

Chen, Y., Trueswell, J. C., & Papafragou, A. (2024). Sources and goals in memory and language: Fragility and robustness in event representation. *Journal of Memory and Language*, 135(0749-596X), 104475–104475.

Title: Sources and Goals in Memory and Language: Fragility and Robustness in Event Representation

Authors: Yiran Chen, John Trueswell, Anna Papafragou

Corresponding Author: Yiran Chen, Department of Linguistics, University of Pennsylvania, 3401 Walnut St, Suite 300C, Philadelphia, PA, 19104, USA. Email: chen39@sas.upenn.edu

Abstract:

Previous research has demonstrated an asymmetry between Sources and Goals in people's linguistic and non-linguistic encoding of motion events: when describing events such as a fairy going from a tree to a flower, people mention the Goal ("to a flower") more often than the Source ("from a tree"); similarly, people are better at detecting Goal than Source changes in memory tests. However, all prior work used a single task to probe memory of Sources and Goals and thus left the nature of the fragility of event components open. Here, we probed memory for Sources and Goals using either a Same-different or a Forced-choice task after participants passively viewed (Experiment 1), viewed and described (Experiment 2) or viewed and heard descriptions of (Experiment 3) the same set of motion events. We robustly replicated the linguistic Source-Goal asymmetry. However, across encoding contexts, the memory asymmetry persisted in the Same-different task but *disappeared* in the Forced-choice task. The Same-different task results did not change even when participants were explicitly asked to attend to Sources (Experiment 4a) and when motion trajectory was removed at test (Experiment 4b), ruling out a purely test-expectation account for the cross-task effect. We conclude that Sources of motion, even when not mentioned in language nor successfully retrieved at memory test, are nevertheless represented as part of a motion event, and their detailed representation can be reinstated at aided retrieval contexts. Our data clarify the nature of event representation and suggest a fine-grained homology between language and event memory.

Keywords: Source-Goal Asymmetry; Goal bias; Event Cognition; Thematic Roles; Psycholinguistics

1. Introduction

Language has long been argued to reflect core aspects of human perception and cognition (Slobin, 1982; Jackendoff, 1996; Levelt, 1989; Pinker, 2013; Gleitman & Papafragou, 2012). In linguistic typology, patterns that align with perceptual and conceptual preferences are more widely attested in world languages than those that do not (Culbertson, 2012; Dixon, 1994; Croft, 2002, among others). This alignment is perhaps most apparent in the way we perceive and communicate about events - segmented and structured units of our continuous and dynamic experience. Oftentimes, the relative prominence of different components of an event - for example, the agent or patient of an action, the goal of an action or the instrument to achieve the action - points to parallels between language and non-linguistic event cognition, with some components being more robustly represented compared to others in both domains (e.g., Wilson, Ünal, Trueswell, & Papafragou, 2014; cf. Lakusta & Landau, 2005, 2007, 2012; Papafragou, 2010; Hayward & Tarr, 1995, Munnich, Landau, & Doshier, 2001; Landau & Hoffman, 2005). For instance, just as instruments used to achieve an action (e.g., hit the ball **with a bat**) are often expressed in optional adjunct positions instead of as obligatory arguments across languages, their visual representations are also less robustly remembered than those of other event components (e.g., the ball) (Ünal, Richards, Trueswell, & Papafragou, 2021).

Despite these findings, there is no consensus about the precise shape of event representation and its interface with linguistic processes. For instance, it remains unclear what exactly it means for an event component to be more robust (or fragile) in both non-linguistic event cognition and language. In the studies cited above, the fragility of an event component was often indexed by less frequent or more peripheral mention in language and/or poorer performance in memory tasks, but no explicit proposal about the nature of such fragility has been made. Relatedly, the connection between the granularity of linguistic construal and that of non-linguistic event representation has rarely been examined since the two kinds of representations have typically been tested separately. Currently, it is an open question whether explicit linguistic encoding might not simply reflect but also affect an existing asymmetry between different components in event cognition. These questions are crucial to understanding how event cognition and language come together and how they diverge (see Papafragou & Grigoroglou, 2019, for a recent review).

In this article, we aim to further our understanding of non-linguistic event cognition and language by addressing the relative robustness of individual event components across the two domains. We focus on a particularly clear test case, the asymmetrical representation of Sources and Goals in motion events, and explore its roots, scope and relation to linguistic encoding.

1.1 Sources and Goals in Language and Memory

According to a now extensive literature, Goals (end points) are prioritized over Sources (starting points) in the representation of motion events. As shown by both production studies in a laboratory setting as well as analyses of naturalistic corpora, when describing simple motion events such as a fairy going from a tree to a flower, people mention the Goal (“to a flower”)

more often than the Source (“from a tree”) (Stefanowitsch & Rohde, 2004; Lakusta & Landau, 2005, 2012; Papafragou, 2010; Regier & Zheng, 2007; Do, Papafragou & Trueswell, 2020). Such a linguistic asymmetry has been observed for adults as well as children (Papafragou, 2010; Lakusta & Landau, 2012; Lakusta, Muentener, Petrillo, Mullanaphy, & Muniz, 2016), in different subtypes of motion events involving Goal and Source paths (e.g., Change of State: Lakusta & Landau, 2005; Attachment/Detachment: Regier & Zheng, 2007, Transfer of possession, Lakusta & Landau, 2005, Chen, Papafragou & Trueswell, 2022) as well as in typologically (even modally) different languages (e.g., Regier & Zheng, 2007; Johanson, Semilis, & Papafragou, 2019; Zheng & Goldin-Meadow, 2002; Ihara & Fujita 2000; Lakusta, Yoshida, Landau & Smith, 2006). Further, this linguistic asymmetry also manifests itself in typological patterns: cross-linguistically, languages have semantically finer-grained and syntactically more prominent devices to encode Goals than Sources (Fillmore, 1997; Ihara & Fujita, 2000; Jackendoff, 1985; Nikitina, 2009; Nam, 2004; Markovskaya, 2006; Regier & Zheng, 2007; Johanson & Papafragou, 2010; Johanson & Papafragou, 2014) and children’s earlier acquisition Goal-encoding language than Source-encoding language (Freeman, Sinha & Stedmon, 1980; Bowerman, de Leon & Choi, 1995; Pleh, 1998; Landau & Zukowski, 2018).

The asymmetry between Sources and Goals in language seems to have conceptual roots (Regier, 1996; Regier & Zheng, 2007; Lakusta & Landau, 2005; 2012; Papafragou, 2010). Much evidence suggests that Goals are more accurately encoded in memory than Sources of motion events: adults and 4-5 year olds detect Goal-changes more accurately than Source-changes (e.g., Papafragou, 2010; Lakusta & Landau, 2012; Regier & Zheng, 2007; Do et al., 2020). Even prelinguistic infants prefer to look at an event with a Goal (e.g., a duck going into a bowl) than one with a Source (e.g., a duck going out of a bowl) (Lakusta & DiFabrizio, 2017) and look longer to a test event where a Goal object has changed from an earlier, familiarization event compared to a test event where the Source object has changed (Lakusta, Wagner, O’Hearn, & Landau, 2007; Lakusta & Carey, 2015).

This parallel asymmetry is often taken as a golden case for a homology between language and non-linguistic cognition: Source, the event role that is less likely to be present in linguistic description, is also less robustly encoded in non-linguistic representation. However, two major issues with regard to the nature of the representation of motion Goals and Sources remain unclear. First, previous literature interprets these findings as demonstrating that Source is less robustly represented as people encode motion events. However, what it means for Source to be less robustly represented is left open. Do people sometimes fail to encode Source entirely and instead represent the motion event as one with only a Goal (e.g., the fairy went to the flower)? This is what linguistic descriptions of events like this might suggest but we lack evidence that the non-linguistic encoding matches this pattern. Alternatively, is Source always encoded in memory, but with a weaker strength or with coarser granularity? Understanding how Source, a well-documented “fragile” component, is represented will shed light on the nature of robustness in event cognition in general.

One way to approach the nature of the Source fragility is to ask whether Source information escapes memory altogether, or is present and can be revived when the retrieval context calls for

it. Prior work has not been able to address this question because, to the best of our knowledge, it has always tested recognition memory of Source and Goal via a single test format - the Same-Different task (Papafragou, 2010; Regier & Zheng, 2007; Regier, 1996; Do et al., 2020; Lakusta et al., 2007). In this task, participants view a set of motion videos and later compare new videos to their memory of the previous target event and make a decision (same or different). Surface failure to recognize a Source change in this task could represent unavailability of Source information in the representation, but could also mean that Source information is harder to access (Tulving & Pearlstone, 1966). Given that the same memory trace can be better retrieved when given a retrieval context better matching that of encoding (Godden & Baddeley, 1975; Tulving & Thomson, 1973), the asymmetrical robustness of Sources and Goals may not surface in the same way in different memory tasks where different cues are available.

A second question related to the stability and nature of the Source-Goal asymmetry is how Source-Goal representation might change as a result of linguistic encoding. Previous research suggests that the Goal bias in memory persists even when the experimental design explicitly inhibits linguistic encoding by introducing a verbal shadowing task as people view the visual stimuli (Lakusta & Landau 2012). However, could the act of describing the event verbally or the act of hearing a linguistic description during encoding affect the Source-Goal asymmetry in memory at a later stage? After all, it has been demonstrated that linguistic labels can affect memory (e.g., Archambault, O'Donnell & Schyns, 1999; Feist & Gentner, 2007; Loftus, 1975, 1979; Loftus & Palmer, 1974; Loftus, Miller, & Burns, 1978; Lupyan, 2008). Within the domain of event cognition, hearing verbs that denote paths of motion before viewing motion events attenuated memory for manners of motion (Skordos et al., 2021). Other studies where participants both provided event descriptions and were tested on event memory have reported that memory of individual event components was stronger when these components were also mentioned (e.g., Do et al., 2020). However, relatively few studies have directly looked at how language production and comprehension modulate event cognition.

1.2 Current Study

To address the two main questions outlined above, the current study tested the stability of the classic Source-Goal asymmetry across different memory tasks and after different levels of explicit linguistic engagement with the visual stimuli.

To examine whether Sources and Goals are consistently remembered asymmetrically, we probed memory for Source and Goal in the Same-different task and compared results to another widely-used recognition test format - the Forced-choice task. In this task, after participants have encoded items in memory, pairs of items - a target and a foil - appear at the same time during the memory test phase, and the participants are asked to select the target. It has long been recognized that, despite the wide use of the two tasks as comparable measures of recognition memory, the assessment of memory can function differently in the two test formats (Deffenbacher, Leu & Brown, 1981; Bastin & Van der Linden, 2003; Aggleton & Shaw 1996, O'Reilly & Rudy, 2000; Holdstock et al., 2002, Westerberg et al., 2006; Kroll et al., 2002; Bayley,

Wixted, Hopkins & Squire, 2008, Khoe et al., 2000; Mayes, Holdstock, Isaac, Hunkin & Roberts, 2002).¹ One critical point of difference between the two test formats is the availability of retrieval cues at test. One possibility is that the direct accessibility to the target at test might boost Goal and Source equally, preserving the Source-Goal asymmetry. Alternatively, the reinstatement of the target could benefit the component that was more weakly encoded (i.e., the Source) to a greater extent, pulling it to the same level of recognition with the more strongly encoded role (the Goal), thus eliminating the bias. The direct comparison of the two memory tasks could thus throw light onto the nature of the Source disadvantage: if Source indeed fails to be encoded at least some of the time, it would not be salvaged by aided retrieval in the Forced-Choice format. However, if it is characterized by a weaker overall trace or coarser granularity of encoding, Source might benefit from the richer retrieval cues in the Forced-Choice format.

To test whether these results would generalize across different contexts of linguistic encoding, we probed memory for Sources and Goals after people viewed events without explicit involvement of language (passive viewing), with the task of describing them (production), or with a verbal description about the events presented to them (comprehension). We also had two more specific aims. First, we wanted to ask whether a linguistic asymmetry in production or comprehension would directly align with the non-linguistic cognitive bias (such that, e.g., Sources that are more poorly remembered would also be less likely to be mentioned). Second, we wanted to probe the effect of language comprehension on the non-linguistic Source-Goal asymmetry. Since Source-only descriptions are unusual in listeners' linguistic and perceptual experience (Lakusta & Landau, 2005; Fisher et al., 1994), and unexpected information is remembered better (Graesser, 2013; Shapiro & Fox 2002), hearing a Source-only description might boost Source memory more robustly compared to the effect of hearing a Goal-only description on Goal memory, thus mitigating the asymmetry.

In sum, we conducted three experiments to compare memory for Sources and Goals, probed with either a Same-different task or a Forced-choice task, after participants passively viewed (Experiment 1), viewed and described (Experiment 2) or viewed and heard descriptions of (Experiment 3) the same set of motion events. In a further experiment, we considered and removed alternative interpretations of our findings (Experiment 4). Together, our data can inform theories of event representation and the language-cognition interface (especially those probing representational asymmetries among more and less prominent event components). Our investigation also makes a crucial methodological contribution by highlighting the need for convergent findings from different tasks probing language and memory across multiple contexts to construct valid theories connecting language to non-linguistic event cognition.

1.3 Data Availability

For all the experiments that we report in what follows, trial-level data and analysis code can be found in an OSF repository: <https://osf.io/neqcw/>.

¹ Researchers take different positions as to whether the two tasks differentially tap into “familiarity” and “recollection”. We do not comment further on this issue since this distinction is less central here.

2. Experiment 1

In Experiment 1, we tested memory of Sources and Goals after participants passively viewed motion events. We expected that the Source-Goal asymmetry in memory would be replicated in the Same-Different task (as prior literature using this task has shown; Regier & Zheng, 2007; Lakusta & Landau, 2012; Papafragou, 2010; Do et al., 2020) but might not surface in the Forced-Choice task where more retrieval cues were provided.

2.1 Methods

2.1.1 Participants

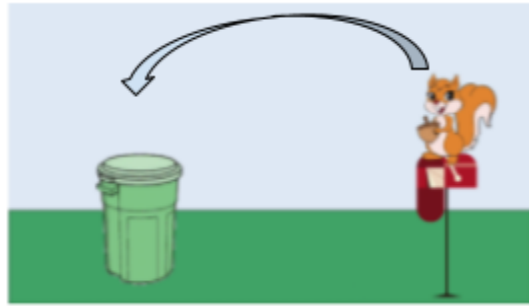
A total of 120 native speakers of American English were recruited on Prolific (www.prolific.co). They participated for compensation at a rate of \$6.5/hour. The number of participants was determined based on a power analysis of effect size on the adult participants reported in Experiment 1 of Papafragou (2010). Given $\alpha = 0.05$ and desired power of 0.80, the projected sample size in each group was 17. Given some small design differences (i.e., we have slightly more items and added one more between-subject manipulation), we decided to recruit 20 participants for each experimental list before carrying out the study.

2.1.2 Materials

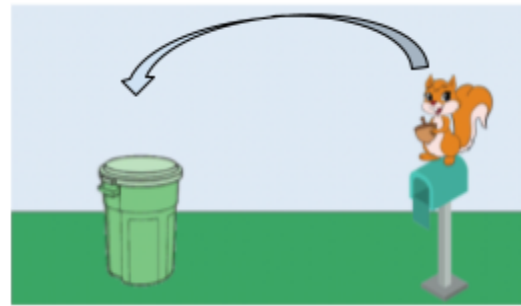
We created 16 critical video clips, each of which depicted an animate Figure moving from an inanimate Source landmark (i.e. the starting point of motion) to an inanimate Goal landmark (i.e. the end point of motion). The Figure, Source and Goal were all represented by clipart images (See Figure 1a for an example of a critical clip). The motion was achieved through Powerpoint Animation. Each clip lasted five seconds.

The direction of the motion was left-right counterbalanced such that half of our clips showed a Figure moving from left to right and the other half showed a figure moving from right to left. We constructed two experimental lists to counterbalance Source and Goal landmarks such that objects which were the Sources in one list were the Goals in the other. Thus, the visual salience and the namability of these objects were completely counterbalanced across lists. We also created 12 filler motion events, which did not involve a Source/Goal path (e.g., A ghost moves around the moon).

To probe speakers' memory of Sources and Goals, we also constructed foil videos that involved either a Source Change or a Goal Change. Source and Goal changes were always within-category (e.g., the mailbox was changed to another mailbox, see comparison of Figure 1a and 1b).



(Figure 1a)



(Figure 1b)

Figure 1. (a) A sample critical event (“the squirrel went from the mailbox to the trash can”). (b) A sample foil for this event involving Source change. The arrows were only included here to illustrate the figure’s motion path. Participants saw animated video clips of the events without the arrows.

2.1.3 Procedure

Participants were directed to this online experiment via a URL link. First, we familiarized participants with all the clipart images that would later appear in either the target video clips or the foil video clips in the memory task. This was done to minimize any visual familiarity effects that could arise in the memory task. These pictures were presented one at a time at the center of the screen and proceeded automatically every two seconds. Then, participants proceeded to the exposure phase, where they were asked to carefully view video clips from cartoons and were told that we would ask them questions about these video clips later. The 28 clips were presented in a pseudo-random order. Each video clip disappeared after it played once. Meanwhile, participants heard a beep and automatically proceeded to the next trial 2.5s after the beep.²

The test phase immediately followed. In that phase, participants were randomly assigned to either the Same-Different task ($n = 80$) or to the Forced-Choice task ($n = 40$). In the Same-Different task, during each test trial, participants were shown either a Source Change or a Goal Change variant of the critical videos. For each participant, half of the critical items had a Source Change and the other half had a Goal Change. Participants were instructed to click ‘Yes’ if the video clip shown here at test was ‘exactly the same’ as the clip that they had originally seen and click ‘No’ otherwise. Which item was a Source or Goal change was counterbalanced across participants. There was no time limit on participants’ response, and the video was on a loop. (See Figure 2 for a sample trial). For fillers, the memory phase displayed the original version of the clips. Thus, correct responses on critical events were always ‘No’ and the correct response for filler events was always ‘Yes’. Events were tested in a different pseudo-random order from the order in which they were viewed.

² This is to make the presentation interval between stimuli comparable to that in Experiment 2 and 3.

Have you seen a video that's exactly the same as this one before?



Figure 2: A Sample Same-Different Memory Test Trial in Experiment 1.

In the Forced-Choice task, on each trial, participants chose which video they had seen from 4 video options: the target, a foil that only differed in the Source, a foil that only differed in the Goal and a foil that differed in both (See Figure 3 for an example memory test trial). The foil landmark images were the same ones across the two task conditions. Given the test format, for the same event, participants are tested on both Goal and Source at the same time. Therefore, the Forced-Choice task only had two experimental lists that contained events that were path reversals of each other (as opposed to the four lists used in the Same-Different task).

Which video did you see? Click on the video you saw.

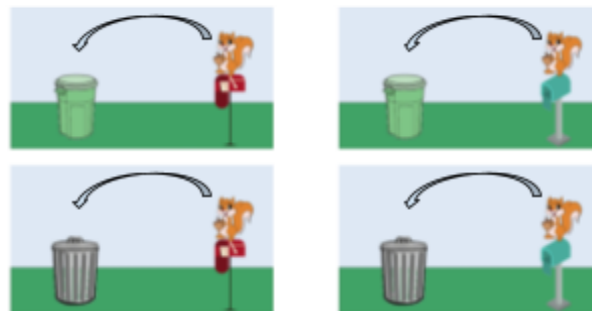


Figure 3: A Sample Forced-Choice Memory Test Trial in Experiment 1.

In the Same-Different task, each participant was randomly assigned to one of four experimental lists - List 1a, List 1b, List 2a and List 2b. Videos in List 1 and 2 were path reversals: while participants in List 1a and 1b saw the video in which the squirrel went from the mailbox to the trash can, participants in List 2a and 2b saw the video in which the squirrel went from the trash can to the mailbox. List 1a and 1b (as well as List 2a and 2b) only differed in whether participants were tested with a Source change or a Goal change on the same event. For example, after having seen the event “the squirrel went from the mailbox to the trash can”, participants in List 1a were tested with the mailbox change (as in Figure 1a vs. Figure 1b), while participants in List 1b were tested with the trash can change. In all lists, half of the target test trials involved a Source change whereas the other half involved a Goal change. In the Forced-Choice task, participants were just assigned to two lists - List 1 and List 2 - that

contained events that were path reversals of each other. The List 1a vs. 1b distinction was no longer needed because participants were tested on Sources *and* Goals as they made a selection.³ Given this design, we were able to compare memory of Sources and Goals within subjects.

2.1.4 Analysis

To ask whether Goal landmarks were remembered more accurately than the Source landmarks and whether test formats modulated such an asymmetry, we built logistic mixed-effect models with R (R Core Team, 2022).

First, we coded the memory data in the two task conditions in a way that allowed us to compare them. For the Same-Different task, we coded whether each test trial was Correct (0 vs. 1) and whether it was a Goal-Change/Source-Change trial. We excluded participants who always selected “Yes” in the Same-Different task from analysis ($n = 2$). In the Forced-Choice task, we coded each test trial for whether the event chosen by the participant contained the correct Goal (0 vs. 1) and/or the correct Source (0 vs. 1) respectively. Therefore, each Forced-Choice trial was converted into two data points - whether it was correct in terms of Goal and whether it was correct in terms of Source. As a result, chance performance is 0.5 for both tasks, allowing for direct comparison of performance across tasks. It is important to note though that given this transformation, participants in the Forced-Choice Task each offer twice as much data as those in the Same-Different task. To ensure the same total number of observations across tasks, there were twice as many participants in the Same-Different task than the Forced-choice task.

Then, we built our models. In the primary model, we predicted whether the response on a critical trial was correct with a fixed effect of Change Type (Goal vs. Source, sum coded), Task (Same-Different vs. Forced-Choice, sum coded) and their interaction. For this model, as well as all the models in the following experiments reported in this study, based on the suggestions by Barr et al. (2013), we started with the maximal random structure and reduced complexity when the model failed to converge and no non-intrusive adjustments (e.g., increasing max number of iterations) could help reach convergence and/or overparameterization is detected (Bates et al., 2015). When simplifying random-effect structures, we reduced random-effect variance components iteratively until there was significant loss to the goodness of fit of the model (Bates et al., 2015). We report the final random effect structure. This first primary model included by-subject and by-item random intercepts, by-subject and by-item random slopes of Change Type and by-item random slopes of Task.

We further built two separate models for each task, predicting whether the response was correct with Change Type (Goal vs. Source, sum coded) as a fixed effect. The final models for both conditions included by-subject and by-item random intercepts.

2.2 Results

³ Experiments 2 and 3 both adopted this design as well.

Figure 3 presents the mean proportion of correct responses as a function of Change Type (Goal vs. Source) and Task (Same-Different vs. Forced-Choice). Our primary model revealed a significant main effect of Change Type such that participants were more accurate for Goals than Sources ($\beta = 0.167$, $SE = 0.044$, $p < 0.001$); a significant main effect of Task, such that participants were more accurate on the Forced-Choice task than the Same-Different task ($\beta = -0.804$, $SE = 0.106$, $p < 0.001$), as well as a significant interaction between them ($\beta = 0.174$, $SE = 0.047$, $p < 0.001$).⁴ Importantly, the significant interaction suggests that the extent to which memory accuracy of Sources and Goals differs was different across the two tasks, which is in line with our prediction that the Source-Goal asymmetry in memory might not manifest itself in aided retrieval contexts.

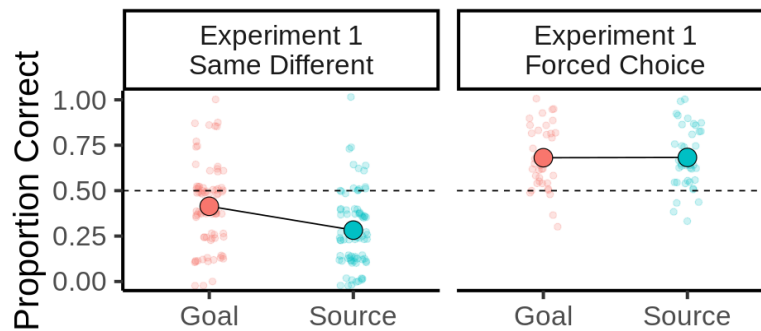


Figure 4: Participants' mean proportion of correct Goal and Source memory responses at test in Experiment 1 (Same-Different task and Forced-Choice task). In the Same-Different task, correct responses refer to successfully detecting the change of Goal/Source. In the Forced-Choice task, correct responses refer to selecting the event that contains the correct Goal/Source.

To gain further insight into the interaction, we turn to the separate models for each task. In line with prior work in the literature, in the model for the Same-Different task, we found that participants were significantly more likely to detect a Goal change when probed with a Same-Different task (Mean Goal = 0.414, $SD = 0.225$; Mean Source = 0.283, $SD = 0.205$; $\beta = 0.335$, $SE = 0.084$, $p < 0.001$) (See Figure 4 Left). The Same-Different task thus fully replicated the Source-Goal asymmetry observed in prior work: Participants were more sensitive to Goal change than to Source change in the same motion event.

Conversely, the model for the Forced-Choice task revealed no effect of Change Type ($p = 0.95$) (See Figure 4 Right). Participants were as likely to choose the event with the correct Goal as they were to choose the event with the correct Source (Mean Goal = 0.681, $SD = 0.166$; Mean Source = 0.683, $SD = 0.169$). The mean proportion of selecting each of the four alternatives was included in the Appendix.

It is notable that, overall, performance was better in the Forced-Choice task (mean proportion Goal correct 68%, Source correct = 68%) than the Same-Different task (mean proportion Goal

⁴ To account for the fact that participants in the Forced-Choice task contributed two data-point on each trial while participants in Same-Different task contributed one, we also ran a model with nested random effects to reflect that (glmer syntax: $Correct \sim Change_type * Task + (1 + Change_type | Data\text{-}point / participant) + (1 + Change_type + Task | Event)$). This model showed the same results with regard to the fixed effects.

correct 44%, Source correct = 28%)⁵. This is in line with previous research comparing these methods (Green & Moses, 1966; Deffenbacher, Leu & Brown, 1981; Macmillan & Creelman, 1994). One might wonder, however, whether the absence of the Goal advantage in the Same-Different task is the product of a simple ceiling effect: as participants approached perfect performance (100% correct) in this task, would any Goal advantage necessarily disappear simply because of the bounded nature of the measure? To address this concern, we split the participants in both tasks into high and low accuracy groups based on whether their overall memory accuracy was above or below median accuracy respectively. As shown in Figure 5 (Row 1), and contrary to this alternative possibility, the Goal bias was attested in both better-performing and worse-performing participants when a Same-Different task was used (left), and not attested in either group when a Forced-Choice task was used (right). Thus, it seems unlikely that the difference across tasks can be reduced to the difference in baseline performance.

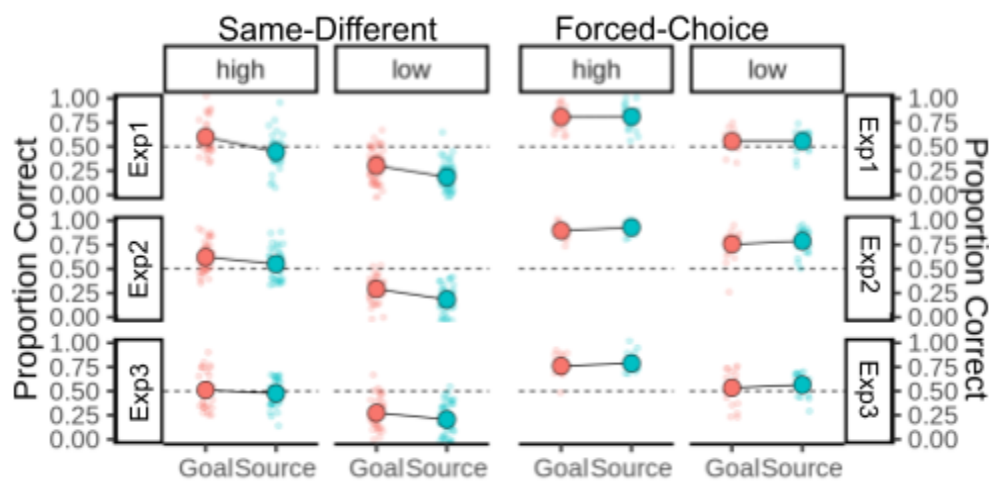


Figure 5: High- versus low-accuracy participants' mean proportion of correct Goal/Source response at memory test in Same-Different task and Forced-Choice task in Experiment 1, 2 and 3.

2.3 Discussion

We observed a striking difference in Source-Goal memory when we used a different means of probing what was remembered. Using the traditional Same-Different task, we replicated the well established Goal advantage: participants were better at detecting Goal changes than Source changes. However, using a Forced-Choice task in which participants were asked to

⁵ Indeed, performance in the Same-Different task was quite low (below chance) (Figure 4 left). This means that our participants had a tendency to say “yes” and had a hard time in general detecting any differences. Their performance was also lower than what was reported in previous studies that used a similar task (e.g., Papafragou, 2010). We attribute this drop in overall accuracy to the increased difficulty in our task due to 1) inclusion of a familiarization phase to minimize a potential pure familiarity effect, 2) increased number of test items, 3) the fact that stimuli were presented in a different random order between memory exposure and test and 4) general instructions about the Same-Different test, just like in the Forced-Choice test (e.g., we did not give examples of what a change to the event might look like). The first three differences granted us increased power compared to previous studies and that the last was a justified choice for the purpose of better comparing with the Forced-Choice test format.

select among the same video and alternatives that changed either the Source, the Goal, or both, we found no Goal advantage. This pattern cannot be explained by the baseline performance difference in the two tasks.

Such a finding is consistent with the claim that Source information is encoded in memory, even in cases where participants' performance in the Same-Different task indicate that they fail to remember it. Because Sources are encoded in memory, though less robustly than Goals, they can be recovered by increased accessibility at test: when all possible combinations of true and false Sources and Goals were present at test, people were equally good at picking the true Goal and the true Source. Given the novelty of this finding, it is important to replicate it in different encoding contexts. We do so in Experiment 2 and 3.

3. Experiment 2

Experiment 2 asked whether the memory patterns in Experiment 1 would be replicated after participants produced descriptions of the events (see Do et al., 2020, for evidence that the Source-Goal asymmetry tested via a Same-Different task persists after a description task). The experiment also probed the relation between the content of individual descriptions and memory performance.

One might expect an even more pronounced Goal bias in memory after linguistic encoding due to a potential additive effect: since English has more available and more finer-grained linguistic devices for describing Goal-path than Source-path (e.g., *onto*, *into*, *unto* for Goal paths, but only *from* for the corresponding Source paths; Lakusta & Landau, 2009), examining an event for the purpose of planning a linguistic description may further highlight the Goal in the event construal. Alternatively, the Goal bias might be reduced after linguistic encoding because, if people have overall more heightened attention when they are tasked with describing the events, peripheral components such as Source might in turn receive more attention than otherwise. Given these mixed predictions about how linguistic encoding might affect the Goal bias in memory, it was not obvious whether we would observe an effect of retrieval conditions in Experiment 2.

3.1 Methods

3.1.1 Participants

A total of 123 native speakers of American English recruited from the University of Pennsylvania subject pool and Prolific participated in the experiment. To better ensure the equivalence of the two pools, we only recruited Prolific participants between the ages of 18-26 who identified themselves as students.⁶ The University subjects received course credit, and the Prolific subjects were compensated at a rate of \$6.5/hour.

3.1.2 Procedure

⁶ These criteria apply to all our experiments reported here that included recruitment from Prolific.

In Experiment 2, instead of passively viewing the events, participants were asked to type a description of the events in a text box, cued by a beep after the video played once and disappeared. The rest of the experiment was identical to Experiment 1. We randomly assigned 83 participants to the Same-Different task and 40 participants to the Forced-Choice task.

3.1.3 Analysis

Our primary interest within the linguistic description data was whether participants would mention Goals more often than Sources, as in prior work (e.g., Landau & Lakusta 2005; Papafragou, 2010; Do et al., 2020; Chen et al., 2022). To do that, we coded whether each utterance included mention of the Source and/or Goal landmark. Then, to determine whether there was a linguistic Source-Goal asymmetry, we built a logistic mixed-effects model. We predicted landmark mention with Role type (Goal vs. Source) as a fixed effect and included by-subject and by-item random intercepts as well as by-subject and by-item random slopes of Role type.

For the memory data, our primary interest was two-fold. First, we wanted to test whether the task-driven modulation of the Source-Goal asymmetry from Experiment 1 would be replicated when there was explicit linguistic involvement during encoding. To achieve this goal, we built the same main model as in Experiment 1 to predict whether a memory response was correct (with respect to Goal/Source) with Change Type (Goal vs. Source, sum coded), Task (Same-Different vs. Forced-Choice, sum coded) and their interaction as fixed effects. The final model included by-subject and by-item random intercepts, and a by-item random slopes for Task. We also built two separate models for each task.

Second, we explored how participants' description of these events interacted with memory. To do that, we added Source Mention - that is, whether participants mentioned the Source in their previous linguistic description for this event, as a fixed-effect predictor in the model for each task.⁷ Therefore, for each task, we predicted whether the response was correct with Change Type, Source Mention and their interaction as fixed effects. The final model for the Same-Different task included by-subject and by-item random intercepts. The final model for the Forced-Choice task also included by-subject and by-item random slopes for Change Type. We applied the same exclusion criterion as in Experiment 1, and excluded five participants from the Same-Different task for always selecting "Yes".

3.2 Results

3.2.1 Language Production

We replicated the Goal bias in language observed in prior work: participants were more likely to mention the Goal than the Source in their descriptions of the events (Mean Goal mention = 0.95,

⁷ Only Source Mention is of interest here because, based on prior work, Goal is almost always mentioned in such motion events (e.g., 93% in Do et al., 2020), thus we expected to see little variability.

Mean Source mention = 0.77; Same-Different: $\beta = 0.661$, $SE = 0.159$, $p < 0.001$; Forced-Choice: $\beta = 1.142$, $SE = 0.241$, $p < 0.001$) (Figure 6 Top Row).

3.2.2 Memory for Sources and Goals

Our main model revealed a significant main effect of Task ($\beta = -1.158$, $SE = 0.129$, $p < 0.001$), and a significant interaction between Change Type and Task ($\beta = 0.183$, $SE = 0.051$, $p < 0.001$). Just like in Experiment 1, participants were more likely to be correct on the memory test when tested with a Forced-Choice format (Figure 6 Bottom row). This was not surprising because it was a much more actively aided retrieval condition.

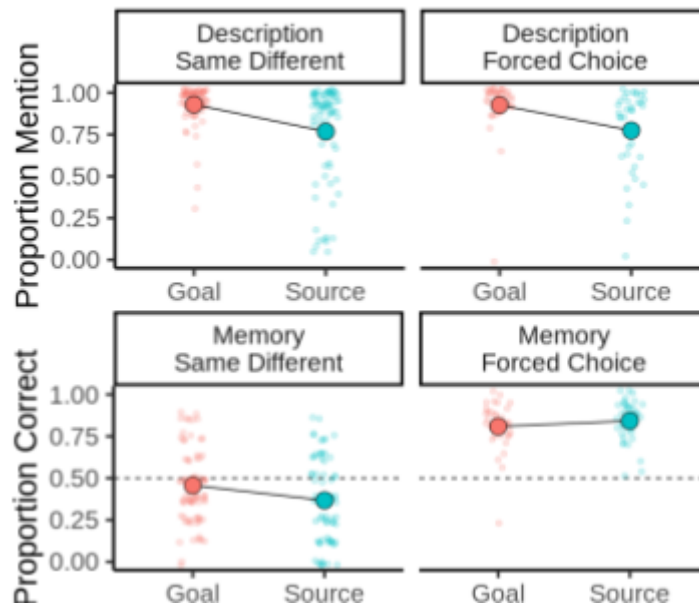


Figure 6: Participants' mean proportion of mentions of Goal/Source in their linguistic descriptions and mean proportion of correct Goal/Source responses at the memory test in Experiment 2.

Unlike in Experiment 1, we did not find a main effect of Change Type in Experiment 2, suggesting that, collapsing across the two tasks, there was no longer a reliable memory advantage for Goals over Sources. However, the significant interaction between Change Type and Task revealed that the Source/Goal memory might pattern differently in the two tasks, and invites further investigation with models for each task.

Our model for the Same-Different task revealed a main effect of Change Type (Mean Goal = 0.457, $SD = 0.221$; Mean Source = 0.367, $SD = 0.233$; $\beta = 0.274$, $SE = 0.083$, $p < 0.001$). As shown in Figure 5 (Bottom Left), just like in Experiment 1, participants were significantly more likely to detect Goal changes than Source changes when probed with a Same-Different memory task. Therefore, this asymmetry held after participants linguistically encoded the events as well.

Conversely, the model for the Forced-Choice task did not reveal an effect of Change Type (Mean Goal = 0.809, SD = 0.134, Mean Source = 0.843, SD = 0.121, $p = 0.892$). As shown in Figure 5 (Bottom-Right), the memory of Sources and Goals was similar when probed by the Forced-Choice task. Mean proportions of choosing each of the four alternatives are provided in the Appendix. To address the concern that such a pattern is a mere byproduct of a ceiling effect in the Forced-Choice task, we performed the same median-split analysis as in Experiment 1 (see Section 2.2) for Experiment 2. As shown in Figure 5 (Row 2), both the higher-performing participants and the lower-performing participants showed a Goal bias in the Same-Different task and no bias in the Forced-Choice task, suggesting that this pattern cannot be reduced to a difference in baseline performance.

Additionally, a main effect of Source Mention also surfaced in models for both tasks: participants performed better on events for which their description earlier included Source (Same-Different: $\beta = 0.287$, SE = 0.104, $p = 0.006$; Forced-Choice: $\beta = 0.419$, SE = 0.102, $p < 0.001$). However, importantly, the interaction between Source Mention and Change Type was not significant in the Same-Different task ($p = 0.570$), but was significant in the Forced Choice task ($\beta = -0.200$, SE = 0.093, $p = 0.031$). That is to say, the discrepancy between the memory of Source and Goal did not seem to be modulated by whether Source was mentioned in the Same-Different task (Figure 7 Top Row), but was so in the Forced-Choice task. Upon further inspection, the significant interaction in the Forced Choice task was likely driven by the different performance on Source: in cases where Source was mentioned, Source memory was in turn much better (Figure 7 Bottom Row). Post hoc pairwise comparisons through the emmeans package confirmed that Source performance was significantly improved when Source was mentioned ($\beta = 1.238$, SE=0.267, $p < .0001$), but Goal performance was not ($\beta = 0.439$, SE=0.283, $p=0.408$).

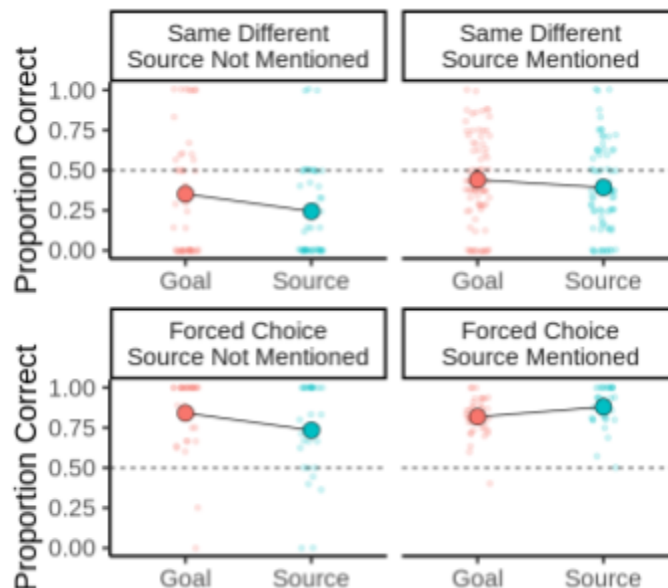


Figure 7: Participants' proportion of correct Goal/Source responses during the memory test on trials that elicited Source mention during description vs. trials that did not in Experiment 2.

3.3 Discussion

Experiment 2 replicated the main finding from Experiment 1 but in a context where participants were actively planning a linguistic description for the events during encoding: the Goal bias in memory was only found when probed with a Same-Different task but not a Forced-Choice task. Whether participants mentioned Source (a more peripheral event component) directly predicted memory accuracy of the event in general in both tasks. This suggests that description planning upon first exposure strengthens event encoding.

However, separate analyses for each task suggest that the effect of linguistic encoding on memory did surface differently in the two tasks: while the Goal bias was not modulated by Source Mention in the Same-Different task, Source Mention brought about a clearer boost for Source and not for Goal in the Forced-Choice task. Memory probed with the Forced-Choice task seemed to more straightforwardly reflect the specific heightened attention on Source during encoding. In other words, the boost of Source brought about by linguistic involvement only surfaced in the more aided retrieval condition.

4. Experiment 3

Experiment 3 asked whether the pattern from Experiment 2 would be replicated when participants' memory of Goals and Sources was tested after receiving, instead of producing, linguistic event descriptions.

4.1 Methods

4.1.1 Participants

A total of 122 native speakers of American English recruited from the University of Pennsylvania subject pool and Prolific participated in the experiment. The University subjects received course credit, and the Prolific subjects were compensated at a rate of \$6.5/hour.

4.1.2 Materials

The visual stimuli were exactly the same as the previous experiments. In order to probe how processing of a linguistic description interacts with the mental representation of these events, we recorded a female speaker reading one-sentence descriptions for the events in English. For each critical event, a Goal-only description and a Source-only description were made. The Source-only sentences all had the verb "came", and the Goal-only sentences all had the verb "went." For example, the Source-only description for our sample event would be "The squirrel came from the mailbox" and the Goal-only description would be "The squirrel went to the trash can." For each critical event, half of the participants heard the Source-only description, and the other half heard the Goal-only description. For each participant, half of the critical events had Source-only descriptions and the other half had Goal-only descriptions. The description of filler events remained the same across lists. Each description lasted for about 2 seconds.

4.1.3 Procedure

Experiment 3 was procedurally identical to Experiment 1 and 2, except for one change: during exposure, instead of passively viewing an event (Experiment 1) or typing a description (Experiment 2), participants heard a sentence describing the event that they just saw after the beep. We randomly assigned 82 participants to the Same-Different task and 40 participants to the Forced-Choice task.

4.1.4 Analysis and Predictions

To analyze the memory data, we ran similar logistic mixed-effect models as in Experiment 1 and 2. First, in the primary model, we predicted the correct response with Change Type (Source vs Goal, sum coded), Task (Same-Different vs Forced-Choice, sum coded), and their interaction as fixed effects. The final model included by-subject and by-item random intercepts, by-subject random slopes for Change Type as well as by-item random slopes for Change Type and Task. Similarly, we further built separate models for the two tasks. To probe the effect of hearing a linguistic description on memory, we added Mention - whether the changed role was mentioned in the linguistic description of this event, as a predictor in the models for each task. Therefore, both models included Change Type, Mention (Mentioned vs Not mentioned, sum coded) and their interaction as fixed effects and by-subject and by-item random intercepts and by-subject random slope of Change Type.

Three predictions were of interest. First, as before, we expected that the Source-Goal asymmetry would surface in the Same-Different task and not in the Forced-Choice task. This should surface as a significant interaction between Change Type and Task in the main model. Change Type would only be a significant predictor in the model for the Same-Different task. Second, we expected to find a main effect of Mention in both tasks, since people may have better memory about the object mentioned in the linguistic prompt due to increased depth of processing. Lastly, given that Source-only descriptions are unusual in listeners' linguistic and perceptual experience (Lakusta & Landau, 2005; Fisher et al., 1994), participants might find them especially salient. And since unexpected information is remembered better (Graesser, 2013; Shapiro & Fox, 2002), we also predicted an interaction between Change Type and Mention in both tasks: the boosting effect of Mention for Source should be stronger than that of Goal due to the salience of a Source-only construal.

4.2 Results

Just like in Experiment 1 and 2, our main model revealed a significant main effect of Task ($\beta = -0.702$, $SE = 0.088$, $p < 0.001$), suggesting higher accuracy in the Forced-Choice task in general. Importantly, our first prediction was borne out: there was a significant interaction between Change Type and Task ($\beta = 0.107$, $SE = 0.048$, $p = 0.025$). We further explored the interaction with a model for each task. As shown in Figure 8 (left), participants were significantly more likely to detect Goal than Source changes in motion events in the Same-Different task (Mean Goal = 0.372, $SD = 0.197$; Mean Source = 0.317, $SD = 0.201$; $\beta = 0.149$, $SE = 0.067$, $p =$

0.027). However, as shown in Figure 8 (right), memory of Source and Goal was equivalent in the Forced-Choice task (Mean Goal = 0.636, SD = 0.175; Mean Source = 0.666, SD = 0.148, $p = 0.396$). See Appendix for the mean proportions of choosing each of the four alternatives. As in Experiment 1 and 2, the median-split analysis (Section 2.2) suggested that the disappearance of the asymmetry in the Forced-Choice task cannot be explained by a ceiling effect (Figure 5, Row 3).

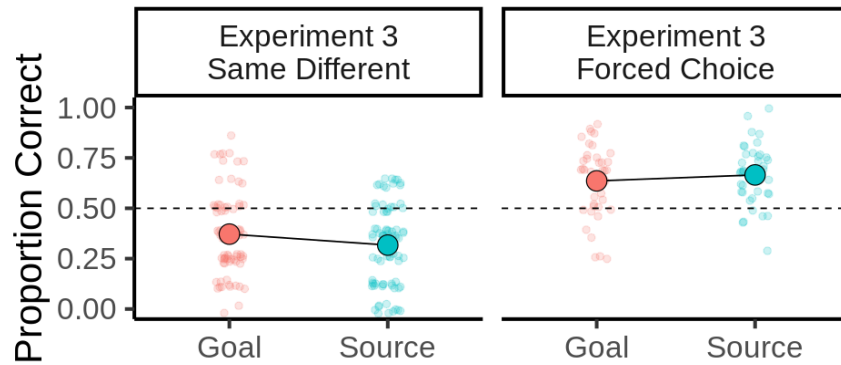


Figure 8: Participants' mean proportion of correct Goal/Source response at the memory test in Experiment 3.

However, our two other hypotheses were not borne out. We observed no main effect of Mention (Same-Different: $p = 0.935$; Forced-Choice: $p = 0.672$) in either task: participants did not necessarily remember the role that was mentioned in the description that they heard better than the non-mentioned role. Specifically, in Figure 9, we plotted the memory performance on events with Goal-only descriptions and those with Source-only descriptions separately in both tasks. As shown in the figure, in the Same-Different task, participants had a slight Goal bias regardless of whether they heard a Goal-only or a Source-only Description, while in the Forced-Choice task, there was no memory difference between Sources and Goals regardless of descriptions. As is also clear from Figure 9, and unlike our expectations, there was not a significant interaction between Mention and Change Type (Same-Different: $p = 0.965$; Forced-Choice: $p = 0.762$): hearing a Source-only description did not particularly boost participants' memory of Source in either condition.

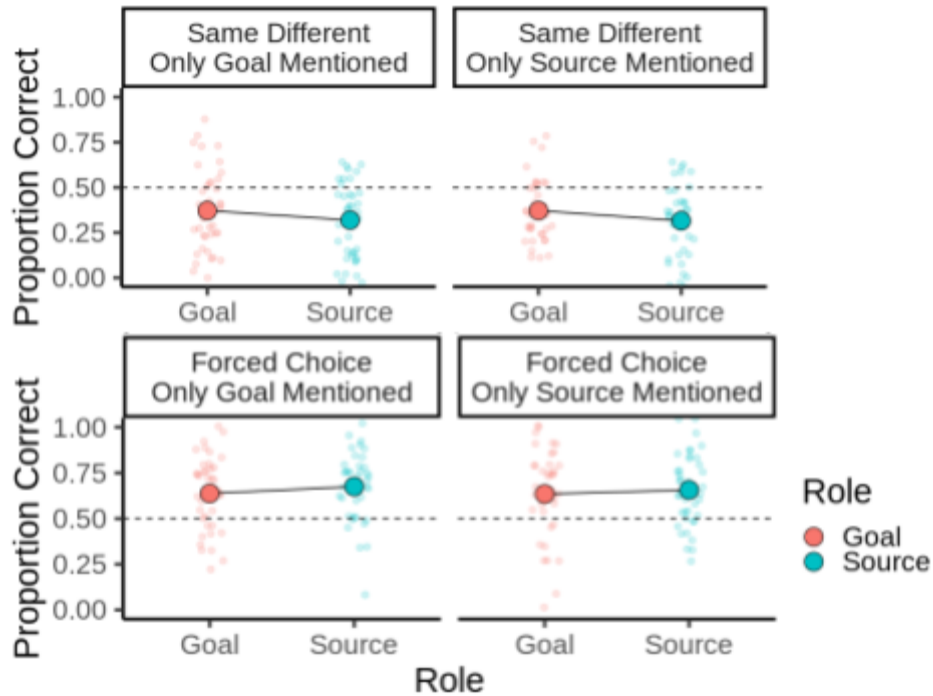


Figure 9: Participants' mean proportion of correct Goal/Source responses when they were exposed to Goal-only vs. Source-only audio prompts in Experiment 3.

4.3 Discussion

Experiment 3 replicated the finding that the Goal bias in memory was only found when probed with a Same-Different task but not a Forced-Choice task: this finding held after participants viewed events and heard linguistic description of them. The exclusive surfacing of the bias suggests that the Goal bias in memory is more malleable than previously thought: despite weaker encoding of Source, the visual access to the exact match in the Forced-Choice task could “save” Source and pull it to the same level of recognition as Goal.

We did not find an effect of exposure to different linguistic descriptions of the same event on memory, possibly due to limitations of our current design. In order to make sure that participants' non-linguistic processing of the events was not disrupted and was comparable to that in Experiment 1 and 2, we played the audio clips of linguistic descriptions after the video ended. By the time participants heard the description, they might have already formed their own construal of the event and might not revise their construal based on the linguistic prompts. Indeed, other studies in which a linguistic prompt was presented prior to the visual stimuli (e.g., Skordos et al., 2020; Fausey & Boroditsky, 2010) did find a significant effect of the prompt on memory of different event components.

5. Experiment 4

Throughout our experiments so far, we have consistently observed that the Source-Goal memory asymmetry was task-specific and have attributed this finding to the presence of retrieval cues in the Forced-Choice task. Could there be an alternative explanation? One might argue that the bias against Source and in favor of Goal in the Same-Different task is not a bias in representation (since people did equally well on both components in the Forced-Choice task) but instead reflects a strategy associated with this task. Put simply, perhaps Sources and Goals are encoded equally well in memory, but when faced with a Same-Different test, participants do not expect to be tested on a possible Source change, and as a result pay less attention to the Source (as compared to the Goal) when making a Same-Different judgment.

Recent findings showing that pragmatic factors modulate the Goal bias in linguistic production lend support to this possibility (e.g., Do et al., 2020). Do and colleagues argued that part of the reason why speakers leave Sources out of their descriptions is that they believe that it is less necessary to mention them to their interlocutors. They found that speakers are indeed more likely to mention Sources when Sources are not in the common ground of the conversation. Similarly, here, it is possible that the Source disadvantage does not result from worse encoding of Source than Goal, but rather, from people's tendency to prioritize Goal (or overlook Source) when searching for a difference in the Same-Different task where every aspect of the event could potentially differ from the original and lead to a "different" decision. In other words, a top-down computation of what is likely to be targeted by the current cognitive manipulation might be responsible for the asymmetry in the Same-Different task. Admittedly, previous demonstrations of the Goal advantage in memory even in motion events with a single landmark (e.g., Papafragou 2010, Experiment 2) render this purely expectation-based account for the difference across test formats less likely. However, it is still worth explicitly testing whether test-expectations about Goals and Sources modulate the asymmetry in the Same-Different task. We next carried out a pair of experiments to do just that.

5.1 Experiment 4a

In Experiment 4a, to give Goals and Sources an equal chance to be considered during the Same-Different decision, we circled the event component that changed and asked specifically whether that element differed. The expectation-based account predicts that the Source disadvantage should be eliminated or mitigated under the current manipulation, because it was unlikely that people would still overlook Source. However, if Source is truly less robustly represented, such a disadvantage should nonetheless continue to surface.

5.1.1 Methods

5.1.1.1 Participants

Seventy-eight native speakers of American English recruited on Prolific participated for compensation at a rate of \$8/hour.

5.1.1.2 Materials

Exposure materials were the same as in Experiment 1 (Same-Different task). However, in order to clearly indicate at test which particular element should be scrutinized for change, we added a red circle around the changed Goal or the changed Source in the previous test stimuli (Figure 10).

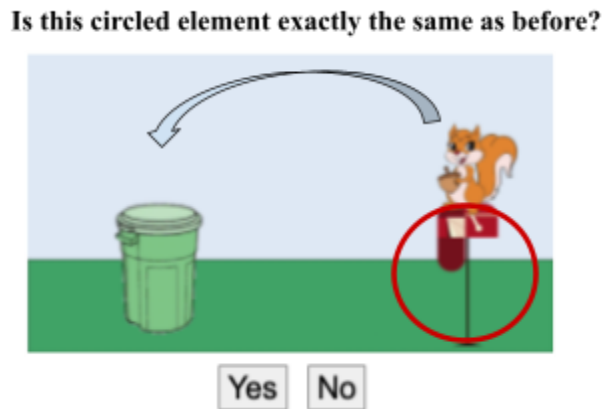


Figure 10: A Sample Same-Different Memory Test Trial in Experiment 4a.

5.1.1.3 Procedure

The procedure was identical to Experiment 1 (Same-Different task) except for one change: at test, instead of being asked to make a general judgment about whether they had seen the test event, people were asked to judge specifically whether the circled element was exactly the same as before (Figure 10). We also included in the test instructions a new event as an example.

5.1.1.4 Analysis

In order to test whether the Goal bias persisted when the change component was highlighted, we built a logistic mixed-effect model, predicting whether the response on a critical trial was correct with a fixed effect of Change Type (Goal vs. Source, sum coded). In order to directly compare with Experiment 1 where participants did not see the critical component being circled out, we also included Experiment (Experiment 1:Same-Different vs. Experiment 4a, sum coded) and the interaction term between Change Type and Experiment as a fixed effect. By-subject and by-item random slopes of Change Type and random intercepts were included as random effects. Participants who always selected “Yes” were excluded from analysis ($n = 1$).

5.1.2 Results and Discussion

We found that participants were still significantly more likely to detect a Goal change than a Source change in a Same-Different task even when it was made explicit where the potential change could be (Mean Goal = 0.572, SD = 0.222, Mean Source = 0.484, SD = 0.239, $\beta = 0.0271$, SE = 0.064, $p < 0.001$) (See Figure 10 right). This time, the Source disadvantage could not be attributed to inattention to Source arising from participant expectations because

participants were explicitly told to focus on the circled element (i.e., the Source on any Source-change trial).

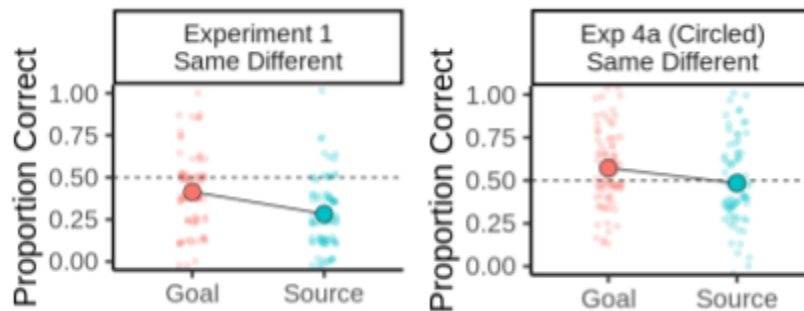


Figure 11: Participants' mean proportion of correct Goal and Source responses at the memory test in Experiment 1 Same-Different task (reproduced from Figure 3 left) and in Experiment 4a.

Furthermore, Experiment was a significant predictor of memory performance ($\beta = 0.441$, $SE = 0.071$, $p < 0.001$). It is unsurprising that performance was better in Experiment 4a than its earlier version in Experiment 1, since in the more recent version the only possible locus of difference was highlighted.

Is it possible that circling the Source might have mitigated (though not fully eliminated) the Source disadvantage by giving Source a better chance to be examined? The answer to this question seems to be no. In particular, we find no evidence suggesting that the extent of the Goal bias was any different under the two tasks in Experiment 4a and the Same-Different task in Experiment 1, since there was no significant interaction between Experiment and Change Type ($\beta = -0.066$, $SE = 0.048$, $p = 0.173$).

5.2 Experiment 4b

In Experiment 4b, to eliminate participants' potential tendency to examine the Goal first at test because of expectations about what was being tested, we removed the landmarks' identity as Goals or Sources by using static images of the motion event as test stimuli. If the Source disadvantage in the Same-Different task indicates less robust representation of Source (that is hidden by the Forced-Choice tasks), the disadvantage should still surface even when people are tested with a static image (since the encoding phase was kept the same). Alternatively, if the Source is in fact well represented, and the Source disadvantage results from the priority given to Goal when visually examining a motion event for a Same-Different judgment, participants should now lack a reason to prefer to examine either landmark in the static picture, and the Source disadvantage would disappear.

5.2.1 Methods

5.2.1.1 Participants

Eighty native speakers of American English recruited from the University of Pennsylvania subject pool participated for course credit.

5.2.1.2 Materials

Participants saw the exact same visual stimuli as in all previous experiments during the exposure phase. However, we removed the moving figure from the previous test stimuli. The resulting new test stimuli were static images consisting of only the background of the motion events, with a change in either of the two landmarks (see Figure 12 for an example).

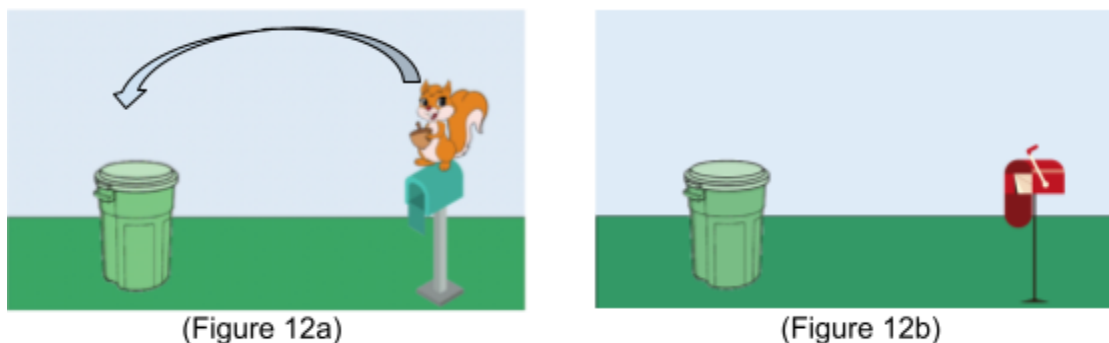


Figure 12. (a) Sample first-frame of the event “the squirrel went from the mailbox to the trash can” at exposure in Experiment 4b. (b) Sample test trial of this event involving Source change.

5.2.1.3 Procedure

The procedure of Experiment 4b was identical to the Same-Different task in Experiment 1 except for one change. In the test phase, instead of being asked whether they had seen this event before, participants were asked whether they had seen this background before.

5.2.1.4 Analysis

In order to test whether the Goal bias persists when probed with event background images, we built a logistic mixed-effect model, predicting whether the response on a critical trial was correct with a fixed effect of Change Type (Goal vs. Source, sum coded). In order to directly compare with Experiment 1a where participants were tested with videos, we also included Experiment (Experiment 1 vs. Experiment 4b, sum coded) and the interaction term between Change Type and Experiment as a fixed effect. By-subject and by-item random slopes of Change Type and random intercepts were included as random effects. Participants who always selected “Yes” were excluded from analysis ($n = 1$).

5.2.2 Results and Discussion

Just as in the Same-Different task in Experiment 1, we found that participants were still significantly more likely to detect a change to a (previous) Goal than a (previous) Source ($\beta = 0.279$, $SE = 0.051$, $p < 0.001$) (See Figure 13 right). Since the attentional preference to Goal

should be removed now, the current disadvantage of Source cannot be solely explained by the expectation-based account.

Participants performed slightly better in Experiment 4b than in the original Experiment 1 (Mean Goal = 0.465, SD = 0.201, Mean Source = 0.369, SD= 0.207; $\beta = 0.175$, SE = 0.065, $p = 0.007$), likely due to the fact that they couldn't be distracted in Experiment 4b by considering whether the Figure was different. However, there's no significant interaction between Experiment and Change Type ($\beta = 0.051$, SE = 0.048, $p = 0.287$), suggesting that the extent of the Goal bias was not different compared to Experiment 1 with videos as test stimuli. Not only did the removal of the motion trajectory at test fail to eliminate the Goal bias, it also failed to attenuate the bias.

In sum, the Source disadvantage here indicates weaker representation of the Source and in turn supports our original account of why the Same-Different and Forced-Choice tasks produce different results.

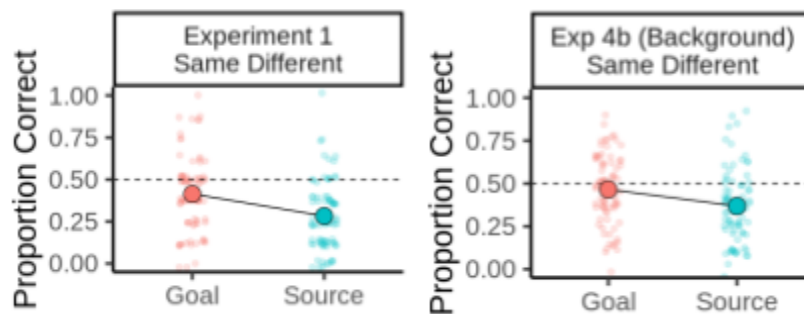


Figure 13: Participants' mean proportion of correct Goal and Source response at memory test in Experiment 1 Same-Different task (reproduced from Figure 3 left) and in Experiment 4b.

6. General Discussion

6.1 Revisiting the Source Disadvantage in Memory

Prior research has accumulated robust evidence showing that Sources and Goals of motion events are not equally likely to be mentioned in event descriptions and are not remembered equally accurately (Papafragou, 2010; Lakusta & Landau 2012; Regier & Zheng, 2007; Regier, 1996; Do et al., 2020). Many commentators have taken this parallel Source-Goal asymmetry in language and memory as strong evidence for finely articulated event architecture in which some event components are more salient compared to others. The current study revisited the nature of the fragility of Source representation. In a striking finding, across a series of experiments, the Source-Goal asymmetry appeared in a Same-Different memory task, a task that was ubiquitously used in previous explorations of this topic, but *not* in a Forced-Choice task, an equally well-established probe for recognition memory.

In the memory literature, similar task discrepancies are not uncommon and are often used to refine our understanding of human memory (e.g., Deffenbacher, Leu & Brown, 1981; Freed, Corkin & Cohen, 1987; Aggleton & Shaw, 1996). The important question, in the current context,

is: what should we take the conflicting results to mean? Is the verdict that we should trust the Same-Different task, and disregard the Forced-Choice task, which is just too insensitive to reflect an underlying conceptual difference? Or should we trust the Forced-Choice task and conclude that there is in fact no conceptual difference between Sources and Goals, but the disadvantage for Source in the Same-Different task stems from specific test demands? Our answer is neither. What provides a negative answer to the first question is the fact that the asymmetry surfaced in the Same-Different task and disappeared in the Forced-Choice task across all baseline performance levels and that performance in the Forced-Choice task was not at ceiling (see Figure 5). Our results in Experiment 4 speak against the second possibility: explicit manipulation of expectations about whether Source would be the target of change did not seem to affect the extent of the Source disadvantage at all, further suggesting that the asymmetry reflected in the Same-Different task is likely to reflect true disparity of the robustness of Sources vs. Goals in non-linguistic conceptual representation.

Such a striking difference across paradigms contributes further to our understanding of the nature of the Source-Goal asymmetry. Taken together, our results suggest that there is a true encoding asymmetry between Source and Goal, as reflected in the Same-Different task, but the presence of the exact target as a retrieval cue at test in the Forced-choice task reinstates all components of the event, thus leading to the recovery of the Source information. This is a novel contribution: although prior research has long noted that Sources exhibit a “representational deficit” (Papafragou, 2010, pg. 1084), the nature of that deficit has not been specified. To the best of our knowledge, we made the first specific commitment about what the Source disadvantage is (or more precisely is not). The fragility of Source does not index an omission at encoding, an irreparable loss of information, or even a loss of granularity but rather suggests weaker overall robustness. In fact, the Source information at the finest granularity that has been examined in this literature (a within category change of one object to a different one of the same kind) can be “reinstated”, to the same extent as Goals, with retrieval aids.

This interpretation is consistent with the cognitively-rooted, yet dynamic nature of the Source-Goal asymmetry, which has been pointed out by many commentators (Lakusta & Landau, 2012; Papafragou, 2010). Previously, it has been shown that many factors internal to the motion events might modulate the Source-Goal asymmetry: the Goal bias in memory disappears for events when the moving Figure is inanimate (e.g., a pen rolling from an eyeglass case onto a book: Lakusta & Landau, 2012, Experiment 3; Lakusta & Carey, 2015) or lacks intentionality (Lakusta & Landau, 2012, Experiment 2), or when the Source might be the cause of the motion (Lakusta et al., 2016). We add to the list that external factors, such as the retrieval conditions of the memory event, can also modulate whether a Goal bias emerges. If asked to retrieve the event in full view of the target, participants are no more prone to choosing the wrong Source than they are to choosing the wrong Goal.

Many questions are still left open about the specific mechanisms through which the two tasks present different opportunities for a difference in conceptual salience to surface. We tentatively propose that the current results are reminiscent of the asymmetrical effect of serial context on retrieval. We know that, in free recall tasks, participants are more likely to recall an item after

successfully retrieving another item in a nearby serial position in the studied list (Kahana, 1996; Howard & Kahana, 2002). Furthermore, such a facilitation effect is asymmetrical such that the successful retrieval of item in position i in the study list has a stronger facilitation effect on item $i+1$ than on item $i-1$. In other words, associations in a forward direction are stronger than those in a backward direction. Going back to our current context, our stimuli were clips of simple motion events involving a Figure moving from a Source to a Goal. On critical trials within a Same-Different task, participants either saw an event with a correct Source and a changed Goal or an event with a correct Goal and a changed Source. Since participants' attention follows the motion trajectory (Trueswell & Papafragou, 2010; Papafragou, Hulbert & Trueswell, 2008), they necessarily examine and encode the event components following the order of Source-Goal. Therefore, according to the asymmetric facilitation effect for forward association, a correct Source is predicted to be a more helpful retrieval cue for the original Goal than a correct Goal to the original Source. This is indeed consistent with our findings in the Same-Different task. On the other hand, within a Forced-choice task, because the retrieval conditions contained an exact match, the temporal context of encoding might have been rendered a less relevant cue. So far, to the best of our knowledge, the asymmetry between forward and backward association has not been investigated for recognition memory but only for recall. Therefore, this connection remains speculative. Future research should investigate further whether such temporal embedding of Sources and Goals contributes to the Source-Goal asymmetry.

How do the current findings generalize to the processing of realistic events? The fact that the Source-Goal asymmetry has been demonstrated with a variety of stimuli ranging from abstract geometric representations (e.g., Tatone et al., 2015) to fully naturalistic events (e.g., videos of Olympic figure skating routines, Levine, Hirsh-Pasek, Pace & Golinkoff, 2017) gives us a basis for believing that the effect we observed here reveals a property of abstract event structure. Nonetheless, we acknowledge the need to investigate the questions about fragility and robustness of event representation in complex dynamic natural scenes. Future investigations should also examine how such an asymmetry in event representation is formed in real-time. Very recent work suggests that the Goal bias emerges early in visual attention during real-time event apprehension and links these real-time signatures with off-line measures in language and memory (Chen, Trueswell & Papafragou, 2023).

6.2 Linguistic and Non-linguistic Encoding of Motion Events

Our results contribute to the understanding of the interaction between linguistic encoding and non-linguistic conceptual encoding of motion events. Recall that the Goal bias in memory emerges in the Same-Different task, but not in the Forced-Choice task, regardless of the type of linguistic task participants perform during and/or right after encoding (Experiments 1-3). This pattern provides direct evidence for the conceptual basis of the Source-Goal asymmetry independent from language-internal factors.

Even though many studies documented a parallel between a Source's omission in linguistic description and its less accurate memory, the current study was the first to directly test whether Sources that are mentioned are remembered better than those that remain unmentioned and

whether such mentioned Sources are still at disadvantage compared to Goals. Experiment 2 spoke directly to this point. With both memory tasks, we confirmed that Sources were indeed remembered better when mentioned. However, the mention of Source was found to modulate the extent of the Source-Goal asymmetry in a Forced-Choice task but not in a Same-Different task. Reinforcement of Sources in memory through linguistic encoding more straightforwardly surfaced when the retrieval conditions involved more memory support.

Similarly, we were among the first to examine whether the comprehension of a linguistic description of the event affects non-linguistic encoding of Sources and Goals (Experiment 3). We found that processing Source-only and Goal-only descriptions did not shift people's memory: even when exposed to a Source-only description of the event, people still showed a Goal bias in memory in the Same-Different task. This pattern could indicate that the phenomenon of having better memory for unexpected information that has been attested in the event memory literature (Graesser, 2013; Shapiro & Fox, 2002) does not necessarily generalize to unexpected perspectives. Admittedly, the lack of an effect of comprehension on motion memory might be attributable to the fact that people heard the linguistic prompts after they had viewed the visual stimuli. We hope to examine this question again in the future by inserting the prompts prior to the presentation of visual stimuli, a manipulation that has the sensitivity to capture effects of linguistic prompts on memory (Skordos et al., 2020).

6.3 Implications for Event Cognition and Language

Taken most broadly, our current findings bear on event cognition and its interface with language by delineating what "fragility" of an event component might mean in the context of the structured unit of the event as a whole. We show that fragile event components can nonetheless be encoded rather than entirely neglected despite being inconsistently mentioned in linguistic tasks - people monitored these event components even when they did not talk about them or notice their change. Not only is this evidence that language is not an exhaustive representation of cognition, it also suggests that memory performance from one test should not be taken as an exhaustive representation of cognition either.

These results invite a re-examination of other event components that have traditionally been considered less salient in both cognition and language. One particularly interesting case might be Instrument, the means used to perform an action (e.g., a girl hit the tennis ball into the basket with a racket). As alluded to already, many experimental findings corroborate the idea that Instruments have a fragile status in event representation. Both adults and children frequently omit instruments in their speech (e.g., Brown & Dell, 1987; Lockridge & Brennan, 2002; Wilson et al., 2014; Grigoroglou & Papafragou, 2019a, 2019b). Moreover, adults as well as 4-5 year olds are less sensitive to changes to Instruments during a change-blindness task (Wilson et al., 2014; Ünal et al., 2021). Our results raise the question of how Instruments, or other less-robustly-represented event components, are encoded and stored in memory. Though it appears that our linguistic and non-linguistic cognitive representation captures events in the world in similar ways, a variety of factors, including the encoding and retrieval conditions that our current results reveal here, modulate how such representations manifest themselves.

Our current results about the flexibility in the representation of event components are also important for language acquisition. In the case of the Source-Goal asymmetry, the acquisition of spatial language reveals a strong Goal bias: young children use significantly more Goal-marking case suffixes compared to Source-marking ones in their utterances (Pleth, 1998). They are also more likely to interpret a transfer verb to mean taking something to it (Goal reading) rather than taking something from it (Source reading) (e.g., take “weed the garden” to mean bringing weeds to the garden instead of removing weeds from the garden, see Papafragou, 2010; Srinivasan & Barner, 2013). Nevertheless, children learn rich Source language (Clark & Carpenter, 1989). The type of representation of Source that our results reveal here is in all likelihood required for the observed pattern in language learning: although less salient event components are often backgrounded, they remain available and can be invoked under the appropriate circumstances.⁸

Lastly, the robust cross-task variation that we show here highlights a straightforward yet important methodological point: multiple means of measuring event memory are needed to provide a clearer picture of a memory deficit, especially when event memory is taken to index the conceptual representation of events. This point of caution has deep roots in the field of cognitive science and extends to many other phenomena at the language-cognition interface. Consider a long-standing debate on space. Several commentators have argued that the spatial language in one’s community influences spatial cognition (Brown & Levinson, 1993; Pederson et al., 1998). In one prominent paradigm, participants were first familiarized with an array of objects on a table; they were then asked to turn around 180 degrees and make an array of the same objects on another table “in the same way as before”. Dutch speakers, whose language uses an egocentric spatial system (e.g., “The house is to your left”), reorganized the objects with respect to their own left/right. Tzeltal Mayan speakers, whose language adopts a geocentric frame of reference (e.g., “The house is to the east”) predominantly reorganized the objects according to cardinal directions.

Later work showed that these memory patterns were task-dependent. When Tzeltal speakers were trained to solve unambiguous spatial rotation tasks (e.g., choose the card with the matching orientation) as opposed to the previous prompt that was more open to interpretation, they successfully adopted both geocentric and egocentric solutions to rotation problems (Li, Abarbanell, Gleitman & Papafragou, 2011; see also Li & Gleitman, 2002). It seems that Tzeltal speakers, due to their linguistic encoding habits, had a clear preference in their understanding of what it meant to make the array “the same as before”, but their spatial memory in a rotation task was flexible and task-dependent (but not limited as a result of their linguistic habits). We

⁸ We may consider an analogy with the developmental speech perception literature. At around 10–12 months, infants begin to attend closely to phonological distinctions in the surrounding language and cease to distinguish non-native contrasts that they once had been able to at 6–8 months (e.g., Werker & Tee, 1984). Initially, these results were interpreted as a loss of early “universal” sensitivity. However, later evidence that children and even adults can acquire the phonological contrasts in additional languages or that adults can distinguish non-native contrasts in certain test contexts (e.g., Best, 1994) has led more commentators to believe that the universal sensitivities may continue to remain present, just “functionally reorganized” depending on how necessary cognitive resources have to be allocated to the phonetic contrasts in the surrounding language in different stages in life or different tasks at hand.

conclude that, whether there is a conceptual difference across language groups, or, in our case, over two co-present event components, relies crucially on how the availability or accessibility of that representation is measured (Tulving & Pearlstone, 1966).

Beyond the domain of events, the current work reiterates the necessity of careful attention to the methods we use to probe conceptual representation, especially the different demands that might arise across contexts and tasks. The importance of cross-task convergence for the study of the homology between language and cognition applies to other domains, including visual categories (e.g., Gibson et al., 2017; Winawer et al., 2007; Gilbert, Regier, Kay & Ivry, 2006; Thierry et al., 2009), numerical representation (e.g., Feigenson, Dehaene & Spelke, 2004; Pica et al., 2004; Gordon 2004; Frank et al., 2008, 2012; Spaepen et al., 2011), time (e.g., Boroditsky, 2001; Chen, 2007; January & Kako, 2007; Tse & Altarriba, 2008; Boroditsky, Fuhrman, & McCormick, 2011) and spatial representation (e.g., Choi & Bowerman, 1991; McDonough, Choi & Mandler, 2003; Pruden, Levine & Huttenlocher, 2011), despite remaining disagreements about the nature of the underlying categories (for recent reviews, see Ünal & Papafragou, 2016; Lupyan, Rahman, Boroditsky & Clark, 2020; Landau, 2022). This work also adds onto the existing rich memory literature on how different retrieval strategies lead to different assessment of memory (Tulving & Pearlstone, 1966; Tulving & Thompson, 1973; Godden & Baddeley, 1975; McDaniel & Masson, 1985, Karpicke & Zoromb 2010; Johansson & Johansson, 2014; Weldon & Roediger, 1987; Bäuml & Trißl, 2022) and call for a better integration of what we know about memory into the study of language and event cognition.

Appendix

Table 1. Percentage of choosing the four options in Experiment 1 (Forced-Choice)

	Correct Source	Incorrect Source	
Correct Goal	49.38%	18.75%	68.13%
Incorrect Goal	18.91%	12.97%	31.88%
	68.29%	31.72%	

Table 2. Percentage of choosing the four options in Experiment 2 (Forced-Choice)

	Correct Source	Incorrect Source	
Correct Goal	67.95%	12.98%	80.93%
Incorrect Goal	16.35%	2.72%	19.07%
	84.3%	15.7%	

Table 3. Percentage of choosing the four options in Experiment 3 (Forced-Choice)

	Correct Source	Incorrect Source	
Correct Goal	43.28%	20.31%	63.59%
Incorrect Goal	23.28%	13.13%	36.41%

	66.56%	33.44%	
--	--------	--------	--

References

- Aggleton, J. P., & Shaw, C. (1996). Amnesia and recognition memory: A re-analysis of psychometric data. *Neuropsychologia*, 34(1), 51-62.
- Archambault, A., O'Donnell, C., & Schyns, P. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science*, 10(3), 249–255.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
- Bayley, P. J., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/no recognition, forced-choice recognition, and the human hippocampus. *Journal of Cognitive Neuroscience*, 20(3), 505-512.
- Bäumli, K. H. T., & Trißl, L. (2022). Selective memory retrieval can revive forgotten memories. *Proceedings of the National Academy of Sciences*, 119(8), e2114377119.
- Boroditsky, L. (2001). Does language shape thought? English and Mandarin speakers' conception of time. *Cognitive Psychology*, 316-335.
- Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently?. *Cognition*, 118(1), 123-129.
- Boroditsky, L. (2018). Language and the construction of time through space. *Trends in Neurosciences*, 41(10), 651-653.
- Bowerman, M., De León, L., & Choi, S. (1995). Verbs, particles, and spatial semantics: Learning to talk about spatial actions in typologically different languages. In *Proceedings of the Twenty-seventh Annual Child Language Research Forum* (pp. 101-110). Stanford, CA: Center for the Study of Language and Information.
- Brown, P., & Levinson, S. C. (1993). "Uphill" and "downhill" in Tzeltal. *Journal of Linguistic Anthropology*, 3(1), 46-74.
- Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, 19(4), 441-472.
- Chen, J. Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition*, 104(2), 427-436.
- Chen, Y., Papafragou, A., & Trueswell, J. (2022). Encoding transfer of possession events. *Proceedings of the Linguistic Society of America*, 7(1), 5290.
- Chen, Y., Trueswell, J., & Papafragou, A. (2022). The Source-Goal asymmetry in motion events: Sources are robustly encoded in memory but overlooked at test. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).
- Chen, Y., Trueswell, J., & Papafragou, A. (In press). The Goal bias emerges early in motion event inspection and speech planning: Evidence from eye-movements. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41(1-3), 83-121.

- Croft, W. (2002). *Typology and universals*. Cambridge University Press.
- Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass*, 6(5), 310-329.
- Davidoff, J., Davies, I., & Roberson, D. (1999). Color categories in a Stone-Age tribe. *Nature*, 398, 203-4.
- Deffenbacher, K. A., Leu, J. R., & Brown, E. L. (1981). Memory for faces: Testing method, encoding strategy, and confidence. *American Journal of Psychology*, 94, 13–26.
- Dixon, R. M. (1994). *Ergativity*. Cambridge University Press.
- Do, M. L., Papafragou, A., & Trueswell, J. (2020). Cognitive and pragmatic factors in language production: Evidence from source-goal motion events. *Cognition*, 205, 104447.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307-314.
- Feist, M., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, 35(2), 283–296.
- Fillmore, C. J. (2006). Frame semantics. *Cognitive linguistics: Basic readings*, 34, 373-400.
- Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92, 333-375.
- Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, 108, 819-824.
- Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, 64(1-2), 74-92.
- Freed, D. M., Corkin, S., & Cohen, N. J. (1987). Forgetting in HM: A second look. *Neuropsychologia*, 25(3), 461-471.
- Freeman, N. H., Sinha, C. G., & Stedmon, J. A. (1981). The allative bias in three-year-olds is almost proof against task naturalness. *Journal of Child Language*, 8(2), 283-296.
- Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 114(40), 10785-10790.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the national academy of sciences*, 103(2), 489-494.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, 66(3), 325-331.
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, 306, 496-499.
- Graesser, A. C. (2013). *Prose comprehension beyond the word*. Springer Science & Business Media.
- Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin*, 66(3), 228.
- Grigoroglou, M., & Papafragou, A. (2019a). Children's (and adults') production adjustments to generic and particular listener needs. *Cognitive Science*, 43, e12790.
- Grigoroglou, M., & Papafragou, A. (2019b). Interactive contexts increase informativeness in children's referential communication. *Developmental Psychology*, 55, 951-966.

- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46(3), 269-299.
- Ihara, H., & Fujita, I. (2000). A cognitive approach to errors in case marking in Japanese agrammatism: The priority of the goal-ni over the source-kara. In *Constructions in cognitive linguistics* (p. 123). John Benjamins.
- Jackendoff, R. S. (1985). *Semantics and cognition* (Vol. 8). MIT Press.
- Jackendoff, R. (1996). How language helps us think. *Pragmatics & Cognition*, 4(1), 1-34.
- January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, 104, 417-426.
- Johanson, M., & Papafragou, A. (2014). What does children's spatial language reveal about spatial concepts? Evidence from the use of containment expressions. *Cognitive Science*, 38(5), 881-910.
- 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, 826-840.
- Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, 25(1), 236-242.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24(1), 103-109.
- Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, 62(3), 227-239.
- Khoe, W., Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia*, 38(10), 1333-1341.
- Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Frederick, C. M. (2002). Separating sensitivity from response bias: Implications of comparisons of yes-no and forced-choice tests for models and measures of recognition memory. *Journal of Experimental Psychology: General*, 131(2), 241.
- Lakusta, L., & Carey, S. (2015). Twelve-month-old infants' encoding of goal and source paths in 'agentive' and 'nonagentive' motion events. *Language Learning and Development*, 11, 152-175.
- Lakusta, L., & DiFrabrizio, S. (2017). And, the Winner Is...A Visual Preference for Endpoints over Starting Points in Infants' Motion Event Representations. *Infancy*, 23, 323-343.
- Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, 96, 1-33.
- Lakusta, L., & Landau, B. (2012). Language and Memory for Motion Events: Origins of the Asymmetry Between Source and Goal Paths. *Cognitive Science*, 36, 517-544.
- Lakusta, L., Muentener, P., Petrillo, L., Mullanaphy, N., & Muniz, L. (2016). Does Making Something Move Matter? Representations of Goals and Sources in Motion Events With Causal Sources. *Cognitive Science*, 41, 1-13.
- 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.

- 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, 179–197.
- Landau, B. (2022). Language and thought: The lexicon and beyond. In A. Papafragou, J. C. Trueswell & L. Gleitman (Eds.), *The Oxford handbook of the mental lexicon*. Oxford University Press.
- Landau, B., & Zukowski, A. (2018). Objects, motions, and paths: Spatial language in children with Williams syndrome. In *Williams Syndrome* (pp. 105-137). Routledge.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levine, D., Hirsh-Pasek, K., Pace, A., & Michnick Golinkoff, R. (2017). A goal bias in action: The boundaries adults perceive in events align with sites of actor intent. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(6), 916.
- Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, 9(3), 550-557.
- Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, 83(3), 265-294.
- Li, P., Abarbanell, L., Gleitman, L., & Papafragou, A. (2011). Spatial reasoning in Tenejapan Mayans. *Cognition*, 120(1), 33-53.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7(4), 560–572.
- Loftus, E. F. (1979). The malleability of memory. *American Scientist*, 67(3), 312–320.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4 (1), 19–31.
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 585–589.
- Lupyan, G. (2008). From chair to “chair:” A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, 137(2), 348–369.
- Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, 24(11), 930-944.
- Macmillan, N. A., & Creelman, C. D. (1994). *Detection theory: A user's guide*. Psychology Press.
- MacWhinney, B. (1977). Starting points. *Language*, 152-168.
- Markovskaya, E. (2006). Goal-source asymmetry and Russian spatial prefixes. *Nordlyd*, 33(2).
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(2), 371.
- McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, 46(3), 229-259.
- Nam, S. (2004). Goal and source: Their syntactic and semantic asymmetry. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 30, No. 1, pp. 304-317).
- Nikitina, T. (2009). Subcategorization pattern and lexical meaning of motion verbs: a study of the source/goal ambiguity.

- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1), 155-184.
- Papafragou, A. (2010). Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive Science*, 34, 1064–1092.
- Papafragou, A., & Grigoroğlu, M. (2019). The role of conceptualization during language production: evidence from event encoding. *Language, Cognition and Neuroscience*, 34(9), 1117-1128.
- Pederson, E., Danziger, E., Wilkins, D., Levinson, S., Kita, S., & Senft, G. (1998). Semantic typology and spatial conceptualization. *Language*, 74(3), 557-589.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigenous group. *Science*, 306, 499-503.
- Pinker, S. (2013). *Learnability and Cognition, new edition: The Acquisition of Argument Structure*. MIT Press.
- Pléh, C. (2010). On the importance of goals in child language. *Language Acquisition Across Linguistic and Cognitive Systems*, 52, 147.
- Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children’s spatial thinking: Does talk about the spatial world matter?. *Developmental Science*, 14(6), 1417-1430.
- Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA: MIT Press.
- Regier, T., & Zheng, M. (2007). Attention to endpoints: A cross-linguistic constraint on spatial meaning. *Cognitive Science*, 31, 705–719.
- Rohde, H., Kehler, A. and Elman, J. (2006). Event Structure and Discourse Coherence Biases in Pronoun Interpretation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 28, 697-702.
- Shapiro, M. A., & Fox, J. R. (2002). The role of typical and atypical events in story memory. *Human Communication Research*, 28(1), 109-135.
- Skordos, D., Bungler, A., Richards, C., Selimis, S., Trueswell, J., & Papafragou, A. (2020). Motion verbs and memory for motion events. *Cognitive Neuropsychology*, 37(5-6), 254-270.
- Spaepen, E., Coppola, M., Spelke, E. S., Carey, S. E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences*, 108(8), 3163-3168.
- Srinivasan, M., & Barner, D. (2013). The Amelia Bedelia effect: World knowledge and the goal bias in language acquisition. *Cognition*, 128(3), 431-450.
- Stefanowitsch, A., & Rohde, A. (2004). The goal bias in the encoding of motion events. In G. Radden & K.-U. Panther (Eds.), *Studies in linguistic motivation*. Berlin: Mouton de Gruyter
- Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. (2009). Unconscious effects of language-specific terminology on pre-attentive colour perception. *Proceedings of the National Academy of Sciences*, 106, 4567–4570.
- Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63(1), 64-82.

- Tse, C. S., & Altarriba, J. (2008). Evidence against linguistic relativity in Chinese and English: A case study of spatial and temporal metaphors. *Journal of Cognition and Culture*, 8(3-4), 335-357.
- Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5(4), 381-391.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352.
- Ünal, E., & Papafragou, A. (2016). Interactions between language and mental representations. *Language Learning*, 66(3), 554-580.
- Ünal, E., Richards, C., Trueswell, J. C., & 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), e13116.
- Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition*, 15(4), 269-280.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19), 7780-7785.
- Zheng, M. & Goldin-Meadow, S. (2002). Thought before language: how deaf and hearing children express motion events across cultures. *Cognition*, 85, 145-175.