in our stimuli. Note that unlike in previous work (Brown & Dell, 1987; Grigoroglou & Papafragou, 2019a; Lockridge & Brennan, 2002), participants were overall more likely to mention typical components as opposed to atypical ones. These findings confirm the conclusion that the relative salience of event components in language is consistent with the asymmetries postulated by the Thematic Hierarchy (Baker, 1997; Jackendoff, 1990). They thus offer a starting point for investigating the non-linguistic prominence of these components in the studies that follow.

3. Experiment 2: Visual search

Experiment 2 was designed to investigate the relative salience of Agents, Patients, Goals and Instruments using a visual search task. This study adapted Griffin and Bock's (2000) "Find the Patient" task, and explicitly asked participants to identify specific event components by looking at them while monitoring their eye-movements around the scene. By comparing fixation probabilities to each event component over time, it would be possible to determine the accuracy and speed at which event components are detected and which event components are more salient than others. Of interest was whether the accuracy and/or speed at which Agents, Patients, Goals and Patients were identified would reflect the asymmetries proposed by the Thematic Hierarchy (Baker, 1997; Jackendoff, 1990).

3.1. Method

Participants. Forty undergraduate students at the University of Delaware (all native speakers of English) participated for course credit (22 females, Mean age = 19;1, SD = 14.82 months). None of these individuals had participated in Experiment 1.

Materials. Stimuli for target events consisted of a subset of 18 images that were used in Experiment 1. Compared to the larger set of 24 events, this subset minimized overlap between areas of interest, especially the overlap between Agents and Instruments. An additional set of 18 caused motion events were used as fillers. Filler events were similar to target events, except that they did not have minimal overlap between the areas of interest, and alternated with target images within the stimuli presentation. Participants viewed the stimuli in one of two orders which were the reverse of each other.

Procedure. Participants were told that they would see images depicting an action or event. Each participant was randomly assigned to one of four conditions with 10 participants in each condition). In the

Agent condition, participants were instructed to look as quickly as possible at "the person or animal who was performing the action," and then press the space bar after they had fixated on the target. In the Patient condition, participants were given the same instruction but told to look at "the object directly affected by the action." In the Goal condition, participants were instructed to look at "the goal or destination of the action," and in the Instrument condition, participants were instructed to look at "the tool or body part used to make the action." Each participant saw a practice image (an archer firing an arrow at a target with a bow) in which the target event component relevant to their condition (Agent, Patient, Goal or Instrument) was circled. Before each of the 36 pictures (18 target events, and 18 filler events), participants were instructed to fixate a cross located at the top of the screen in the center, and to press the space bar when they were fixating it. The image was presented immediately after the participants pressed the spacebar. After each image, participants viewed one of two images (randomly selected) depicting two static frogs for 3000 ms. The purpose of these pictures was to encourage participants to make eye-movements around the screen.

Participants' eye-movements were tracked using a Tobii T60 eye-tracker. Eye gaze was sampled at 60 Hz (every 16.67 ms). Participants were seated approximately 60 cm from the screen. At the start of the experiment, participants' eye-movements were calibrated using a five-point calibration procedure, in which they followed a red dot that moved to the four corners of the screen and then to the center of the screen. If calibration was incomplete, the procedure was repeated. Typically, participants required only one calibration. The experiment took approximately 10 min.

3.2. Results

Accuracy data. As a first step, we checked if participants followed the instructions and indeed looked at the target event components and then responded after they had fixated them. A response was classified as correct if the participant had looked at the target event component during the last 150 ms leading up to that response. If participants were incorrect in more than 45% of the trials, this would indicate that they had difficulty following the instructions. Data from 6 participants (3 in the Patient condition and 3 in the Instrument condition) who met this criterion were excluded from further analyses.

Next, we tested whether the accuracy with which participants located the target event role differed across conditions. Fig. 3 shows the mean accuracy of target identification across conditions. Data were analyzed with generalized binomial linear mixed effects modelling (*glmer*) with crossed random intercepts for Subjects and Items. The dependent measure consisted of binary values for accuracy (1 = accurate, 0 = not accurate) at the item level. The fixed effect of Condition was contrast coded using the forward difference coding. The analysis revealed that Agents were identified equally accurately with Patients ($\beta = 0.399$, $SE = 0.601$, $z = 0.664$, $p = .507$), which were identified equally accurately with Goals ($\beta = -1.046$, $SE = 0.636$, $z = -1.646$, $p = .099$). However, Goals were identified more accurately than Instruments ($\beta = 1.985$, $SE = 0.619$, $z = 3.209$, $p = .001$).[2] In the next set of analyses, we investigated the allocation of attention to target event components across time once the participants were able to identify them correctly.

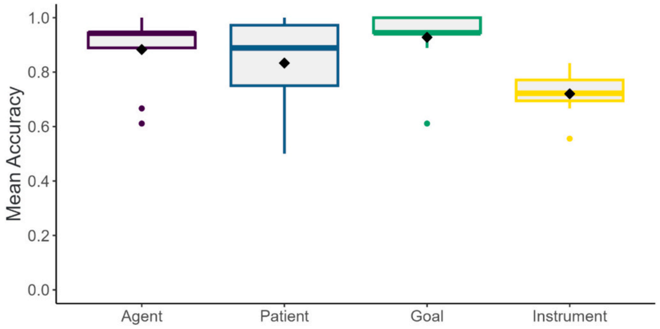


Fig. 3. Accuracy target event component identification across conditions (Experiment 2).
Note. Black diamonds represent the group mean. Horizontal colored bars represent the median, boxes represent the interquartile range (25th and 75th percentile), whiskers represent the range excluding the outliers, and colored dots represent outlier participants.

Eye-movement data. Next, we used the eye movement data to investigate whether participants identified target event roles at different speeds. We reasoned that if participants identified an event component more quickly, then they would spend more time looking at that component, hence the proportion of fixations would be higher. Thus, our dependent variable was proportion of fixations to the target event component out of all fixations.

In each image, four non-overlapping Areas of Interest (AOIs) were defined (Agent, Patient, Goal, Instrument) using the Tobii Studio AOI tool. In cases where the Agent was holding an Instrument, the Agent AOI was defined as the area of the Agent's torso and head, and the Instrument as the tool or instrument itself, as well as the hand and wrist of the Agent (if the Agent was performing a kicking action, the foot and ankle were considered the Instrument AOI). We computed whether a fixation fell into one of the AOIs in each successive sample from the onset until the end of the trial using the Tobii Studio software. Trials with greater than 30% trackloss were excluded from the analysis (2% of the data).

The time course of eye movements was analyzed only for the trials in which the participants had correctly identified the target event component based on the criteria described for the accuracy analysis. Eye movement data from incorrect trials were excluded from the analysis (15.5% of the data).

Analysis of eye movement data focused on a subset of overall time course. Specifically, we focused on the window spanning 150 ms to 1150 ms from trial onset. We excluded the eye movement data from the first 150 ms of the trial since it takes about 150 ms for participants to plan and land a saccade (Allopenna et al., 1998; Matin et al., 1993), and analyzed the next 1000 ms of eye movement data. Furthermore, because participants were instructed to look at the target and then press the space bar after they identified the target, the length of the trial varied across participants. Specifically, in 53% of the trials, participants identified the target before 1150 ms of eye gaze were collected. This led to a drop in the proportion of fixations to the target event component over time. Importantly, this drop was observed not because participants were fixating at the target less. Rather, it was because the data from completed trials ended up being missing as time proceeded since participants had responded and the trial ended. To resolve this problem, following Allopenna et al. (1998), we artificially extended out the correct trials that ended before 1150 ms so that all trials would have 1150 ms of eye gaze data. This step was completed after the eye movement data from incorrect trials were excluded, so all of the additional samples were added to the correct trials only. We reasoned that, if the participants were fixating at the target when they pressed the space bar, then presumably they would have continued to look at the target if they had

---

[2] The reaction times for the accurate button presses for identifying Agents (M = 993.53 SE = 70.05), Patients (M = 1173.56, SE = 64.64), Goals (M = 1264.31, SE = 45.27) and Instruments (M = 1067.33, SE = 47.38) were compared to each other with a linear-mixed effects model with random intercepts for Subjects and Items. The model revealed that the reaction times did not differ across any of the conditions (Agents vs. Patients: $\beta = -247.67$, $SE = 300.67$, $t = -0.823$, $p = .416$; Patients vs. Goals: $\beta = 38.40$, $SE = 300.58$, $t = 0.128$, $p = .899$; Goals vs. Instruments: $\beta = 167.84$, $SE = 300.80$, $t = 0.558$, $p = .581$). However, because participants were asked to look at the component as quickly as possible and then press the space bar after they have fixated – instead of pressing the space bar as quickly as possible – the reaction time data may not be a precise measure of how fast participants identified a target event component. The eye-movement data is more informative in this respect.

to wait for a fixed period of time for the trial to end (e.g., 1150 ms/69 samples). Thus, in the extended dataset, the additional samples were assigned a value of 1 for looks to the target event role (e.g., Agent AOI in the Agent condition, Patient AOI in the Patient condition, etc.) and a value of 0 for all other event roles (e.g., Patient AOI in the Agent condition, Goal AOI in the Agent Condition, etc.). Fig. 4 shows the proportion of fixations to the target event component across conditions for this extended dataset.

The time course data of eye movements were divided into two 500 ms windows. Previous work using stimuli similar to ours had used 1000 ms windows (Bunger et al., 2016, 2021). Because our stimuli remained visible for a shorter duration, we chose to halve these windows of analysis in order to increase sensitivity. We reasoned that, since event apprehension is quite rapid (Griffin & Bock, 2000; Hafri et al., 2013), eye movements in the earlier time window can more accurately reflect visual identification of target event roles. Thus, asymmetries in the speed of target role identification might be more prominent in the first 500 ms window compared to the second 500 ms window. We tested whether the speed at which participants identified the target event role differed across conditions and time windows with a linear mixed effects model (*lmer*). The dependent measure was elogit-tranformed mean proportion of target fixations at the item level. Subjects and Items were added as random intercepts, and Condition, Time and interaction between Condition and Time were added as fixed factors. The fixed effect of Condition was tested with planned contrasts using forward difference coding and the fixed effect of Time was tested with centered contrasts ($-1/2$ = First 500 ms, $1/2$ = Last 500 ms). Parameter estimates from the model are presented in Table 1. Parameter estimates for the planned contrasts testing the fixed effect of Condition in the First 500 ms and Second 500 ms are separately reported in Table 2.

The analyses revealed an interaction between the contrast for Patients vs. Goals and Time and a trend approaching significance for the interaction between the contrast for Agents vs. Patients and Time. However, the interaction between the contrast for Goals vs. Instruments and Time was not statistically significant. That is, in the First 500 ms time window, the overall proportion of time spent looking at the target region was greater when the target was the Agent compared to when the target was the Patient and when the target was the Patient compared to when it was the Goal. However, the overall proportion of time spent looking at the target did not differ when the target was the Goal as opposed to when it was the Instrument. Unlike in the First 500 ms window, in the Second 500 ms window, the overall proportion of the time spent looking at the target did not differ when the target was the Agent vs. the Patient and when it was the Patient vs. the Goal. However, similar to the First 500 ms window, in the Second 500 ms window there was no difference in the overall time spent looking at the target when the target was the Goal vs. the Instrument.

One might ask whether the relative speed with which participants identified a target event component could be affected by the size and the location of the AOIs. For example, proportion of looks to the target might be higher if the area taken up by that event component is larger. Similarly, participants might be faster in moving their eyes to an event component if it is located closer to the center of the screen – i.e., where the fixation cross was presented immediately before the picture. To rule out these possibilities, we conducted a set of exploratory analyses. As described in Experiment 1, size of each AOI using the Tobii Studio AOI tool. The overall sizes of AOIs in the 18 images used in this experiment also differed ($F(3,68) = 33.22$, $p < .001$). As in the larger set of 24 images, Goals ($M = 13.81$, SD = 6.39) were larger than Agents ($M = 5.17$, SD = 1.57), Patients ($M = 3.54$, SD = 1.61), and Instruments ($M = 3.63$, SD = 2.49; all $p < .001$ based on pairwise comparisons with corrections). We measured location of AOI as the distance between the center of the screen and the X and Y coordinates of the center of mass of the AOI. We compared the locations of Agents (M = 457.40 SE = 4.33), Patients (M = 188.53, SE = 4.79), Goals (M = 506.01, SE = 8.17) and Instruments (M = 221.73, SE = 6.20). A one-way ANOVA revealed a main effect of Component on the Distance from Center ($F(3,68) = 39.40$, $p < .001$). Pairwise comparisons with corrections for multiple comparisons revealed that both Patients and Instruments were closer to the center of the screen than both Agents and Goals (all $p < .001$). No other differences were significant.

Next, we tested a linear mixed effects model on the elogit-tranformed mean proportion of target fixations at the item level with crossed random intercepts for Subjects and Items. The model included Size and Location of the AOI (for that specific item) as fixed factors. The model

Table 1
Parameter estimates of the fixed effects of condition and time for the proportion of target fixations.

	β	*SE*	*t*	*p* value
Intercept	0.508	0.047	10.838	< 0.001
Agent vs. Patient	0.274	0.108	2.540	0.016
Patient vs. Goal	0.121	0.107	1.127	0.268
Goal vs. Instrument	0.134	0.108	1.241	0.223
Time (First 500 ms vs. Last 500 ms)	0.732	0.031	23.443	< 0.001
Agent vs. Patient x Time	−0.17	0.087	−1.959	0.050
Patient vs. Goal x Time	−0.230	0.086	−2.687	0.007
Goal vs. Instrument x Time	−0.096	0.088	−1.071	0.285

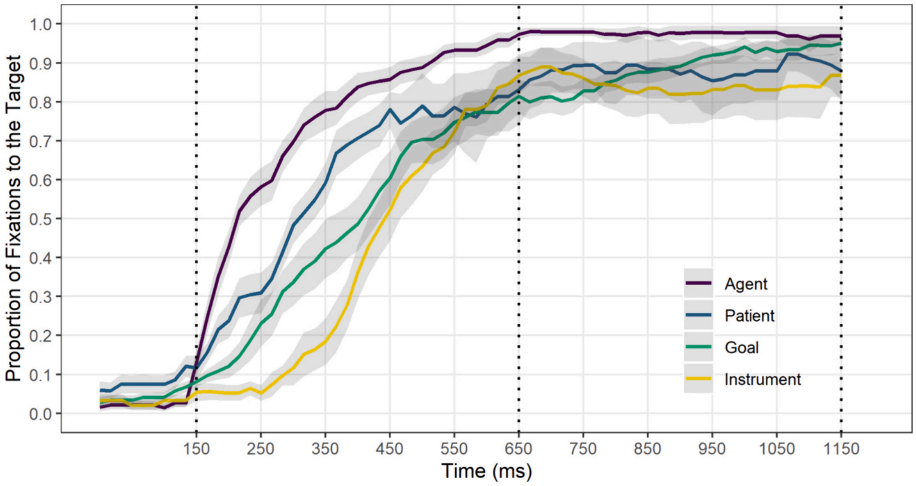


Fig. 4. Looks to the target event component across conditions (Experiment 2, correct trials only).
Note. Shaded areas indicate standard error of participant means.

Table 2

Parameter estimates of the fixed effects of condition on for the proportion of target fixations in the first and second 500 ms windows.

	β	*SE*	*t*	p value
First 500 ms window				
Agent vs. Patient	0.359	0.116	3.088	0.004
Patient vs. Goal	0.236	0.116	2.042	0.047
Goal vs. Instrument	0.182	0.117	1.557	0.126
Second 500 ms window				
Agent vs. Patient	0.189	0.116	1.624	0.118
Patient vs. Goal	0.006	0.116	0.052	0.959
Goal vs. Instrument	0.086	0.117	0.736	0.465

revealed that both size and location predicted proportion of fixations (Size: $\beta = -0.022$, $SE = 0.008$, $t = -2.880$, $p = .004$; Location: $\beta = 0.007$, $SE = 0.0003$, $t = 2.860$, $p = .005$). We compared this model to a model that also included fixed effects of Condition and Time and the interaction between Condition and Time. The second model fit the data better (χ^2 (7) = 484.15, $p < .001$). However, in this new model neither the Size ($\beta = -0.0097$, $SE = 0.007$, $t = -1.350$, $p = .178$) nor the Location ($\beta = 0.0004$, $SE = 0.0002$, $t = 1.384$, $p = .167$) of the AOI predicted looks to the target event component. However, all of the originally reported differences from fixed effect of Condition (except for the difference between Patients and Goals in the first 500 ms) remained statistically significant. Thus, although size and location of the AOIs predicted overall time spent looking at the target, these effects became non-significant once we took into account which event component was the target. Therefore, the size and the location of the event components in the images is unlikely to have affected the relative speed with which participants have identified different event components.

As a final check, we inspected the eye-gaze patterns to each event component both when that event component is the target and in non-target event conditions. Fig. 5 displays the proportion of looks to Agents, Patients, Goals and Instruments across the four conditions. We reasoned that looks to an entity would differ when that event component is the target as opposed to non-target conditions. Furthermore, if the relative salience of event roles differs and in ways that conform to the thematic hierarchy, then roles that are ranked higher in the hierarchy could interfere with the looks to the roles that are ranked lower in the hierarchy even when they are not the target.

In the Agent condition, the looks to the target (Agent) increased quite rapidly and there were no visible increases in the looks to the other event components (Patient, Goal or Instrument). This shows that Agents can be identified quite rapidly and without any interference from other roles. In the Patient condition, there was a rapid increase in the proportion of looks to the target (Patient). However, there was also an increase in the looks to the Agent (between 150 and 300 ms). There were no visible increases in the looks to the Goal or the Instrument. This suggests that identification of Patients in visual search could be susceptible to interference from Agents but not from Goals or Instruments. In the Goal condition, the increase in the proportion of fixations to the target (Goal) was accompanied by a slight increase in the looks to the Agent and Patient, but not the Instrument. Thus, identifying Goals might be susceptible to interference from both Patients and Goals but not Instruments. Finally, in the Instrument condition, the increase in the looks to the target (Instrument) was less rapid than it was in the other conditions. Furthermore, there was an early increase in the looks to the Agent and the Patient, but not the Goal. This suggests that identifying Instruments might be susceptible to interference from Agents and, especially Patients. However, Goals did not seem to interfere with identifying Instruments—a pattern deviating from the asymmetries predicted by the Thematic Hierarchy. The conclusions drawn based on the inspection of the looks to the components across target and non-target conditions is consistent with the findings from the main analyses based on target looks across conditions.

3.3. Discussion

Experiment 2 sought to investigate the processing of event components in a visual search task. Consistent with Experiment 1, we discovered asymmetries between event components: not all event components were identified with equal accuracy or speed. Finding an Agent in a

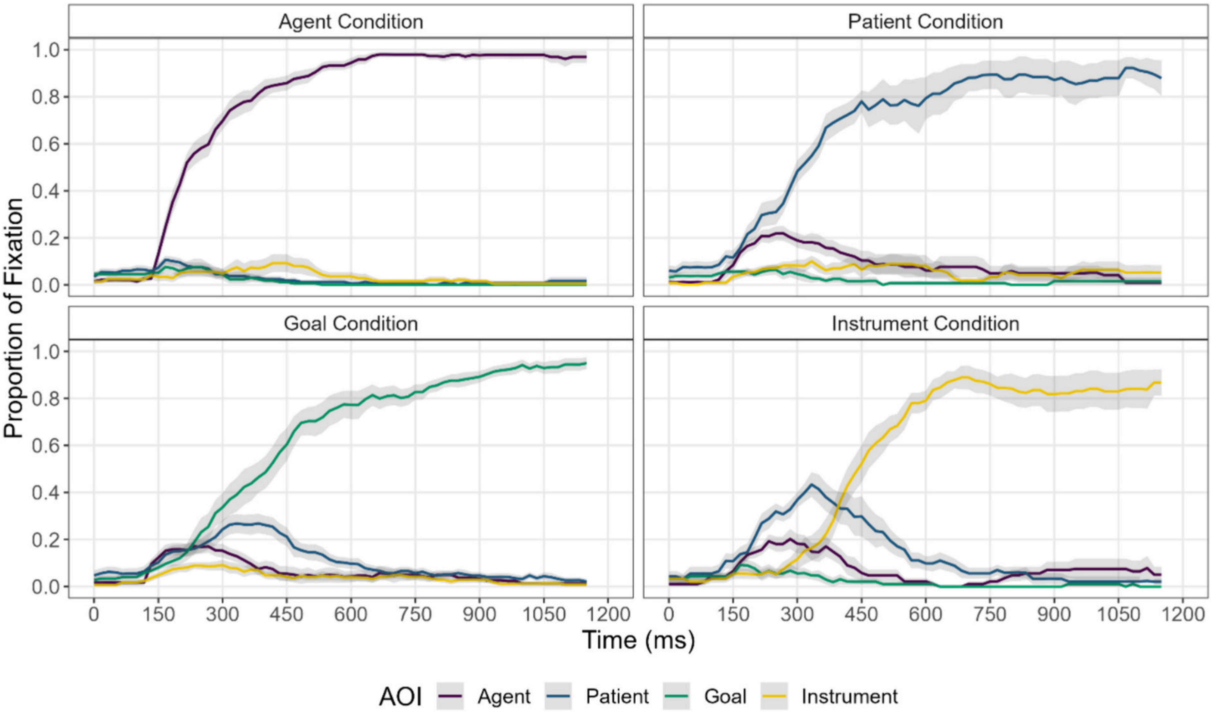


Fig. 5. Looks to each event component across Agent, Patient, Goal and Instrument conditions (Experiment 2, correct trials only).
Note. Shaded areas indicate standard error of participant means.

causative event occurred more quickly than finding a Patient which occurred more quickly than finding a Goal. Although Goals were numerically identified faster than Instruments, this comparison did not reach significance (we speculate on possible explanations below). In the accuracy data, none of our planned comparisons reached significance except that Instruments were identified less accurately than Goals. We speculate that accuracy, being a binary measure, may not be as sensitive as the continuous eye movement measure. Nevertheless, this measure was sensitive to Instrument identification being particularly inaccurate. In light of Experiment 1, where Instruments were mentioned less frequently than other roles, this might be because Instruments are considerably less salient than other roles. Further, the low accuracy of Instrument identification may also explain why the difference between Goals and Instruments did not reach significance in the eye movement measure. Since the eye movement measure is based on correct trials only, data from incorrect trials—those during which the Instrument was not fixated at the end of the trial—are by definition removed from the analysis. Given the high proportion of incorrect trials in the Instrument condition, the eye movement data includes substantially fewer data points than in the other conditions, possibly reducing the sensitivity needed to capture the differences in the speed of Goal vs. Instrument identification.

Another aspect of our findings concerns the fact that Agents were the only animate event component. It seems unlikely that the asymmetry between Agents and other event components can be merely explained by animacy given recent evidence from eye-tracking studies that the preference to fixate on Agents over Patients persists when the Patient is both animate and inanimate (Isasi-Isasmendi et al., 2023). Nevertheless, since we cannot definitely differentiate between animacy and agency in our experiments, this component is henceforth referred to as (animate) Agents.

Although Experiment 2 revealed asymmetries in the relative speed with which different event components were identified in a visual search task, two aspects of this task suggests that findings from this task need to be interpreted with caution. First, at the beginning of the task, participants were given a verbal description of the target event component (e.g., *the object affected by the action* for Patient). This might have encouraged them to (implicitly) use linguistic encoding while searching for the event components. It is possible that the asymmetries in how fast different event components were identified and their similarity to the patterns in the linguistic description task might be a reflection of linguistic encoding, at least to some extent. Second, target event component was manipulated between-subjects. This was done to eliminate the need to provide different verbal instructions to participants on a trial-by-trial basis. However, because each participant searched for only one event component, they might have become experts on identifying that component. A non-linguistic task that uses a within-subjects design would provide a more stringent test of the relative salience of event components in cognition.

4. Experiment 3: Change blindness

Experiment 3 was designed to extend findings from Experiment 2 on the relative cognitive salience of different event components (Patients, Goals, Instruments, and secondarily, animate Agents) while minimizing the explicit (including linguistic) demands of task and eliminating the limitations described above. To do so, we exploited the well-documented change blindness phenomenon (e.g., Aginsky & Tarr, 2000; Hollingworth & Henderson, 2000; Rensink et al., 1997). Change blindness is known as the difficulty, or at times even the inability, to see changes to a scene. Rensink et al. (1997) used a flicker paradigm, in which participants were asked to detect changes to a display which was flickering between an image and an altered version of that image (e.g., a version where the color or location of an item were changed). They found that changes to items of central interest in a scene were detected more quickly than changes to items of marginal interest. Rensink et al. (1997) proposed that the ability to detect a change to an item is linked to the attention allocated to that item, and that, by comparing the relative speeds at which changes to individual items are detected, it is possible to determine the degree to which attention is allocated to different components in a scene. For present purposes, change blindness offers a particularly straightforward way of measuring relative prominence of event participants: event components to which changes are detected faster would be allocated more attention or be more salient. Unlike Experiment 2 that asked participants to follow verbal instructions to identify a specific event component, this task simply required participants to identify the object that changed color without guiding them towards a specific event component. Furthermore, this task used a within-subjects design to manipulate which event component changed on a trial by trial basis. Hence, this task could be considered to be more implicit and giving more insight into the relative salience of event components in cognition. Furthermore, to ensure that the speed with which participants detected changes to event components was determined by their specific role in the event and not the physical salience of the entities filling these roles, we included controls for this task in Experiment 4 (see also Supplementary Material for additional controls).

4.1. Method

Participants. Twenty undergraduate students at the University of Delaware, all native speakers of English, participated for class credit (13 females, Mean age = 19;4, SD = 7.5 months). None of them had participated in Experiment 1 or 2. Four additional participants were excluded from the analysis because they inaccurately identified the changed object in more than four trials.

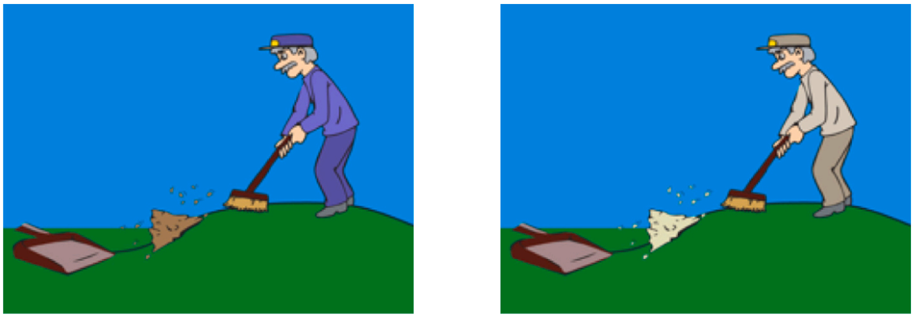
Materials. The same 24 images that served as the test stimuli of Experiment 1 were used. For each of the original images, a set of four variants was created, in which the color of either the Agent, Patient, Goal or Instrument was changed (Fig. 6). For Agents, color changes affected the Agent's clothes.

Procedure. The design was based on Rensink et al.'s (1997) flicker paradigm. Participants viewed flicker sequences composed of an original image (e.g., Fig. 1) followed by a variant (Fig. 6), with a mask (i.e., a blank grey screen) between the images. Participants were told that they would see an image which would flicker and change, and that the change would be something in the scene changing color. Participants were told that they should hit the space bar as soon as they were able to identify the change. They would then be asked to name the object which was changing. Participants saw two practice items with events that did not involve caused motion, after which understanding of the task was checked. We measured whether participants were able to identify the change accurately, as well as the reaction time needed to detect the change.

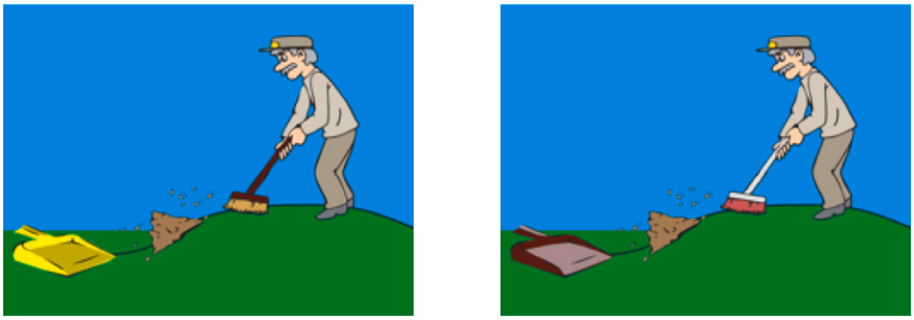
The stimuli were presented using *E*-Prime software on a Dell laptop. Participants were seated in front of the laptop, about 60 cm from the display. Following Rensink et al. (1997), images were displayed for 240 ms and the grey screen for 80 ms. Thus, one complete cycle (original image-grey screen-variant-grey screen) had a duration of 640 ms. Each participant saw 6 relevant variants for each of the 4 event components (Agent, Patient, Goal, Instrument) for a total of 24 original event-variant pairs. Variants for each event were counterbalanced across participants, so each participant saw only one type of variant for each event. The stimuli were presented in a different randomized order for each participant using the randomization feature in *E*-Prime. The experiment took about 10 min to complete.

4.2. Results

Trials where the participant failed to identify the changed object correctly were excluded from the analysis (3.95% of all trials). Reaction times which were more than 3 SD from the mean were also excluded from the analysis (1.52% of the accurate trials). There were no reaction



Agent Change Patient Change



Goal Change Instrument Change

Fig. 6. Examples of change items (Experiment 3).

times of less than 1 cycle (i.e., 640 ms). Fig. 7 shows the mean reaction time for each event component. Data were analyzed with linear mixed effects modelling (*lmer*) with crossed random intercepts for Subjects and Items. The dependent measure was Reaction Times at the item level. The fixed effect of Condition was contrast coded using forward difference coding. The model revealed that participants were faster in identifying changes to Patients than changes to Goals ($\beta = -328.36$, $SE = 106.17$, $t = -3.093$, $p = .002$) and changes to Goals than changes to Instruments ($\beta = -348.23$, $SE = 108.19$, $t = -3.219$, $p = .002$). However, somewhat surprisingly, they were also slower in identifying changes to Agents than changes to Patients ($\beta = 376.54$, $SE = 106.64$, $t = 3.531$, $p = .004$).

Finally, as in previous experiments, we checked whether the asymmetries in the speed at which changes to individual event roles were detected were not simply a byproduct of the physical properties of the stimuli such as the size of event components. It is possible that the changes that were detected faster could be to the event components that were larger. To rule out this possibility, we compared a linear mixed effects model on Reaction Times that included Size of the event component (for that specific item) as a fixed factor to a model that included fixed effects of both Size and Component. The second model fit the reaction time data better compared to the model that included only Size as a fixed effect ($\chi^2(3) = 31.194$, $p < .001$). The model revealed that Size of the component did not significantly predict reaction times ($\beta = -0{,}398$, $SE = 8.923$, $t = -0.045$, $p = .964$). Furthermore, all of the previously reported differences in reaction times in detection of changes to the event components remained the same. Thus, the asymmetries in the relative speed with which changes to different event components were detected was independent of the physical attributes of the event components.

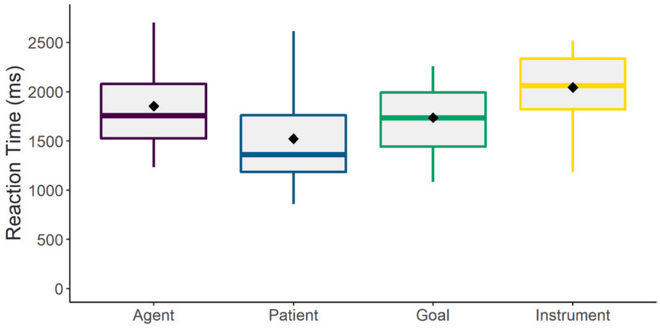


Fig. 7. Mean reaction times in the change blindness task (Experiment 3).
Note. Black diamonds represent the group mean. Horizontal colored bars represent the median, boxes represent the interquartile range (25th and 75th percentile), whiskers represent the range excluding the outliers, and colored dots represent outlier participants.

4.3. Discussion

Experiment 3 aimed to determine the relative salience of (animate) Agents, Patients, Goals and Instruments in causative events by measuring the time taken to detect changes to each of these components. Changes to individual event components were detected at different speeds. Specifically, changes to Patients were identified more quickly compared to changes to Goals; furthermore, changes to Goals were identified more quickly compared to changes to Instruments. These findings parallel the linguistic prominence patterns observed in Experiment 1, which ranked Patients above Goals and Goals above Instruments.

Somewhat unexpectedly, changes to (animate) Agents were identified more slowly than changes to Patients. The Agent findings contrast with the description data in Experiment 1 which showed a large preference to encode Agents linguistically as well as the eye-movement data in Experiment 2 which revealed speedy identification of Agents in visual search. We believe that the Agent findings in the change blindness task were due to the fact that, unlike the other three (inanimate) components, the color change in Agents applied to an attribute of the participants (the Agents' clothes), not the participants themselves. Because of this asymmetry, the results of this task might have limited potential to bear on the conceptual prominence of Agents as event participants.

Although we ruled out size as an alternative, we need to address two additional explanations for the patterns in the data before the theoretical significance of the findings from Experiment 3 can be assessed further. First, some of the entities in the events may have been physically more salient relative to other entities regardless of their specific involvement in the event. For instance, in the event in Fig. 1, the dirt may be more prominent than the broom regardless of the role these objects play in the event (Patient vs. Instrument respectively). Second, some of the color changes performed on individual objects may have been more salient than others regardless of the contribution of each object to the structure of the event. To further ensure that the observed asymmetries between event entities were due to the role these entities played in the depicted events rather than the salience of individual entities or the changes to these entities, an additional study was performed (Experiment 4) where the entities in Experiment 3 were dissociated from the events they appeared in.

5. Experiment 4: Removing event structure

Experiment 4 modified the materials from Experiment 3 such that each to-be-changed component was isolated from the event, but retained the same size, position and background as in the original event. For instance, the stimulus in Fig. 1 was replaced by a test item depicting only a single object (e.g., the Goal - the dustpan) in the same position and background as before, and a changed version where that object changed color. If changes to isolated entities reveal asymmetries similar to those in Experiment 3, the conclusion that event participants in causative events enjoy differential salience would be challenged. Alternatively, if the asymmetries observed in Experiment 3 disappear in these new tasks or if new asymmetries emerge, the conclusion that participants in causative events have different degrees of salience would remain valid.

Experiment 4 included two different tasks. In the Change Blindness Single-Object task, we used a method similar to that in Experiment 3, except that participants were asked to report the nature of the color change (e.g., red to green) after they had detected the change. In the Magnitude Estimation Single-Object task, we asked another group of participants to judge the magnitude of the visual difference between the stimuli (original vs. variant) using a Magnitude Estimation rating procedure (see Sorace, 2010, among others). The data from both tasks are compared to the data from the original Change Blindness experiment with the complete events.

5.1. Method

Participants. Participants were drawn from the same population as Experiments 1–3 but had not participated in the earlier studies. Twenty participants (11 females, Mean age 19;7, SD = 8 months) participated in the Change Blindness Single-Object task and another group of twenty-two participants (9 females, Mean age 19;6, SD = 10 months) participated in the Magnitude Estimation Single-Object task. All participants were native speakers of English.

Materials. The stimuli used in Experiment 3 were adapted so that each event component was isolated, while retaining the same position, size and general background from the original image. For instance, to create a pair of trials that corresponded to the Goal variant in the event in Fig. 1, the dustpan (Goal) was isolated from the original event and stood alone against the same background and in the same position as before, such that the color change in the variant of the scene affected this single object. See Fig. 8 for an Example. For ease of reference, we continue referring to these entities as Agents, Patients, Goals and Instruments in this experiment, even though they no longer participate in an event.

Procedure. In the Change Blindness Single-Object task, the procedure was as in Experiment 3 except that participants were asked to report the nature of the color change (e.g., brown to yellow) rather than the object that was changing after they had detected the change.

In the Magnitude Estimation Single-Object task, the stimuli were presented sequentially for 500 ms each using *E*-Prime. For each pair of stimuli, participants were asked to rate the magnitude of the change using a Magnitude Estimation judgment (Sorace, 2010). In a practice phase, participants were trained to make proportional judgments as required by Magnitude Estimation by estimating the relative length of lines. The practice phase had two purposes. First, it provided training in making proportional judgments. Second, it allowed participant judgments to be compared to an objectively measurable quantity so that it was possible to determine whether participants had understood the task and exclude those participants who had not. Data from an additional 4 participants whose judgments of line lengths had a correlation less than $r = 0.75$ with the actual line length were excluded from the analysis because it was not clear that they fully understood the nature of the proportional judgments.

5.2. Results

Change Blindness Single-Object Task. Data were analyzed with linear mixed effects modelling with crossed random intercepts for Subject and Item. Dependent variable was Reaction Times at the item level. Fixed effect of Condition was contrast coding using forward difference coding. The mean reaction times for change detection across conditions is presented in Fig. 9A. Participants were equally fast in detecting changes to Agents vs. Patients ($\beta = 112.36$, $SE = 97.17$, $t = 1.156$, $p = .248$), Patients vs. Goals ($\beta = 77.13$, $SE = 97.17$, $t = 0.794$, $p = .428$), and Goals vs. Instruments ($\beta = -84.04$, $SE = 97.17$, $t = -0.865$, $p = .388$).

Since the stimuli in this task were simpler than in the original task, it is possible that performance on this task might be at ceiling. To address this possibility, we compared the change detection data from Experiment 3 to the data from the Single-Object task. We used a mixed effects model on Reaction times to test the fixed effects of Experiment and Component. Overall, participants were 109.571 ms faster in the Single-Object task compared to the original task ($\beta = -154.9$, $SE = 61.9$, $t = -2.502$, $p = .013$). However, the difference in reaction times between experiments was specific to Goals ($\beta = 232.7$, $SE = 115.7$, $t = 2.011$, $p = .045$) and Instruments ($\beta = 481.5$, $SE = 117.2$, $t = 4.108$, $p < .001$). Participants had similar reaction times when detecting changes to Agents ($\beta = 80.6$, $SE = 116.2$, $t = 0.693$, $p = .488$) and Patients ($\beta = -175.3$, $SE = 115.1$, $t = -1.522$, $p = .128$) across the Single-Object task and the original task. Thus, it is unlikely that the performance on the Single-Object task was at ceiling since (at least for Agents and Patients) participants took as much time when detecting changes to isolated objects as they did when the same objects were integrated in a way to form a coherent event.

Magnitude Estimation Single-Object Task. Participants' ratings were standardized by taking the natural logs of ratings, and then calculating the z scores. Data were analyzed with linear mixed effects modelling with crossed random intercepts for Subject and Item. Dependent variable was z scores of logs of ratings at the item level. Fixed effect of Component was contrast coding using forward difference coding. The mean ratings for the magnitude of the changes across conditions is presented in Fig. 9B. Participants indicated that the magnitude of

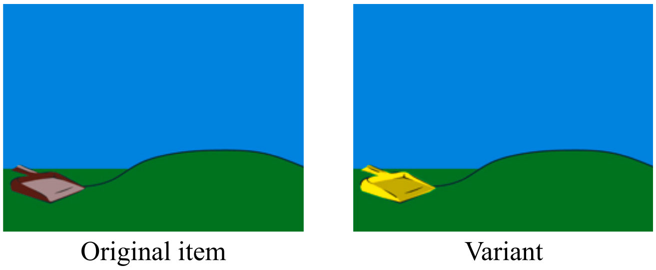


Fig. 8. Examples of original items and variants used in single-object tasks (Experiment 4).

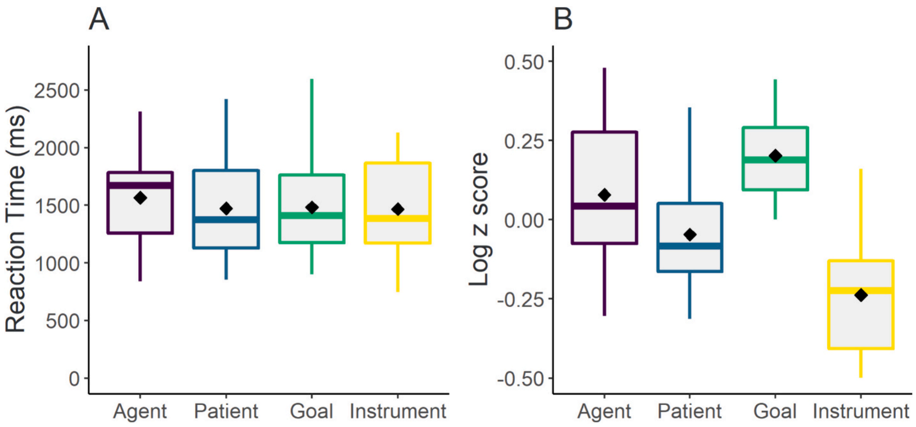


Fig. 9. Data from (A) change blindness single-object and (b) magnitude estimation single-object tasks (Experiment 4).
Note. Black diamonds represent the group mean. Horizontal colored bars represent the median, boxes represent the interquartile range (25th and 75th percentile), whiskers represent the range excluding the outliers, and colored dots represent outlier participants.

changes to Patients was smaller than the magnitude of the changes to Goals (0.361, $SE = 0.152$, $t = 2.372$, $p = .021$). They also indicated that the magnitude of changes to Goals was bigger than the magnitude of changes to Instruments ($\beta = 0.516$, $SE = 0.152$, $t = 3.392$, $p = .001$). However, there were no differences between the magnitude of the changes to Agents vs. Patients ($\beta = 0.168$, $SE = 0.152$, $t = 1.107$, $p = .273$).

Finally, we conducted an exploratory analysis on the data from the change blindness task in Experiment 3 by including the magnitude estimation ratings from Experiment 4 as a control variable. We compared a model that only had the fixed effect of Magnitude Estimation Ratings to a model that included both Magnitude Estimation Ratings and Component. The second model had a better fit for the Reaction Time data (χ^2 (3) $= 38.309$, $p < .001$). In addition to the previously reported asymmetries in Experiment 3, there was a fixed effect of magnitude of the change ($\beta = -442.04$, $SE = 100.57$, $t = -4.395$, $p < .001$). Therefore, across all items, changes that were rated to be bigger were also detected faster. However, this effect did not account for the asymmetries in reaction time data in the original task. Thus, differences in the magnitude of the change are unlikely to explain the asymmetries in the data from the change blindness task.

5.3. Discussion

Recall that Experiment 3 found that changes to Patients were detected more quickly than changes to Goals, which were in turn detected more quickly than changes to Instruments. The results from both Single-Object studies in Experiment 4 suggest that these asymmetries are unlikely to be due to physical features related to individual entities or individual color changes, since these asymmetries mostly disappear when event entities are isolated from the event. There was only one case where the Single-Object studies did reveal an asymmetry in the salience of event components: the magnitude of the change to Goals was estimated to be greater than that for both Patients and Instruments. Note that the former indicates an asymmetrical pattern that is in the opposite direction than in Experiment 3. Nevertheless, changes to Agents and Patients that were detected at different speeds in Experiment 3 did not differ in salience in Experiment 4. Furthermore, the asymmetries in Experiment 3 remained the same after controlling for the estimated magnitude of the change. This shows that visual salience defined over physical features of entities does not necessarily correspond to conceptual salience defined over relations between entities in an event. On the basis of the Single-Object studies, we can conclude that the differences observed between event entities in Experiment 3 truly emerged as a consequence of the way these entities participated in relatively more complex causative events. (See Supplementary Material for converging evidence from additional control tasks with inverted or scrambled event images that also failed to replicate the asymmetries with the original stimuli.)

6. General discussion

How does language encode events? The purpose of the present study was to explore the correspondence between linguistic and conceptual event structure by comparing the relative salience of event roles across linguistic descriptions and non-linguistic measures of event cognition for the same events. We tested the possibility that asymmetries in the prominence of thematic roles reflect asymmetries in the conceptual prominence of event components (Baker, 1997; Dowty, 1991; Jackendoff, 1990; Levin & Pinker, 1991). Unlike previous linguistically inspired work on event cognition that has focused on simple, two-participant events (Griffin & Bock, 2000; Hafri et al., 2013, 2018; Rissman & Lupyan, 2022; Sauppe & Flecken, 2021), we examined multi-component caused motion events that are relatively more complex. Specifically, we examined the relative salience of four event components, (animate) Agents, Patients, Goals and Instruments, within a single event as a way of determining their contribution to the internal structure of events in both language and cognition. We began by exploring the relative salience of event components in language by focusing on the frequency with which speakers mentioned different event components in a description task (Experiment 1). We then explored the relative salience of event components in cognition by examining the relative accuracy and speed at which they identify these components in a visual search task (Experiment 2) and the relative speed at which speakers detected changes to the same event components (Experiments 3 and 4). In all experiments we sought to explore whether the relative salience of event components in language and cognition would conform to the Thematic Hierarchy proposed in linguistic theory (Baker, 1997; Jackendoff, 1990).

6.1. Event role prominence in language and cognition

In language production (Experiment 1), there were asymmetries among components of caused motion events. As expected, (animate) Agents were mentioned more frequently than Patients in causative event descriptions. Patients were followed by Goals which were mentioned more frequently than Instruments. These asymmetries in the relative salience of event roles in language are consistent with the predictions of the Thematic Hierarchy (Baker, 1997; Jackendoff, 1990). They also provided a basis for investigating the proposal that these asymmetries are grounded in the conceptual saliency of event components.

Our non-linguistic measures of event cognition (Experiments 2 and

3) revealed that the event components filling these roles in language can be identified rapidly and accurately. In the visual search task, participants identified Agents, Patients, Goals and Instruments after viewing the event for almost less than a second. In the change blindness task, changes to the same event components were correctly identified after viewing the original item and the variant less than four times (i.e., within two seconds). These findings cohere with prior work on the rapid extraction of event roles relevant for language in simple events with two-participants (Dobel et al., 2007; Griffin & Bock, 2000; Hafri et al., 2013, 2018; Webb et al., 2010). Our findings also connect to recent evidence showing that the distinction between verbs denoting symmetrical vs. asymmetrical event role relations (e.g., kiss vs. punch) is expressed with visually motivated forms in sign languages (Gleitman et al., 2019) and hearing individuals with no knowledge of sign languages also have access to this mapping (Hafri et al., 2022). Furthermore, we extend prior work on two-participant events that have focused on Agents and Patients to caused motion events that involve additional participants, namely Goals and Instruments. The fact that the event roles relevant for language can be rapidly identified in our experiments supports the idea that the structure of events in language is strikingly similar to conceptual event structure in cognition.

Crucially, these non-linguistic measures of event cognition also revealed that not all event components were identified equally rapidly. There were asymmetries in the speed or accuracy with which target event components were identified in visual search (Experiment 2). Furthermore, changes to event components were detected at different speeds (Experiment 3). These findings suggest that event components differentially contribute to structured event representations in cognition. Patients tended to be more salient than Goals and Goals tended to be more salient than Instruments. While (animate) Agents tended to be more salient than Patients in visual search, this asymmetry was not replicated in change detection. These asymmetries in the conceptual saliency of event components are consistent with predictions of the Thematic Hierarchy (Baker, 1997; Jackendoff, 1990) and are similar to the role prominence asymmetries for the same events in language production. The similarities across linguistic and non-linguistic measures provide another piece of evidence for the idea that the thematic roles in language build on and reflect the conceptual organization of event structure. These homologies are reminiscent of similar homologies in the literature, such as the Source-Goal asymmetry that characterizes both language production and non-linguistic event cognition (Lakusta et al., 2007; Lakusta & Landau, 2005, 2012; Papafragou, 2010; Regier & Zheng, 2007). Our findings cohere with and extend these asymmetries demonstrated among Goals and Sources to (animate) Agents, Patients and Instruments.

Nevertheless, the homologies between linguistic and non-linguistic measures of event role salience were not perfect. As mentioned earlier, there were deviations from the general pattern of asymmetries observed across the three tasks. Although the asymmetries in language production aligned with the Thematic Hierarchy, participants in the visual search task were not faster at identifying Goals as compared to Instruments. These same participants were however less accurate at identifying Instruments than Goals, which is in line with the Thematic Hierarchy predictions. It is possible that the binary vs. continuous nature of these two measures as well as the fact that a relatively high number of (inaccurate) trials are excluded from eye movement measure in the Instrument condition might have limited the sensitivity needed to tease apart the relative salience of these two roles. The change blindness task appears to be more sensitive from this respect. However, one caveat of the change blindness task was that the color changes in the Agent condition applied only to a property of the component instead of the whole component.[3] This might be a possible explanation for why changes to Agents were not detected the fastest. In fact, this pattern is consistent with prior change blindness work in the visual cognition literature showing that central changes are detected more easily compared to marginal changes (Levin & Simons, 1997; Rensink et al., 1997; Simons, 2000). Nevertheless, the fact that this pattern did not emerge in the control experiments strongly suggests that this factor is unlikely to be the only explanation for this pattern.

Crucially, however, detection of the changes to Patients, Goals and Instruments in Experiment 3 conformed to the asymmetries predicted by the Thematic Hierarchy. This pattern of asymmetries was unlikely to be explained by low level features such as the physical size of the component (Experiment 3) or the perceived magnitude of the change (Experiment 4, Magnitude Estimation Single Object task). Furthermore, the asymmetries in the speed of change detection disappeared once event structure was removed from the stimuli by presenting the entities filling in these event roles individually (Experiment 4, Change Blindness Single Object Task).

Alternatively, these minor misalignments between the linguistic and cognitive salience of event roles might be explained by the fact that language production and event cognition are characterized by different pressures. For instance, in language production, speakers need to take into account language-specific constraints on argument realization (Levelt, 1989) as well as pragmatic factors (Papafragou & Grigoroglou, 2019). Furthermore, the cognitive salience of event roles is susceptible to the influence of additional conceptual factors. For instance, the Goal-Source asymmetry in memory is more likely to emerge for intentional events performed by animate agents (e.g., a man hopping from a table to a ladder) as opposed to physical events involving inanimate objects (e.g., a paper blowing from a container to a candle), even though this asymmetry emerges in language production for both kinds of events (Lakusta & Landau, 2012).

Finally, earlier we raised the possibility that the similarity between the patterns in the linguistic description task and the visual search task being a reflection of linguistic encoding of the scenes. However, the overall pattern of asymmetries was largely similar across the change blindness task that had minimal linguistic demands and the visual search task that was relatively more explicit (and had higher linguistic demands). Therefore, this possibility is unlikely when the findings of the two non-linguistic tasks are considered together.

Together, our findings suggest that event role prominence asymmetries are largely similar across language and cognition, although not identical. Thus our findings suggest that event roles in language are grounded in non-linguistic representation of the entities filling those roles, at least to a certain extent. However, the correspondence between linguistic and non-linguistic representation of event roles is more complex and less straightforward than as previously suggested in linguistic theory (Baker, 1997; Dowty, 1991; Jackendoff, 1990; Levin & Pinker, 1991).

6.2. *Implications for the relation between event language and cognition*

The present findings cohere with other evidence that systematic patterns of linguistic encoding reveal core aspects of event representation. To take one example, linguistic telicity systems encode the distinction between events that have inherent endpoints (*draw a circle*) vs. events that do not (*draw circles*; Filip, 2012; van Hout, 2016; Wilbur, 2003). A homologous distinction between temporally bounded vs. unbounded events has recently been shown to characterize non-linguistic cognition (Ji et al., 2022; Ji & Papafragou, 2020a, 2020b, 2022; Kuhn et al., 2021; Strickland et al., 2015).

Could the homology in the present data indicate a relation between language and cognition that is in the opposite direction from the one we have considered? Namely, could it be that the salience of event roles in English has shaped their conceptual salience such that those roles that are ranked higher in the thematic hierarchy are also more salient in our conceptual tasks? We know that speakers apprehend events in a way

---

[3] Future work can use other agents, e.g., aliens or robots whose color change completely to address this limitation.

that is consistent with language-specific features when planning linguistic descriptions (e.g., Gleitman et al., 2007; Griffin & Bock, 2000; Konopka & Meyer, 2014; Meyer et al., 1998; Papafragou et al., 2008; Sauppe, 2017; Sauppe et al., 2013; Ünal et al., 2022; Ünal & Papafragou, 2016; van de Velde et al., 2014). For example, speakers of languages with Subject-Verb-Object (SOV) word order (e.g., English) tend to fixate on Agents during early stages of language planning (Gleitman et al., 2007; Griffin & Bock, 2000), but this preference is reduced in languages with Verb-Object-Subject (VOS) word order and languages that have a flexible word order (Norcliffe et al., 2015; Nordlinger et al., 2022) or based on recent linguistic experience (Sauppe & Flecken, 2021). Could the way a language encodes event roles influence how salient the event participants filling these roles are in tasks such as ours that do not require people to produce linguistic descriptions? Cross-linguistic evidence is needed to evaluate this possibility. In a recent demonstration, learners of English were compared to learners of Turkish—a pro-drop language that has a relatively flexible word order—on linguistic and non-linguistic tasks similar to the ones in the present study (Ünal, Richards, et al., 2021). This work shows that, although Agents were linguistically less prominent in Turkish than in English, these cross-linguistic differences did not surface in a non-linguistic change detection task: changes to Agents were detected at similar accuracy levels across Turkish- and English-speaking children. Further work with adults, especially with speakers of VOS languages or languages that have a flexible word order, is necessary to complete the cross-linguistic picture. Such work would also need to consider the contributions of pragmatic and other factors alongside the cognitive salience of event roles to language production (recall from Experiment 1 that the pragmatic factor of typicality affected the probability of mention of all event components regardless their salience).

The present findings raise a question about the developmental continuity of the homologies between language and event cognition. Several studies with pre-linguistic infants inspired by the linguistic analysis of events demonstrated evidence for developmental origins of these homologies (Göksun et al., 2010; Wagner & Lakusta, 2009). Recent developmental work also shows that the asymmetries in the linguistic and cognitive salience of event roles observed with English-speaking preschoolers are largely similar to the patterns observed with English-speaking adults in the current study (Ünal, Richards, et al., 2021). Further work with learners of different languages from a wider age range is necessary to explore whether and to what extent these homologies change during language acquisition.

Similar to previous work on event roles (e.g., Griffin & Bock, 2000; Hafri et al., 2013, 2018; Rissman & Lupyan, 2022), the present study used static images to investigate the salience of event participants in language and cognition. This seems reasonable given neuroimaging evidence that action category recognition is similar across viewing of still photographs and dynamic videos (Hafri et al., 2017). Future versions of our study could generalize the present results to naturalistic dynamic events.

Finally, from the present investigation it might appear as though the organization of events in language maps onto visually observable entities in the physical world in a more or less straightforward way. However, many events lexicalized in language have few, if any, visual correlates (e.g., *think*, *decide*, *realize*). Moreover, blind individuals who have never had visual experience in their lives do have a thematic hierarchy (Landau & Gleitman, 1985). Even though visually presented of caused motion events offers a convenient empirical tool for present purposes, we would expect our conclusions to generalize to abstract events and their structure.

7. Conclusions

The present investigation offers novel evidence supporting the idea that linguistic and conceptual organization of events are largely similar. Nevertheless, the similarities between event structure in language and cognition seem to be less strict than previously hypothesized by linguistic theory. Together, these findings inform our understanding of the link between event cognition and language production and have important implications for language acquisition and how language interfaces with other aspects of cognition.

CRediT authorship contribution statement

Ercenur Ünal: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **Frances Wilson:** Methodology, Investigation, Conceptualization. **John Trueswell:** Writing – review & editing, Funding acquisition, Conceptualization. **Anna Papafragou:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Data availability

The data and code that support the findings of this study are available at https://osf.io/rtc6m/

Acknowledgments

This work was supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development under grant #R01HD055498 to A.P and J.T. Thanks to Helene Intraub, Ann Bunger, and the Language and Cognition Lab for helpful discussion, James Delorme, Madison Paron and Catherine Richards for assistance in data collection, Rick Chalton for assistance in preparation of the stimuli, and Alon Hafri for assistance with *E*-Prime. An earlier report of Experiment 2 appeared in the Proceedings from the 33rd Annual Meeting of the Cognitive Science Society.

Appendix A. Target stimuli

Items 1–24 were used in Experiments 1, 3 and 4, items 1–18 were used in Experiment 2.

1. A soldier firing a bouquet at a castle with a catapult.
2. A girl hitting a ball of wool towards a hat with a croquet mallet.
3. A man hitting a golf ball into a bucket with a golf club.
4. A man hitting an apple into a basketball hoop with an umbrella.
5. A man kicking a boot into a suitcase.
6. A man pulling a log towards a fire with a rope.
7. A ninja kicking a phone towards a grandfather clock.
8. A man pulling a block towards a pyramid with a rope.
9. A man raking leaves into a basket.
10. A soldier pushing a wheel to a truck with a stick.
11. A man hitting a ball into a goal with a mop.
12. A man shovelling gold into a sack.
13. A man shovelling manure into a truck.
14. A man sweeping dirt into a dustpan with a broom.
15. A man pulling a tree towards a house with a rope.
16. A man kicking a can into a wheelie bin.
17. A woman hitting a ball into a basket with a tennis racquet.
18. A man pulling a television into a cave with a chain.
19. An archer shooting an arrow with a bow towards a target.
20. A caveman hitting some meat towards a fire with a club.
21. A clown firing a bomb at a paddling pool with a cannon.
22. A mouse pulling a slice of cheese to a hole with a rope.
23. A man hitting some paper into a bin with a bat.
24. A cricketer hitting a present into a bowl with a cricket bat.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.105868.

References

Aginsky, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition, 7*(1), 147–162. https://doi.org/10.1080/135062800394739

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Baker, M. C. (1997). Thematic roles and syntactic structure. In L. Haegeman (Ed.), *Elements of grammar: Handbook of generative syntax* (pp. 73–137). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-5420-8.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 249–278). Psychology Press. https://doi.org/10.4324/9780203488430.

Brown, P., & Dell, G. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology, 19*, 441–472. https://doi.org/10.1016/0010-0285(87)90015-6

Bunger, A., Skordos, D., Trueswell, J., & Papafragou, A. (2021). How children attend to events before speaking: Cross-linguistic evidence from the motion domain. *Glossa: A Journal of General Linguistics, 6*(1), 28. https://doi.org/10.5334/gjgl.1210

Bunger, A., Skordos, D., Trueswell, J. C., & Papafragou, A. (2016). How children and adults encode causative events cross-linguistically: Implications for language production and attention. *Language, Cognition and Neuroscience, 31*(8), 1015–1037. https://doi.org/10.1080/23273798.2016.1175649

Bunger, A., Trueswell, J., & Papafragou, A. (2012). The relation between event apprehension and utterance formation in children: Evidence from linguistic omissions. *Cognition, 122*, 135–149. https://doi.org/10.1016/j.cognition.2011.10.002

Carlson, G. (1998). Thematic roles and the individuation of events. In S. Rothstein (Ed.), *Events and grammar* (pp. 35–51). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-3969-4_3.

Chen, Y., Trueswell, J., & Papafragou, A. (2024). Sources and goals in memory and language: Fragility and robustness in event representation. *Journal of Memory and Language, 135*. https://doi.org/10.1016/j.jml.2023.104475. Article 104475.

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. *Developmental Psychology, 29*(3), 421–433. https://doi.org/10.1037/0012-1649.29.3.421

Cohen, L. B., Rundell, L. J., Spellman, B. A., & Cashon, C. H. (1999). Infants' perception of causal chains. *Psychological Science, 10*(5), 412–418. https://doi.org/10.1111/1467-9280.00178

Davis, A. R. (2011). Thematic roles. In C. Maienborn, K. von Heusinger, & P. H. Portner (Eds.), *Vol. 1. Semantics: An international handbook of natural language meaning* (pp. 399–420). De Gruyter Mouton. https://doi.org/10.1515/9783110226614.399.

Do, M. L., Papafragou, A., & Trueswell, J. (2020). Cognitive and pragmatic factors in language production: Evidence from source-goal motion events. *Cognition, 205*. https://doi.org/10.1016/j.cognition.2020.104447. Article 104447.

Do, M. L., Papafragou, A., & Trueswell, J. (2022). Encoding motion events during language production: Effects of audience design and conceptual salience. *Cognitive Science, 46*(1). https://doi.org/10.1111/cogs.13077. Article e13077.

Dobel, C., Glanemann, R., Kreysa, H., Zwitserlood, P., & Eisenbeiss, S. (2011). Visual encoding of coherent and non-coherent scenes. In J. Bohnemeyer, & E. Pederson (Eds.), *Event representation in language: Encoding events at the language cognition interface* (pp. 189–215). Cambridge University Press. https://doi.org/10.1017/CBO9780511782039.

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica, 125*(2), 129–143. https://doi.org/10.1016/j.actpsy.2006.07.004

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*(3), 547–619. https://doi.org/10.2307/415037

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science, 33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Filip, H. (2012). Lexical aspect. In R. I. Binnich (Ed.), *The Oxford handbook of tense and aspect* (pp. 721–751). Oxford University Press.

Fillmore, C. J. (1968). The case for case. In E. Bach, & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). Holt, Rinehart and Winston.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition, 32*(2), 99–156. https://doi.org/10.1016/0010-0277(89)90001-2

Gleitman, L., Senghas, A., Flaherty, M., Coppola, M., & Goldin-Meadow, S. (2019). The emergence of the formal category "symmetry" in a new sign language. *Proceedings of the National Academy of Sciences, 116*(24), 11705–11711. https://doi.org/10.1073/pnas.1819872116

Gleitman, L. R. (1990). The structural sources of verb learning. *Language Acquisition, 1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language, 57*, 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Göksun, T., George, N., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Forces and motion: Causal understanding of preschoolers. *Child Development, 84*(4), 1285–1295. https://doi.org/10.1111/cdev.12035

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). Trading spaces: Carving up events for learning language. *Perspectives on Psychological Science, 5*(1), 33–42. https://doi.org/10.1177/1745691609356783

Golinkoff, R. M., & Kerr, J. L. (1978). Infants' perception of semantically defined action role changes in filmed events. *Merrill-Palmer Quarterly of Behavior and Development, 24*(1), 53–61. http://www.jstor.org/stable/23084058.

Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Grigoroglou, M., & Papafragou, A. (2019a). Children's (and adults') production adjustments to generic and particular listener needs. *Cognitive Science, 43*(10), Article e12790. https://doi.org/10.1111/cogs.12790

Grigoroglou, M., & Papafragou, A. (2019b). Interactive contexts increase informativeness in children's referential communication. *Developmental Psychology, 55*(5), 951–966. https://doi.org/10.1037/dev0000693

Grimshaw, J. (1981). Form, function and the language acquisition device. In C. L. Baker, & J. J. McCarthy (Eds.), *The logical problem of language acquisition* (pp. 165–182). MIT Press.

Grosz, B., Joshi, A., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics, 21*(2), 203–225. https://aclanthology.org/J95-2003.

Hafri, A., Gleitman, L. R., Landau, B., & Trueswell, J. (2022). Where word and world meet: Language and vision share an abstract representation of symmetry. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001283

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General, 142*(3), 880–905. https://doi.org/10.1037/a0030045

Hafri, A., Trueswell, J. C., & Epstein, R. A. (2017). Neural representations of observed actions generalize across static and dynamic visual input. *Journal of Neuroscience, 37*(11), 3056–3071. https://doi.org/10.1523/JNEUROSCI.2496-16.2017

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition, 175*, 36–52. https://doi.org/10.1016/j.cognition.2018.02.011

Hale, K., & Keysar, S. (1993). On argument structure and the lexical expression of syntactic relations. In K. Hale, & S. Keysar (Eds.), *The view from building 20: Essays in honor of Sylvain Bromberger*. Oxford University Press.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition, 7*(1), 213–235. https://doi.org/10.1080/135062800394775

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

van Hout, A. (2016). Lexical and grammatical aspect. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford handbook of developmental linguistics*. Oxford University Press.

Isasi-Isasmendi, A., Andrews, C., Flecken, M., Laka, I., Daum, M. M., Meyer, M., … Sauppe, S. (2023). The agent preference in visual event apprehension. *Open Mind, 7*, 240–282. https://doi.org/10.1162/opmi_a_00083

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 1–30). MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

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

Jackendoff, R. S. (1990). *Semantic structures*. MIT Press.

Ji, Y., Andrews, C., Sauppe, S., Flecken, M., Zariquiey, R., Daum, M. M., … Bickel, B. (2022). Universality and diversity in event cognition and language. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual conference of the cognitive science society* (pp. 12–14). https://escholarship.org/uc/item/9gw479mq.

Ji, Y., & Papafragou, A. (2020a). Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition, 197*. https://doi.org/10.1016/j.cognition.2020.104197. Article 104197.

Ji, Y., & Papafragou, A. (2020b). Midpoints, endpoints and the cognitive structure of events. *Language, Cognition and Neuroscience, 35*, 1465–1479. https://doi.org/10.1080/23273798.2020.1797839

Ji, Y., & Papafragou, A. (2022). Boundedness in event cognition: Viewers spontaneously represent the temporal texture of events. *Journal of Memory and Language, 127*, Article 104353. https://doi.org/10.1016/j.jml.2022.104353

Johanson, M., Selimis, S., & Papafragou, A. (2019). The source-goal asymmetry in spatial language: Language-general vs. language-specific aspects. *Language, Cognition and Neuroscience, 34*(7), 826–840. https://doi.org/10.1080/23273798.2019.1584323

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology, 73*, 1–40. https://doi.org/10.1016/j.cogpsych.2014.04.001

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 748–756. https://doi.org/10.1037/a0022637

Kuhn, J., Geraci, C., Schlenker, P., & Strickland, B. (2021). Boundaries in space and time: Iconic biases across modalities. *Cognition, 210*. https://doi.org/10.1016/j.cognition.2021.104596. Article 104596.

Lakusta, L., & Carey, S. (2015). Twelve-month-old infants’ encoding of goal and source paths in agentive and non-agentive motion events. *Language Learning and Development, 11*(2), 152–175. https://doi.org/10.1080/15475441.2014.896168

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition, 96*(1), 1–33. https://doi.org/10.1016/j.cognition.2004.03.009

Lakusta, L., & Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive Science, 36*(3), 517–544. https://doi.org/10.1111/j.1551-6709.2011.01220.x

Lakusta, L., Spinelli, D., & Garcia, K. (2017). The relationship between pre-verbal event representations and semantic structures: The case of goal and source paths. *Cognition, 164*, 174–187. https://doi.org/10.1016/j.cognition.2017.04.003

Lakusta, L., Wagner, L., O’Hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language Learning and Development, 3*(3), 179–197. https://doi.org/10.1080/15475440701360168

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press.

Landau, B., & Zukowski, A. (2003). Objects, motions, and paths: Spatial language in children with Williams syndrome. *Developmental Neuropsychology, 23*(1–2), 105–137. https://doi.org/10.1080/87565641.2003.9651889

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). Wiley.

Lenth, R. V. (2021). Emmeans: Estimated marginal means, aka least-squares means. R Package (Version 1.5.5-1) https://CRAN.R-project.org/package=emmeans.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*(3), 265–288. https://doi.org/10.1016/S0010-0277(87)80006-9

Levelt, W. (1989). *Speaking*. MIT Press.

Levin, B. (2014). Semantic roles. In M. Aronoff (Ed.), *Oxford bibliographies in linguistics*. Oxford University Press. https://doi.org/10.1093/obo/9780199772810-0141.

Levin, B., & Pinker, S. (1991). Introduction to special issue of cognition on lexical and conceptual semantics. *Cognition, 41*(1–3), 1–7. https://doi.org/10.1016/0010-0277(91)90030-8

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge University Press. https://doi.org/10.1017/CBO9780511610479

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4*(4), 501–506. https://doi.org/10.3758/BF03214339

Lockridge, C. B., & Brennan, S. E. (2002). Addressees’ needs influence speakers’ early syntactic choices. *Psychonomic Bulletin & Review, 9*(3), 550–557. https://doi.org/10.3758/BF03196312

Mathis, A., & Papafragou, A. (2022). Agents’ goals affect construal of event endpoints. *Journal of Memory and Language, 127*, Article 104373. https://doi.org/10.1016/j.jml.2022.104373

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372–380. https://doi.org/10.3758/BF03206780

McRae, K., Ferretti, T., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language & Cognitive Processes, 12*, 137–176. https://doi.org/10.1080/016909697386835

Meyer, A. S., Sleiderink, A., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66*(2), B25–B33. https://doi.org/10.1016/S0010-0277(98)00009-2

Miller, G., & Johnson-Laird, P. (1976). *Language and perception*. Harvard University Press.

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience, 30*(9), 1187–1208. https://doi.org/10.1080/23273798.2015.1006238

Nordlinger, R., Rodriguez, G. G., & Kidd, E. (2022). Sentence planning and production in Murrinhpatha, an Australian’free word order’language. *Language, 98*(2), 187–220. https://doi.org/10.1353/lan.2022.0008

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development, 5*(2), 193–207. https://doi.org/10.1016/0885-2014(90)90026-P

Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science, 34*(6), 1064–1092. https://doi.org/10.1111/j.1551-6709.2010.01107.x

Papafragou, A., & Grigoroglou, M. (2019). The role of conceptualization during language production: Evidence from event encoding. *Language, Cognition and Neuroscience, 34*(9), 1117–1128. https://doi.org/10.1080/23273798.2019.1589540

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*, 155–184. https://doi.org/10.1016/j.cognition.2008.02.007

Paul, H. (1886/1970). The sentence as the expression of the combination of several ideas. In A. L. Blumenthal (Ed.), *Language and psychology: Historical aspects of psycholinguistics* (pp. 34–37). Wiley.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press.

Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundations for Statistical Computing. https://www.Rproject.org/.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.

Rappapport Hovav, M. (2017). Argument realization in syntax. In *Oxford encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.126.

Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science, 31*(4), 705–719. https://doi.org/10.1080/15326900701399954

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Rissman, L., & Lupyan, G. (2022). A dissociation between conceptual prominence and explicit category learning: Evidence from agent and patient event roles. *Journal of Experimental Psychology: General, 151*(7), 1707–1732. https://doi.org/10.1037/xge0001146

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review, 26*(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Rissman, L., & Rawlins, K. (2017). Ingredients of instrumental meaning. *Journal of Semantics, 34*(3), 507–537. https://doi.org/10.1093/jos/ffx003

Rissman, L., Rawlins, K., & Landau, B. (2015). Using instruments to understand argument structure: Evidence for gradient representation. *Cognition, 142*, 266–290. https://doi.org/10.1016/j.cognition.2015.05.015

Rissman, L., Rawlins, K., & Landau, B. (2019). Event participants and verbal semantics: Non-discrete structure in English, Spanish and Mandarin. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual conference of the cognitive science society* (pp. 960–966). Cognitive Science Society.

Rissman, L., van Putten, S., & Majid, A. (2022). Evidence for a shared instrument prototype from English, Dutch, and German. *Cognitive Science, 46*(5). https://doi.org/10.1111/cogs.13140. Article e13140.

Sauppe, S. (2017). Word order and voice influence the timing of verb planning in German sentence production. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01648. Article 1648.

Sauppe, S., & Flecken, M. (2021). Speaking for seeing: Sentence structure guides visual event apprehension. *Cognition, 206*. https://doi.org/10.1016/j.cognition.2020.104516. Article 104516.

Sauppe, S., Norcliffe, E., Konopka, A. E., Van Valin, R. D., & Levinson, S. C. (2013). Dependencies first: Eye tracking evidence from sentence production in Tagalog. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 1265–1270). Cognitive Science Society.

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10-and 12-month-old infants. *Psychological Science, 16*(12), 995–1001. https://doi.org/10.1111/j.1467-9280.2005.01649.x

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*. https://doi.org/10.1016/j.jml.2019.104038. Article 104038.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-prime user’s guide*. Psychology Software Tools, Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-prime reference guide*. Psychology Software Tools, Inc.

Shipley, T. F. (2008). An invitation to an event. In T. Shipley, & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 1–30). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195188370.001.0001.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7*(1–3), 1–15. https://doi.org/10.1080/135062800394658

Slobin, D. I. (1996). From “thought and language” to “thinking for speaking.”. In J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

Sorace, A. (2010). Using magnitude estimation in developmental linguistic research. In E. Blom, & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 57–72). John Benjamins. https://doi.org/10.1075/lllt.27.

Strickland, B. (2017). Language reflects “core” cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive Science, 41*(1), 70–101. https://doi.org/10.1111/cogs.12332

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences, 112*(19), 5968–5973. https://doi.org/10.1073/pnas.1423080112

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description: Grammatical categories and the lexicon* (pp. 36–149). Cambridge University Press.

Talmy, L. (2000). A typology of event intergration. In *, Vol. 2. Toward a cognitive semantics: Typology and process in concept structuring* (pp. 213–288). The MIT Press.

Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science, 13*(1), 224–242. https://doi.org/10.1111/tops.12475

Ünal, E., Manhardt, F., & Özyürek, A. (2022). Speaking and gesturing guide event perception during message conceptualization: Evidence from eye movements. *Cognition, 225*, Article 105127. https://doi.org/10.1016/j.cognition.2022.105127

Ünal, E., & Papafragou, A. (2016). Interactions between language and mental representations. *Language Learning, 66*(3), 554–580. https://doi.org/10.1111/lang.12188

Ünal, E., Richards, C., Trueswell, J., & Papafragou, A. (2021). Representing agents, patients, goals and instruments in causative events: A cross-linguistic investigation of early language and cognition. *Developmental Science, 24*(6). https://doi.org/10.1111/desc.13116. Article e13116.

van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. *Journal of Memory and Language, 71*(1), 124–144. https://doi.org/10.1016/j.jml.2013.11.001

Wagner, L., & Lakusta, L. (2009). Using language to navigate the infant mind. *Perspectives on Psychological Science, 4*(2), 177–184. https://doi.org/10.1111/j.1745-6924.2009.01117.x

Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica, 133*(1), 51–56. https://doi.org/10.1016/j.actpsy.2009.09.001

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wilbur, R. (2003). Representations of telicity in ASL. In *Proceedings from the annual meeting of the Chicago linguistic society* (pp. 354–368). Chicago Linguistic Society.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology. General, 136* (1), 82–111. https://doi.org/10.1037/0096-3445.136.1.82

Wolff, P. (2008). Dynamics and the perception of causal events. In T. Shipley, & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 555–587). Oxford University Press.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Wundt, W. (1900/1970). The psychology of the sentence. In A. L. Blumenthal (Ed.), *Language and psychology: Historical aspects of psycholinguistics* (pp. 20–31). Wiley.

Yates, T. S., Sherman, B. E., & Yousif, S. R. (2023). More than a moment: What does it mean to call something an 'event'? *Psychonomic Bulletin & Review, 30*(6), 2067–2082. https://doi.org/10.3758/s13423-023-02311-4

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science, 26*(8), 1316–1324. https://doi.org/10.1177/0956797615588753

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zwitserlood, P., Bölte, J., Hofmann, R., Meier, C. C., & Dobel, C. (2018). Seeing for speaking: Semantic and lexical information provided by briefly presented, naturalistic action scenes. *PLoS One, 13*(4). https://doi.org/10.1371/journal.pone.0194762. Article e0194762.