

ARTICLE

Speaking but not gesturing predicts event memory: a cross-linguistic comparison

Marlijn ter Bekke^{1,2*} , Aslı Özyürek^{1,2,3} and Ercenur Ünal^{2,3,4} 

¹Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, the Netherlands;

²Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; ³Center for Language Studies, Radboud University, Nijmegen, the Netherlands; ⁴Ozyegin University, Istanbul, Turkey

*Corresponding author. Email: Marlijn.terBekke@donders.ru.nl

(Received 31 July 2020; Revised 08 February 2022; Accepted 09 February 2022)

Abstract

Every day people see, describe, and remember motion events. However, the relation between multimodal encoding of motion events in speech and gesture, and memory is not yet fully understood. Moreover, whether language typology modulates this relation remains to be tested. This study investigates whether the type of motion event information (path or manner) mentioned in speech and gesture predicts which information is remembered and whether this varies across speakers of typologically different languages. Dutch- and Turkish-speakers watched and described motion events and completed a surprise recognition memory task. For both Dutch- and Turkish-speakers, manner memory was at chance level. Participants who mentioned path in speech during encoding were more accurate at detecting changes to the path in the memory task. The relation between mentioning path in speech and path memory did not vary cross-linguistically. Finally, the co-speech gesture did not predict memory above mentioning path in speech. These findings suggest that how speakers describe a motion event in speech is more important than the typology of the speakers' native language in predicting motion event memory. The motion event videos are available for download for future research at <https://osf.io/p8cas/>.

Keywords: motion events; event cognition; event memory; cross-linguistic differences; co-speech gesture; multimodality; Turkish; Dutch

1. Introduction

Language is a powerful tool to describe and share the events we experience (Radvansky & Zacks, 2014). Even though our experience of events might be similar, how languages describe events with speech varies across languages. An important question for event cognition is whether the way language encodes events influences memory for the events described. Even though previous research has investigated this question in relation to broad typological differences between languages, it has not

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

explored whether the way speakers describe events within a language predicts memory. Furthermore, language is a multimodal phenomenon and in face-to-face interaction, information is communicated not only through speech, but also through bodily signals such as meaningful hand gestures (Kendou, 2004; McNeill, 2005). In event descriptions, hand gestures can depict semantic information related to what is expressed in speech (Özyürek, 2017), and gesture and speech form a tightly integrated system both in production (Kita & Özyürek, 2003) and comprehension (Kelly et al., 2010). More importantly, languages are known to differ not only in terms of lexical and syntactic encoding of event components, but also in the gestures that depict these components (Kita & Özyürek, 2003). However, it is not known whether and to what extent language-specific encodings in gestures across different languages influence memory beyond the differences we see in relation to speech. Therefore, in the current study, we combine cross-linguistic and multimodal approaches to investigate whether the speech and co-speech gestures that speakers use in event descriptions relate to their memory, and how this changes cross-linguistically.

Several scholars acknowledge that the information encoded in linguistic descriptions might be linked to memory (Gentner & Goldin-Meadow, 2003; Landau et al., 2010; Lupyan, 2008, 2012; Papafragou et al., 2002; Wolff & Holmes, 2011). More specifically, whether or not an event component is mentioned in speech can predict memory for that event component. This is because, according to some speech production models, if a speaker describes a certain event component, they tend to allocate more attention to this component, indicating that the speaker has this component as part of conceptualization of the event for speaking (Levelt, 1989; see also Papafragou et al., 2008). Thus, those event components mentioned in the linguistic description and thus have been attended to and included in event conceptualization may be more likely to be remembered.

One might also expect co-speech gestures to be related to memory for several reasons. First, gesture production itself may improve memory. This could be because the use of the hands during gesturing allows for visually motivated, iconic depictions of events that exhibit a high degree of resemblance between form (e.g., an iconic running gesture) and meaning (e.g., actual running; Perniss et al., 2010). This multimodal encoding that is iconic in gesture and arbitrary, abstract and categorical in speech could in turn create richer, stronger, or longer-lasting memory due to a dual (i.e., visual and verbal) encoding of the same information (Paivio, 1990). Second, gesture production is also considered to be part of event conceptualization for message preparation, as indicated by the fact that gestures package information in ways tightly linked to how the same information is linguistically encoded in the accompanying speech (Kita & Özyürek, 2003). Furthermore, gesture production is also linked to visual attention during message preparation and this link emerges after controlling for the effects found for speech production (Ünal et al., *under review*). Crucially, it is unknown whether gesture production enhances memory over and above speech production. This question is especially important given the disagreements about the relative benefits of abstract and categorical as opposed to analogue and iconic encodings (Lupyan, 2012) and in light of previous findings showing that use of categorical language can support spatial cognition (Dessalegn & Landau, 2008; Feist & Gentner, 2007; Gentner et al., 2013).

In this study, our goal is to test these proposals on the relation between speech and gesture production and memory in the domain of motion events and ask whether this relation is modulated by language typology. Before laying out the specific possibilities

we are going to evaluate, we describe cross-linguistic differences in the verbal and gestural encoding of motion and their possible relation to motion event memory.

1.1. Cross-linguistic variation in motion event encoding in speech and gesture

Two core components of intransitive motion events are the path that the motion follows (e.g., into the gazebo) and the manner of motion (e.g., running). Speakers of verb-framed languages (e.g., Turkish, Greek, Spanish, and Japanese) tend to encode path in the main verb and can optionally encode manner in adverbial phrases or subordinate verbs (see sentence (1) from Turkish; Talmy, 2000, see also papers in Bylund & Athanasopoulos, 2015 and Ibarretxe-Antuñano, 2017). For satellite-framed languages (e.g., Dutch, English, Russian, and Swedish), manner is typically encoded in the main verb and path in other structures, such as prepositional phrases (see sentence (2) from Dutch). One key difference between verb-framed and satellite-framed languages is that speakers of satellite-framed languages typically encode both path and manner in speech, while speakers of verb-framed languages are more likely to omit manner (Slobin, 2003).

- (1) Kadın (koş-arak) çardağ-a gir-(i)yor
 woman (run-CONN) gazebo-DAT enter-PRES
 Noun phrase (Verb) Noun phrase Verb
 Figure (Manner) Ground Path
 ‘The woman is entering the gazebo runningly.’

- (2) De vrouw rent het prieel in
 the woman run-PRES the gazebo into
 Noun phrase Verb Noun phrase Preposition
 Figure Manner Ground Path
 ‘The woman runs into the gazebo.’

Although these are the most typical and frequent linguistic patterns, this does not mean that speakers of verb-framed and satellite-framed languages exclusively use the above-mentioned constructions. For example, speakers of verb-framed languages can also use manner verbs (e.g., Greek: *treho* ‘run’) and speakers of satellite-framed languages can also use path verbs (e.g., English: *enter*; Papafragou et al., 2002, 2006). Yet how often they use them in a given description might vary according to the typology of the language. Furthermore, there are fewer manner verbs in verb-framed languages and hence the same verb can be used to describe different manners (e.g., *jump* to describe jumping, hopping, and skipping). Conversely, there are fewer path verbs in satellite-framed languages and hence the same verb can be used to describe different paths (e.g., *go* to describe movement towards and away from a landmark). In sum, motion event descriptions vary typologically across languages, while there is also within-language variation in terms of frequencies, types of verbs used, and the specificity with which these verbs are used to distinguish different manners and paths.

Motion event descriptions are often accompanied by iconic co-speech gestures (Kita et al., 2007). As with speech, gestures accompanying motion event descriptions differ both within and across languages (Kita & Özyürek, 2003). Specifically, speakers

(A) Gesture: Path only



Speech: Path + Manner
 bu sefer de bisiklet levhası olan bir
 yerden tekrar parmak ucunda yavaş
 yavaş yürüyor uzaklaşıyor oradan
 Manner Path

'she is walking on her toes slowly from
 a place with a bicycle sign. she is
 leaving that place'

Event: A woman tiptoes from a traffic
 sign

(B) Gesture: Manner only



Speech: Manner
 ay böyle çok huzurlu tropikal bir
 sahilde böyle yine hafif tempo
 koşuyor
 Manner

'she is running in a slow tempo on
 a peaceful tropical beach'

Event: A woman runs from a palm
 tree

(C) Gesture: Path + Manner



Speech: Path + Manner
 er staat een boom aan de zijkant en
 daar is ze naartoe aan het joggen
 Path Manner

'there is a tree on the side and she is
 jogging towards it'

Event: A woman runs to a tree

Fig. 1. Gestures can represent only path (A), only manner (B), or both manner and path (C). The gesture stroke occurred during the underlined speech.

of verb-framed languages (e.g., Turkish and Japanese) that encode path and manner in separate verbal clauses in speech (as in sentence (1)) tend to produce separate gestures that represent either only path (Fig. 1A) or only manner (Fig. 1B). By contrast, speakers of satellite-framed languages (e.g., English) that encode path and manner in a single clause in speech (as in sentence (2)) tend to conflate manner and path within a single gesture (Fig. 1C; Kita & Özyürek, 2003; Özçalışkan et al., 2016; Özyürek et al., 2005). Nevertheless, as in the case of speech, there are also deviations from the gesture patterns described above. For example, how speakers gesture is also affected by the information they express in speech: if speakers express path but not manner they also gesture more about path than manner (Özyürek et al., 2005), in line with speech and gesture as an integrated system. Even if speakers express both manner and path in speech, they might express only path in gesture (Akhavan et al., 2017; Chui, 2009; Gullberg et al., 2008; Mamus et al., 2021) as path can be considered as a core event component (Radvansky & Zacks, 2014). However, speakers may also express complementary information in speech and gesture: McNeill and Duncan (2000) report examples of Spanish-speakers producing gestures that conflate manner and path even if they only mention path in speech, although they do not report any quantitative data. Therefore, it is important to take into account co-speech gesture when investigating how differences in motion event descriptions relate to motion event memory.

1.2. The relation between motion event memory and speech and gesture

Given that motion event descriptions in speech and gesture vary both across and within languages, an important question concerns whether this variation has consequences for motion event memory. Most prior work that related motion event descriptions to memory has focused on speech, and found no cross-linguistic

differences in how speakers of verb-framed and satellite-framed languages remember manner and path after they watched and described motion events (Engemann et al., 2015; Gennari et al., 2002; Papafragou et al., 2008; Papafragou et al., 2002; but see Filipović, 2011 for differences using complex motion events).

Another line of work asked if encoding certain motion event components in event descriptions predicts memory for those components. Here, previous studies provide mixed evidence. When participants had to describe motion events by writing a single verb, those who used a path verb to describe a particular event later remembered this path better (Billman et al., 2000). A more recent study found that when speakers had to describe motion events by saying a single verb, those who produced a path verb were less accurate at remembering the manner regardless of whether the speaker's native language was verb-framed (Greek) or satellite-framed (English; Skordos et al., 2020). However, two other studies in which child and adult participants could freely describe events found that descriptions did not predict subsequent memory (Bunger et al., 2012; Papafragou et al., 2002). Thus, it is still unclear whether motion event descriptions predict motion event memory across typologically different languages.

Little is known about how gestures accompanying motion event descriptions relate to motion event memory. The most relevant evidence comes from the study of action events, which found that gesturing while describing action and motion events enhances event memory (Cook et al., 2010). Moreover, which action event information is encoded in gesture predicts which information is remembered (Koranda & MacDonald, 2015). Finally, compared to only reading descriptions of action events, reading descriptions, and performing these actions improves memory (see Cohen, 1989 for a review). While these results point to the importance of taking co-speech gestures into account, it remains unknown whether spontaneous co-speech gestures help memory in a domain where gestures may potentially depict information in a less embodied way. For example, path gestures that trace the trajectory of motion could be different than performing actions on objects. Furthermore, it remains to be seen if gestures also relate to motion event memory and whether this is influenced by cross-linguistic differences in gesture patterns described above.

1.3. *The present study*

The aim of the present study was to investigate how multimodal motion event descriptions (i.e., speech and co-speech gesture) relate to motion event memory and whether this relation varies within and across speakers of different languages. To this end, participants watched and described motion events in which a figure moved with a distinct manner and path. We used a surprise recognition task to measure memory of manner and path. We chose this measure to keep our methodology similar to previous cross-linguistic work on motion event memory. In order to examine within-language variation, we test motion event memory by taking into account how speakers have described those very same events, and specifically whether or not an event component is mentioned in speech and gesture. In order to examine cross-linguistic variation, we compare speakers of two typologically different languages that encode motion differently: Dutch (satellite-framed language) and Turkish (verb-framed language). Furthermore, in order to test how event encodings in speech predict memory, we zoom into those cases where participants

specifically encode that event component. For example, for the path we chose descriptions where participants specifically encoded the trajectory of motion with respect to a landmark with a verb, spatial noun, or pre-/postposition to ensure that path of motion is encoded in speech. Our multimodal approach to studying the link between descriptions and event memory is completely novel. Given previous work on cross-linguistic variation in motion event gestures, and that gesture has been linked to event memory, this is an important extension of previous work that focused on the relation between native language and memory, or between speech and memory.

In terms of speech and gesture production, we expected Dutch-speakers to mention manner in speech more often than Turkish-speakers, due to the optional encoding of manner in Turkish (Talmy, 2000). A similar pattern was expected for co-speech gesture, with Dutch-speakers gesturing more about manner than Turkish-speakers, and Turkish-speakers gesturing more about path than Dutch-speakers in line with the idea that gesture and speech form a tightly integrated system (Kita & Özyürek, 2003).

Regarding memory, if encoding information in descriptions benefits memory, then encoding a motion event component in speech should predict better memory for that component. Similarly, if the linguistic encoding of a motion event component in speech is accompanied by an iconic gesture depicting that component, then this should predict even stronger memory for that component. Finally, if the effect of speech and/or gesture production interacts with language typology, mentioning manner in speech and/or gesture should be linked to even better manner memory for Dutch-speakers than Turkish-speakers. Conversely, mentioning path in speech and/or gesture should be linked to even better path memory for Turkish-speakers than Dutch-speakers.

2. Method

The stimuli are available at the Open Science Framework Repository <https://osf.io/p8cas/>.

2.1. Participants

The sample consisted of 19 adult native speakers of Dutch (15 females, $M_{\text{age}} = 23$) and 22 adult native speakers of Turkish (16 females, $M_{\text{age}} = 21$). Dutch-speakers were recruited from the Max Planck Institute for Psycholinguistics participant database and received monetary compensation for their participation. Turkish-speakers were students at Özyeğin University in Istanbul and received course credit for their participation. Data from six additional participants were discarded due to experimenter error ($n = 2$), equipment error ($n = 1$), knowledge of Sign Language of the Netherlands ($n = 1$), and motion memory accuracy of more two SDs below the mean ($n = 1$). All participants provided written consent.

For Dutch-speakers, Dutch was their only native language. Around half ($n = 11$) of the Dutch participants knew verb-framed languages, mostly French or Spanish. They all learnt these languages after age 11, usually in high school, and used them never or rarely. They rated their speaking fluency in the verb-framed language as very bad ($n = 4$), bad ($n = 4$), mediocre ($n = 1$), or reasonable ($n = 2$). None of the participants rated themselves as fluent or very fluent.

For Turkish-speakers, Turkish was their only native language. Many ($n = 20$) of the Turkish-speakers knew satellite-framed languages, mostly English. The large majority learnt such languages in school after age 10. Most used these languages often. They rated their speaking fluency in the satellite-framed language as very bad ($n = 2$), bad ($n = 2$), mediocre ($n = 7$), or reasonable ($n = 9$). None of the participants rated themselves as fluent or very fluent.

2.2. Materials

In the study phase, the target events were 16 silent video clips that depicted a female actor moving in a certain manner, along a certain path with respect to a landmark object (e.g., a woman hopped to a cactus). Each clip (2,500 ms) was digitally created by combining four spontaneous manners of motion (run, hop, twirl, and tiptoe) with four paths (to, into, from, and out of). Manners of motion were filmed in a studio at Radboud University for the purpose of this study. The actors performed the manners of motion against a green background. The video clips were edited in Adobe Premiere Pro CC 2015. First, each clip was cut to last 2,500 ms. Then, the green background was removed from the video using the ultra-key feature. Next, motion paths were created by combining the moving actor with a landmark object. For *to* and *into* paths, the landmarks were placed near the final location of the actor's motion. For *into* paths, the actor entered the landmark. For *from* and *out of* paths, the landmarks were placed near the starting location. For *out of* paths, the actor exited the landmarks. The landmark objects were selected such that they were similarly familiar to Dutch- and Turkish-speakers. Finally, in order to create a scene, each manner-path combination was matched with a different background and floor, which could be inside or outside. The backgrounds were appropriate for the landmark, for example, for the palm tree, the background was a beach. A pilot study confirmed that the backgrounds were not so salient that speakers would only mention the backgrounds instead of the landmark objects.

Sixteen additional video clips (2,500 ms) depicting transitive events were used as fillers (e.g., a woman cutting an apple). They were filmed at the same studio as the motion events. Actors performed the actions on a gray table against the same green background. Video-clips were edited in Adobe Premiere Pro CC 2015. First, each clip was cut to last 2,500 ms. Then, the green background was removed and replaced with one of two backgrounds (a white brick wall, or a textured light pink wall). A list of all events can be found in Appendices A and B.

For the memory task, participants were shown 31 videos. Half of them were identical to the videos shown during the description task, and for the other half one aspect had been changed. Of the 16 motion events, 8 remained the same, 4 involved a manner-change, and 4 involved a path-change. The changed motion events were created in Adobe Premiere Pro. For manner-changes, the manner of motion changed, while the spatial relation between the agent in motion and the landmark remained the same. The location of the landmark object and the direction of motion (left-right or right-left) also remained the same. Manner-changes were created in the following way: *running* became *hopping*, *hopping* became *tiptoeing*, *tiptoeing* became *twirling*, and *twirling* became *running* (see Fig. 2A for an example of how *hopping* become *tiptoeing*). For path-changes, the spatial relation between figure in motion and the landmark changed, while manner of motion and location of the landmark

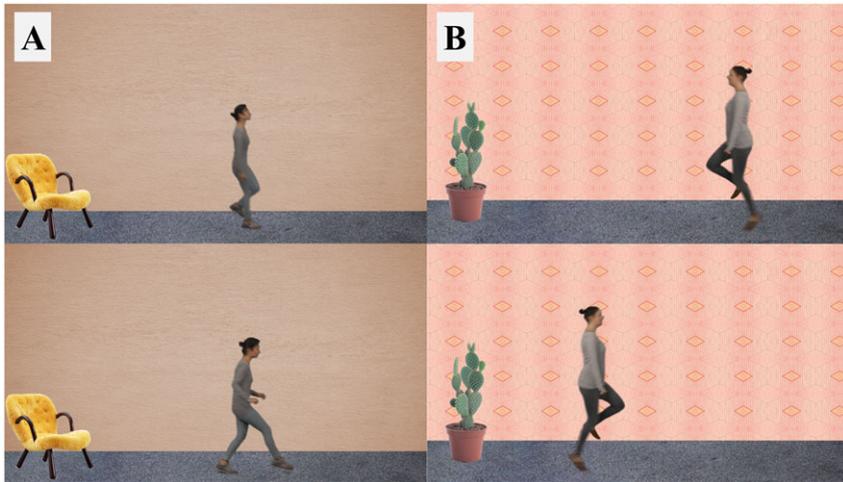


Fig. 2. Example of a manner-change (hop became tiptoe; panel A) and a path-change (to became from; panel B)

object on the screen remained the same. The direction of motion was always reversed for the path-changes. Path-changes were created in the following way: *into* became *out of*, *out of* became *into*, *to* became *from*, and *from* became *to* (see Fig. 2B for an example of how *to* became *from*). Within the eight motion event changes, the two different actors were counterbalanced across path, manner and type of change. Participants were also shown 15 filler events. Half of the fillers remained the same and half involved an object change (e.g., a woman cutting an apple changed to a woman cutting a lemon). These events with changed objects were filmed in the same way as the original transitive events. (Due to an error in the script, 15 fillers were presented instead of 16. For 19 participants (11 Turkish), the missing filler was an object change and for 22 participants (11 Turkish) it was a no-change item.)

2.3. Procedure

Each participant was tested in a quiet room at their university campus in their native language by a native speaker. They were tested together with a confederate addressee whom they believed to be a naïve participant. First, the participant and the addressee performed a short game together that served to familiarize the participant with the addressee and with using their hands. During the game, the participant had to describe four objects (shampoo, hammer, piano, mascara) without using a set of forbidden words, and the addressee had to guess the object. The participant was allowed to use other words from their native language, sounds, or their hands. The data from this task was neither recorded nor analyzed.

During the study phase, participants viewed 16 target and 16 filler events. Each trial consisted of a fixation screen (1,000 ms), followed by an event (2,500 ms), and a gray screen which prompted the participants to describe “what happened in the video” to the addressee. This addressee was present to create a more natural, communicative context. The addressee listened to the descriptions, supposedly in preparation for later questions. The addressee did not say anything meaningful but

was allowed to indicate that they had understood the description (e.g., by nodding). To initiate the next trial, the addressee clicked the computer mouse. The study phase was videotaped for later coding.

Directly after the study phase, the memory task was presented. This task was kept a surprise for the participants to prevent the possibility of the memory task affecting linguistic productions. In the memory task, participants viewed another set of events. For each event, they indicated by button press whether they had seen this exact video before (a green button for 'yes' a red button for 'no'). Since participants had to wait until the end of the video to respond, we did not use reaction time data and only focused on the accuracy of responses. During the study and memory phases, the events were presented to each participant in a different randomized order.

After the motion event memory task, participants performed two working memory tasks, to test if there was group-level variation in general working memory capacity on an independent measure (following Sakarias & Flecken, 2019). The Corsi block-tapping task measured visuospatial working memory (Corsi, 1972; Kessels et al., 2000). On the screen, nine blue blocks were distributed irregularly. One by one, some of these blocks turned red for a short time. Participants had to memorize which blocks turned red in which order, and repeated this sequence by clicking on the blocks in that order. Participants' Corsi-span was calculated as the longest sequence of blocks they reproduced correctly. The digit-span task measured verbal working memory (Wechsler, 1944). Participants were presented with a sequence of digits appearing one-by-one on the screen. They had to keep the sequence in memory and type it on the keyboard once the sequence ended. Participants' digit-span was calculated as the longest sequence of numbers they reproduced correctly.

2.4. Coding

For each motion event description, a native speaker of the relevant language coded the presence of path and manner information in speech and gesture using ELAN (Lausberg & Sloetjes, 2009). In speech, path or manner information was coded as specific, unspecific, or absent. Information was coded as unspecific if it did not disambiguate between the various paths or manners. Manner information was coded as specific if how the motion was performed was encoded with a manner verb (e.g., *rennen* 'running' – mostly in Dutch) or a manner verb subordinated to a path verb via a connective (e.g., *koşarak* 'run-Connective' – mostly in Turkish). Manner information was coded as unspecific if participants used the manner verbs 'to walk' or 'to run' when the manner was not walking or running. This is unspecific because is not clear which of the four manners it describes.

Path information was coded as specific if the change of location with respect to the landmark or the left-right axis was encoded with prepositions or spatial/directional nouns (e.g., *naar* 'to', *içine* 'inside') or path verbs (e.g., *gir* 'enter', *yaklaş* 'approach'). Path information was coded as unspecific if it did not indicate or imply the trajectory of the figure with respect to the landmark or the left-right axis. This included the use of the Turkish unspecific path verbs *ilerle* 'advance' or *git* 'go' because these could for example be used both to describe an *into* path and an *out of* path. Dutch-speakers did not use these unspecific path verbs. In Dutch, use of the word *weg* 'away' was coded as unspecific path information (see Supplementary Material for examples).

In gesture, manner information was coded as present if a gesture depicted the motion in a nonlinear way. Manner gestures could be in third-person perspective (as in Fig. 1C, where the inverted index and middle finger move across space to represent running legs from a third-person perspective). Manner gestures could also be a first-person enactment of the manner (as in Fig. 1B, where the speaker moves her arms as if running herself). Path information was coded as present if the speaker chose a body part to represent the figure (e.g., the index finger), and deliberately traced the change of location with this body part. Tracing could be in the lateral axis (with correct or incorrect direction) or in the sagittal axis (moving toward or away from the body). Points to the landmark location were not included as path gestures. A gesture could include one motion element (path-only as in Fig. 1A or manner-only as in Fig. 1B) or both elements conflated (Fig. 1C).

To obtain reliability, an additional native Dutch coder and native Turkish coder each coded speech and gesture data from four participants (21% of Dutch data and 18% of Turkish data). For speech coding at the clause level, high reliability was obtained (Dutch: 97.6% agreement, $\kappa = 0.964$; Turkish: 94.1% agreement, $\kappa = 0.957$). Similarly, for gesture coding the reliability was high (Dutch: 95.3% agreement, $\kappa = 0.906$; Turkish: 94.7% agreement, $\kappa = 0.913$).

2.5. Statistical analysis

Data were analyzed with generalized binomial linear mixed-effects modelling using the *glmer* function from the *lme4* package (version 1.1.26; Bates et al., 2015) in *R* (version 3.5.3; R Core Team, , 2019) with the optimizer *bobyqa* (Powell, 2009). This mixed-effects approach takes into account the random variability due to having different items and participants. We started off with the maximal random effects structures justified by our design (Barr et al., 2013). When a maximal model failed to converge, we removed random effects, removing interactions first, and choosing between two possible structures based on the lowest Akaike's Information Criterion (AIC). For each fixed effect factor, sum-to-zero contrast coding was used (e.g., for Language: Turkish -0.5 , Dutch $+0.5$). The first mentioned factor level was always coded as -0.5 , and the second as $+0.5$. For all analyses, three trials were excluded (two Dutch) in which the addressee talked and affected the speaker's speech production. Data and analysis code are available at <https://osf.io/p8cas/>.

3. Results

3.1. Speech production

First, we tested whether the frequency of encoding path or manner in speech differed cross-linguistically (Fig. 3). Overall, Manner was mentioned very often by both Dutch-speakers (in 297/302 descriptions; 98%) and Turkish-speakers (in 337/351 descriptions; 96%). Path was mentioned less frequently by both Dutch-speakers (in 201/302 descriptions; 67%) and Turkish-speakers (in 267/351 descriptions; 76%). This pattern was tested with a model including Language (Turkish and Dutch), Component (Path and Manner) and their interaction on binary values for mention in speech (0 = no and 1 = yes) at the item level. It revealed only a main effect of Component ($\beta = 4.46$, $SE = 1.45$, $z = 3.07$, $p < 0.01$), with speakers mentioning Manner more often than Path.

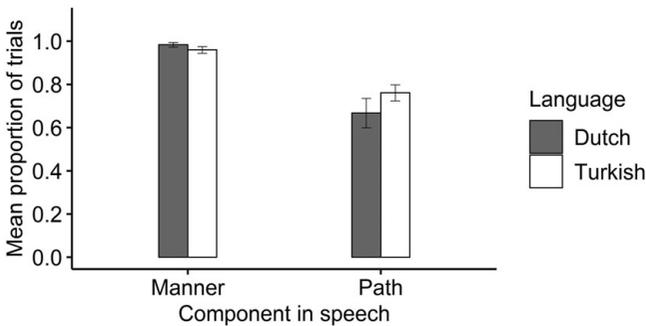


Fig. 3. Speakers of both languages described Manner more often than Path. For visualization, we calculated for each participant the proportion of trials in which they described Path and Manner. The figure shows the mean proportions, separated by Component and Language. Error bars represent the standard error.

Second, we tested whether the descriptions in which participants specifically encoded an event component (i.e., path or manner) were equally frequent across the speakers of two languages. When mentioning Manner in an event description, specific descriptions were almost always used by both Dutch-speakers (in 289/297 manner descriptions; 97%) and Turkish-speakers (in 325/337 manner descriptions; 96%). However, when mentioning Path, Dutch-speakers almost always used specific descriptions (in 196/200 path descriptions; 98%), whereas Turkish-speakers used specific descriptions less frequently (in 198/267 path descriptions; 74%). This pattern was tested with separate models for Manner and Path. For each model, we only included trials in which participants had mentioned that component in speech. Both models included Language (Turkish and Dutch) as a predictor for binary values for whether the mention in speech was specific (0 = unspecific and 1 = specific). The Manner model revealed no effect of Language ($\beta = -0.33$, $SE = 1.66$, $z = -0.20$, $p > 0.05$), but the Path model did ($\beta = 3.97$, $SE = 1.00$, $z = 3.96$, $p < 0.01$). Thus, although speakers of both languages were equally often specific about Manner, Turkish-speakers were less often specific about Path than Dutch-speakers.

3.2. Gesture production

Next, we tested whether the frequency of encoding path or manner in gesture differed cross-linguistically (Fig. 4). Overall, Turkish-speakers gestured more often than Dutch-speakers. Path was gestured more often by Turkish-speakers (in 168/351 descriptions; 48%) than by Dutch-speakers (in 67/302 descriptions; 22%). Manner was also gestured more often by Turkish-speakers (in 165/351 descriptions; 47%) than by Dutch-speakers (in 91/302 descriptions; 30%). This pattern was tested with a model including Language (Turkish and Dutch), Component (Path and Manner), and their interaction on binary values for whether a component was encoded in gesture (0 = no and 1 = yes) at the item level. Note that when an event description contained gestures about both Path and Manner (either in separate or conflated gestures), it contributed to both categories (both the Path and Manner bars in Fig. 4). The model revealed a main effect of Language ($\beta = -1.36$, $SE = 0.45$, $z = -3.01$, $p < 0.01$), indicating that Turkish-speakers gestured more often than Dutch-speakers. No other effects or interactions were significant.

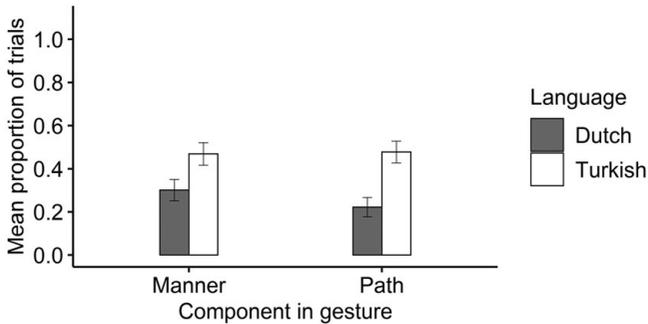


Fig. 4. Turkish-speakers gestured more often than Dutch-speakers. For visualization, we calculated for each participant the proportion of trials in which they gestured about Path and Manner. The figure shows the mean proportions, separated by Component and Language. Error bars represent the standard error.

To better understand the speech context in which these gestures were produced, we looked at the relation between the event components depicted in gesture and the event components described in speech. When path was depicted in a gesture, in the vast majority of the descriptions (82.6%) it was also described in speech. Similarly, when manner was depicted in gesture, it was also described in speech in 99.2% of the cases. Thus, these path and manner gestures typically accompanied path and manner speech, respectively.

3.3. Memory performance

Before analyzing motion event memory, we compared the Dutch- and Turkish-speakers' performance for filler item memory, the digit-span task (verbal working memory), and the Corsi block-tapping task (visuospatial working memory). For filler event memory, both Dutch-speakers ($M = 0.99$) and Turkish-speakers ($M = 0.95$) reached very good accuracy. A model tested the effect of Language (Turkish and Dutch) on binary values for whether a filler item was remembered (0 = no and 1 = yes) revealed no significant effect of Language ($\beta = 1.32$, $SE = 1.08$, $z = 1.22$, $p > 0.05$). For the digit-span task, Dutch participants ($M = 7.63$) and Turkish participants ($M = 7.91$) had similar scores. A linear regression model tested the effect of Language (Turkish and Dutch) on integer digit-span values also did not revealed any differences between Dutch-speakers and Turkish-speakers in verbal working memory capacity, $t(39) = -0.81$, $p > 0.05$. A similar analysis for Corsi-spans revealed that Dutch participants ($M = 7.53$) had higher Corsi-spans than Turkish participants ($M = 6.82$), $t(39) = 2.04$, $p = 0.048$. Thus, Dutch participants had higher visuospatial working memory capacity. To test whether this difference in Corsi-spans was important for our motion event memory analysis, a Spearman-rank correlation was calculated between each participant's memory accuracy for the motion events and their Corsi-span. Results revealed no correlation between Corsi-span and motion event accuracy, $r_s = -0.00$, $p = 0.98$. Therefore, the Corsi-span was not further taken into account in the analyses.

Next, we analyzed motion event memory for the three different types of memory items: Path-changes, Manner-changes, and No-change items. Collapsed across language groups, memory for Path-changes ($M = 0.68$, $SD = 0.26$, $t(40) = 4.29$, $p < 0.001$)

and No-change items ($M = 0.78$, $SD = 0.15$, $t(40) = 11.98$, $p < 0.001$) were significantly higher than chance level, but Manner-change memory was not higher than chance ($M = 0.40$, $SD = 0.26$, $t(40) = -2.49$, $p = 0.99$). This pattern held for the majority of participants, with 34 out of 41 participants not reaching Manner-change accuracy above the chance level of 0.5. Thus, we did not attempt to predict manner memory using speech, gesture and native language because that would be predicting behavior that is indistinguishable from guessing (which would lead to participants responding correctly only half of the time).

Finally, we analyzed how path speech, path gestures and native language related to path memory. Recall that path information in speech could be specific, unspecific or absent. For predicting path memory, unspecific path mentions in speech that only used unspecific verbs (e.g., to advance in Turkish) or adverbs (e.g., away in Dutch; $n = 48$) were analyzed together with trials in which path was not mentioned at all ($n = 146$) and were contrasted to specific path mentions with prepositions, spatial/directional nouns or path verbs (see also Section 2.4; $n = 283$). There were not enough unspecific descriptions to create a separate category. We reasoned that the effect of these unspecific descriptions on memory would be most similar to not mentioning path at all because they could be used regardless of the trajectory of motion. That is, they matched both the original event and its path change, and thus would likely not help to detect changes to the trajectory. Similarly, there were not enough path gestures in the sagittal axis ($n = 34$) to create a separate category. Like unspecific path descriptions, these sagittal gestures matched both the original event and its path change, and thus would likely not help to detect changes to the trajectory. Therefore, they were analyzed together with no gesture trials ($n = 310$) and were contrasted to path gestures in the lateral axis with the correct direction ($n = 133$). Trials describing the incorrect path ($n = 1$) or with lateral path gestures in the incorrect direction ($n = 2$) were excluded, as they might hinder memory. Moreover, we excluded trials in which the addressee talked ($n = 7$), as that could have affected the participant's memory. Fig. 5 shows path memory across Dutch and Turkish speakers separated by Path in speech and Path in Gesture.

A glmer model tested the effects of Path in speech (No mention and Path mention), Path in gesture (No gesture and Path gesture), Language (Turkish and

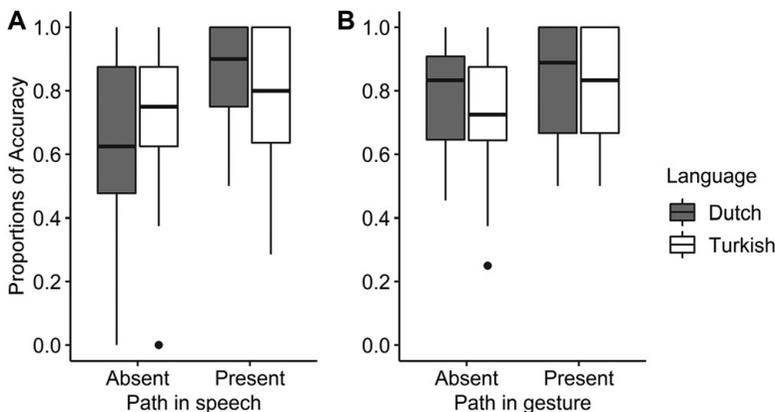


Fig. 5. Proportions of accurate path memory response in memory task. (A) Data separated by Path in speech and Language. (B) Data separated by Path in gesture and Language.

Dutch), and Condition (No-change and Path-change) on binary values for whether an item was remembered (0 = no and 1 = yes). We started with a four-way interaction model, which did not converge. The interaction was simplified into four three-way interactions, none of which was significant. Searching for a more parsimonious model, we first removed the interaction of which the removal resulted in the lowest AIC. Next, nonsignificant predictors were removed if that improved the model fit. Finally, we attempted to add random slopes, but this resulted in convergence issues.

The best-fitting model revealed only a significant interaction between Condition and Path in speech. Parameter estimates from the model are presented in Table 1. The interaction between Condition and Path in speech indicated that for No-change items, participants had similar memory accuracy regardless of whether the Path was mentioned in speech, while for Path-changes, participants were better at detecting changes when they had mentioned that Path in speech compared to when they had not (Fig. 6). No other main effects or interactions were significant. Notably, there

Table 1. Fixed effects from the mixed model predicting path memory

Effect	β	SE	Z	p
(Intercept)	1.353	0.311	4.346	<0.001***
Path in speech <i>Absent vs. Present</i>	0.311	0.393	0.791	0.429
Path in gesture <i>Absent vs. Present</i>	0.612	0.387	1.579	0.114
Language <i>Turkish vs. Dutch</i>	0.455	0.408	1.115	0.265
Condition <i>No-change vs. Path-change</i>	-0.707	0.512	-1.382	0.167
Path in speech * Path in gesture	-0.594	0.759	-0.783	0.434
Path in speech * Language	0.234	0.755	0.310	0.756
Path in gesture * Language	0.318	0.752	0.423	0.673
Path in speech * Condition	1.259	0.512	2.460	0.014*
Path in speech * Path in gesture * Language	-2.675	1.509	-1.773	0.076

Estimates (β), standard errors (SE), z-values, and p-values are given. Significance codes: * $p < 0.05$, *** $p < 0.001$. Formula in R: `glmer (Accuracy ~ Path in speech*Path in gesture*Language + Path in speech*Condition + (1|Subject) + (1|Item), family = binomial, glmerControl (optimizer="bobyqa", optCtrl = list(maxfun = 1,000,000)))`.

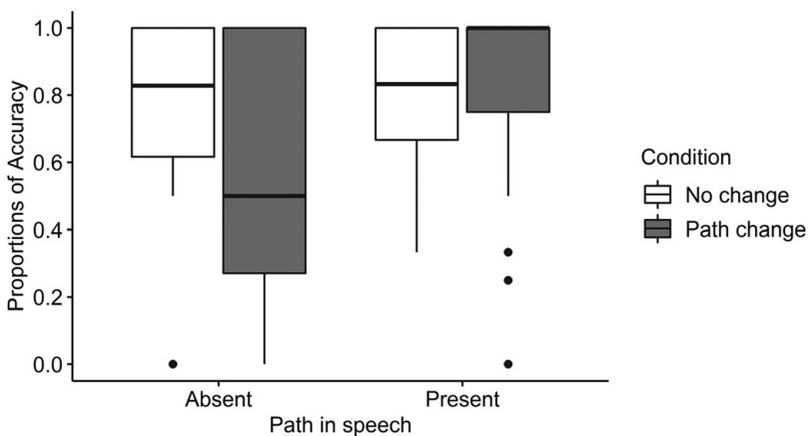


Fig. 6. Proportions of accurate path memory response in memory task, separated by Path in speech and Condition. Data from Dutch-speakers and Turkish-speakers are collapsed. For the Path change condition, memory was more accurate when Path had been specifically mentioned in speech. For the No change condition, memory was similar regardless of whether Path had been described.

were no significant effects involving Language, indicating that similar patterns were found for Dutch- and Turkish-speakers. Moreover, there were no significant effects involving Gesture. Thus, gesturing about path did not correspond to better memory for path. We turn to the significance of these findings below.

4. Discussion

Our first goal in the present study was to test whether encoding certain motion event components in speech predicts better memory for the very same event components. Our second goal was to test whether iconic depictions of event components in gesture would predict even better memory above and beyond speech. Throughout our investigation, we also asked whether the relation between multimodal event encodings in speech and gesture and memory varies cross-linguistically. Below, we summarize and discuss our main findings on cross- and within-language differences in multimodal motion event encodings and their relation to motion event memory.

4.1. Motion events in speech and gesture

In order to motivate our investigation of how motion event encodings in speech and gesture predict memory, we first explored multimodal encodings of motion events by Dutch- and Turkish-speakers. In speech, both Dutch- and Turkish-speakers almost always mentioned the manner and more than the path, but no differences were found between the two languages in preferring one component over the other. This seems to go against the classic typological finding that speakers of verb-framed languages omit manner more often than speakers of satellite-framed languages (Slobin, 2003; see also Özyürek et al., 2008 for a comparison between Turkish and English). This difference could be due to the stimuli: the manners in our study were rather salient (tiptoe, twirl, and hop) as they were not a default way of changing location and were more unusual than the manners used in previous work (e.g., walk, run, carry in Gennari et al., 2002; Papafragou et al., 2002). It is plausible that Turkish-speakers deemed it important to mention the manners because they were salient and contrastive across trials. Indeed, speakers of Greek, another verb-framed language, mention the manner of motion almost twice as often when it is not easily inferable for a listener (Papafragou et al., 2006). Thus, although cross-linguistic differences in manner omission are well-established, these results suggest that within-language encoding flexibility can diminish these cross-linguistic differences under certain conditions.

A more exploratory finding was that Dutch- and Turkish-speakers differed in how specifically they described the path. Dutch-speakers almost always described the spatial relation between the figure and the landmark, or the motion in the left-right axis, in a way that the description clearly disambiguated between *to* and *from* paths. By contrast, Turkish-speakers regularly used the unspecific path verbs *ilerle* 'to advance' or *git* 'to go'. These cross-linguistic differences are reminiscent of previously demonstrated differences between Dutch and other languages (e.g., French) in the semantics of placement verbs (Gullberg, 2011). Together, these findings highlight the relevance of more fine-grained cross-linguistic analyses and moving beyond frequencies of mentioned components. Moreover, the cross-linguistic differences in speech specificity between Dutch and Turkish could be a domain for further research to explore subtle consequences of speech on memory.

Regarding gestures, there were no differences in the frequencies of encoding path and manner across languages, apart from a general trend of more gestures in Turkish than Dutch both for path and manner. This is in line with previous studies that took speech content into account when looking at gesture differences. For example, when speaking about manner, English- and Japanese-speakers do not differ in their frequency of gesturing about manner (Brown & Chen, 2013). Similarly, when looking at either manner-only sentences or path-only sentences, also no cross-linguistic differences were found in the likelihood of English- and Turkish-speakers to gesture about manner and/or path (Özyürek et al., 2005). Indeed, we also found that speakers typically gestured about an event component if they also spoke about it, showing a tight link between speech and gesture. Thus, it appears that when speakers of verb-framed and satellite-framed languages speak about the same components, they also gesture about the same components in line with the Interface Hypothesis (Kita & Özyürek, 2003; but see Brown & Gullberg, 2008).

There is one difference with previous literature that deserves highlighting. Previous studies found that speakers of different languages gesture more about path than manner (Farsi: Akhavan et al., 2017; Mandarin Chinese: Chui, 2009; French: Gullberg et al., 2008; Turkish: Mamus et al., 2021). However, we found that frequency of path and manner gestures did not differ cross-linguistically and both languages gestured more about manner than path. Again, this could be due to the use of salient manners in the current experiment, since speakers are more likely to gesture about highly salient manners than about less salient manners (Yeo & Alibali, 2018). Thus, using these salient manners may have skewed both our speech and gesture production results. For speech, it may have increased the likelihood of manner mention, such that Turkish-speakers reached the same (ceiling) frequency as Dutch-speakers, thus eliminating cross-linguistic differences. For gesture, it may have increased the likelihood for speakers of both languages to gesture about manner, thus removing a previously found path gesture preference. This highlights the importance of stimuli construction when investigating (cross-linguistic) motion event descriptions.

4.2. *Speech predicts path memory in Dutch and Turkish*

One key aim of this study was to test whether mentioning an event component in speech would predict better memory for that component. Consistent with this possibility, speakers who mentioned path in speech were better at detecting changes to that path. One previous study also found that speaking about path predicts better memory for path (Billman et al., 2000), but others did not (Bunger et al., 2012; Papafragou et al., 2002). These conflicting findings could be the result of the heterogeneity of stimuli, procedures, and participant groups. However, the link between speech and memory is consistent with prior findings from other domains, demonstrating relations between how speakers describe and remember visual stimuli (e.g., eye-witness-memory: Marsh et al., 2005; picture recognition: Zormpa et al., 2019). Our results suggest that speech and memory are related, possibly because speaking about a component indicates that the speaker has this event component as part of their event conceptualization (Levelt, 1989; Papafragou et al., 2008).

Another aim of this study was to see whether the link between motion event speech and memory varied across typologically different languages. We found that this was not the case: speaking about path predicted better memory for path changes

for speakers of Dutch and Turkish. Although this finding was not consistent with our predictions, it is reminiscent of the findings of a recent study comparing speakers of other verb-framed (Greek) or satellite-framed (English) languages (Skordos et al., 2020). In that study, producing motion verbs affected motion event memory similarly across language groups. Our findings extend this pattern to another pair of typologically different languages, which was not studied before in this respect. Furthermore, our findings show that the relation between describing and remembering events does not vary cross-linguistically even if speakers describe motion events in full utterances instead of only single verbs (see also Karadöller et al., 2021 for similar developmental evidence in the domain of static spatial relations).

4.3. Using co-speech gesture on top of speech does not predict path memory

A second key aim of the present study was to test whether gesturing about a motion event component predicts memory over and above speaking due to dual encoding of information in an iconic way (Paivio, 1990; Perniss et al., 2010) or by being a part of event conceptualization (Kita & Özyürek, 2003) and in turn affecting memory. However, we found no relation between path gestures and path memory. Importantly, these path gestures typically co-occurred with path speech indicating that path memory was equally accurate for paths encoded in both speech and gesture, compared to paths encoded only in speech. This suggests that dual encoding of motion event components in speech and gesture does not enhance motion event memory further. Furthermore, even though gesture production is linked to attention allocation during message conceptualization (Ünal et al., *under review*) it was not linked to better memory for the gestured components.

A possible explanation for why gesture does not enhance memory above verbal encoding concerns the way speech and gesture encodes information (cf. Lupyan, 2008, 2012). While gesture is analogue and allows information to be conveyed imagistically, speech is categorical and relies on discrete units (Cook et al., 2012). Encoding information in a categorical way could be more helpful for memory, in line with previous studies showing benefits of categorical language on spatial cognition (Dessalegn & Landau, 2008; Feist & Gentner, 2007; Gentner et al., 2013). In order to fully evaluate this possibility, further research should test whether co-speech gestures benefit memory when the information depicted in gesture is complementary, rather than redundant.

The absence of a link between path gesture and memory may seem surprising, given prior work showing a link between gesture production and event memory (Cook et al., 2010; Koranda & MacDonald, 2015). This discrepancy might be attributed to an important difference between these studies and ours: while the present study used motion events only, these previous studies either collapsed motion events with actions (Cook et al., 2010) or used actions only (Koranda & MacDonald, 2015). Gestures depicting actions may differ from gestures depicting motion paths. For example, action gestures might involve stronger motor simulation than tracing path gestures as they are more likely to be enacted from a first-person perspective (Hostetter & Alibali, 2008). On the other hand, path gestures trace the trajectory of motion from a third-person perspective. Further work is needed to more precisely estimate whether there is a hierarchy in gestures that depict event information in more vs. less embodied ways in terms of predicting event memory. Another possible

explanation for the discrepancy is the type of task used to assess memory. The present study used a recognition memory task in which participants responded nonverbally to visually presented stimuli. By contrast, Cook et al. (2010) used a free and cued recall task in which participants verbally described the event they previously seen. Further research is necessary to pinpoint the contributions of these factors to the relation between gesture production and memory.

4.4. Memory for manner versus path

Our study was the first to directly compare memory for manners and paths using intransitive events. While manners were not remembered, path memory was much better. This path-advantage has also been found in prior work that used instrumental motion events, where the manner changes were object changes (e.g., roller skates change into a skateboard; Bunger et al., 2012; Skordos et al., 2020; but see Engemann et al., 2015). Such a path-advantage has also been found developmentally, where infants are better and earlier at categorizing path compared to manner (Pruden et al., 2012, 2013). This might have to do with path being a core aspect of an event (Radvansky & Zacks, 2014) providing information about the intentionality of motion (Pourcel, 2004). A similar link between motion event memory and intentionality has been found for memory for goal-paths versus source-paths. People remember goals better than sources, potentially because goals provide more information about animate figures' intentions (Lakusta & Landau, 2012).

Notably, while participants did not remember manners, speakers of both languages almost always spoke about manner and gestured about it considerably. This speech-gesture-memory dissociation can be interpreted in two ways. One possibility is that there is no strong link between language and manner memory. To test this, it is necessary to increase manner memory accuracy to above chance. Another possibility is that the motion event information that is important to communicate to a recipient is not the same as the information that is important to encode in memory. For communicating about motion, manner may be important when it is salient and/or not inferable. Conversely, for memory, path may be important as it relates to intentionality of an agent's motion.

4.5. Methodological implications

Before we conclude, we would like to highlight one aspect of our findings that has implications for future cross-linguistic investigations of event memory and, more generally, cognition. In the present study, following practices common in prior work (Sakarias & Flecken, 2019), we used two measures of working memory to eliminate group-level differences that might potentially explain differences in motion event memory. Nevertheless, we found that these working memory measures did not correlate with our event memory measure. This opens up discussions about the suitability of these measures for establishing group-level similarities in general cognitive ability in cross-linguistic research. In fact, in previous work, the correlations between these measures and the main measures of interest are rarely tested because the majority of these studies did not find cross-linguistic differences on these working memory tasks and data are dropped from further analyses. We suggest that an alternative approach for establishing group-level similarities could be building in

controls within the main memory task by including items that are not expected to be affected by the cross-linguistic distinctions of interest (e.g., the filler and object-memory items in the current study, see also Ünal et al., 2016 for a similar approach).

5. Conclusions

To conclude, the present study reveals that how people describe an event in speech predicts their memory for that event. However, gesturing about those event components does not seem to enhance event memory on top of speaking. Furthermore, the relation between speaking and remembering motion events did not vary across typologically different languages. Together these findings suggest that how speakers describe a motion event is more important than the typology of the speakers' native language in predicting motion event memory.

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/langcog.2022.3>.

Acknowledgments. We thank Özge Baturlar, Melis Hazır, Merve Ağırbaşı, Elif Balcıoğlu, Leandra Mulder, Roos van de Boom, Milou van Helvert and Melle ter Bekke for assistance with data collection and annotation. We thank Yuka Bekkers and Şevval Cihankaya for assistance with reliability coding.

Funding Statement. This work is supported by a VICI grant, awarded by the Dutch Research Council (NWO) (grant number 277-70-013) to Aslı Özyürek.

Data Availability Statement. The stimuli, data and analysis scripts are available at the Open Science Framework Repository <https://osf.io/p8cas/>

Conflict of Interest. The authors declare no conflicts of interest.

References

- Akhavan, N., Nozari, N., & Göksun, T. (2017). Expression of motion events in Farsi. *Language, Cognition and Neuroscience* 32, 792–804. <https://doi.org/10.1080/23273798.2016.1276607>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Billman, D., Swilley, A., & Krych, M. (2000). Path and manner priming: Verb production and event recognition. Proceedings of the 22nd Annual Conference of the Cognitive Science Society, 615–620.
- Brown, A., & Chen, J. (2013). Construal of Manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics* 24, 605–631. <https://doi.org/10.1515/cog-2013-0021>
- Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition* 30, 225–251. <https://doi.org/10.1017/s0272263108080327>
- Bunger, A., Trueswell, J. C., & Papafragou, A. (2012). The relation between event apprehension and utterance formulation in children: Evidence from linguistic omissions. *Cognition* 122, 135–149. <https://doi.org/10.1016/j.cognition.2011.10.002>
- Bylund, E., & Athanasopoulos, P. (2015). Introduction: Cognition, motion events, and SLA. *The Modern Language Journal* 99, 1–13. <https://doi.org/10.1111/j.1540-4781.2015.12175.x>
- Chui, K. (2009). Linguistic and imagistic representations of motion events. *Journal of Pragmatics* 41, 1767–1777. <https://doi.org/10.1016/j.pragma.2009.04.006>

- Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review* 1, 57–80. <https://doi.org/10.1007/bf01326550>
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language* 63, 465–475. <https://doi.org/10.1016/j.jml.2010.07.002>
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes* 27, 594–610. <https://doi.org/10.1080/01690965.2011.567074>
- Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. McGill University.
- Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science* 19, 189–195. <https://doi.org/10.1111/j.1467-9280.2008.02066.x>
- Engemann, H., Hendriks, H., Hickmann, M., Soroli, E., & Vincent, C. (2015). How language impacts memory of motion events in English and French. *Cognitive Processing* 16, 209–213. <https://doi.org/10.1007/s10339-015-0696-7>
- Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, 35, 283–296. <https://doi.org/10.3758/BF03193449>
- Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism* 15, 466–485. <https://doi.org/10.1177/1367006911403062>
- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition* 83, 49–79. [https://doi.org/10.1016/s0010-0277\(01\)00166-4](https://doi.org/10.1016/s0010-0277(01)00166-4)
- Gentner, D., & Goldin-Meadow, S. (eds). (2003). *Language in mind: Advances in the study of language and thought*. MIT Press.
- Gentner, D., Özyürek, A., Gürçanlı, Ö., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition* 127, 318–330. <https://doi.org/10.1016/j.cognition.2013.01.003>
- Gullberg, M. (2011). Language-specific encoding of placement events in gestures. In J. Bohnermeyer & E. Pederson (eds), *Event representation in language and cognition*, 166–188. Cambridge University Press. <https://doi.org/10.1017/CBO9780511782039.008>
- Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language* 28, 200–236. <https://doi.org/10.1177/0142723707088074>
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review* 15, 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Ibarretxe-Antuñano, I. (ed.). (2017). *Motion and space across languages: Theory and applications*, vol. 59. John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.59>
- Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2021). Spatial language use predicts spatial memory of children: Evidence from sign, speech, and speech-plus-gesture. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar-Raible (eds), *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*, 672–678. Cognitive Science Society.
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21, 260–267. <https://doi.org/10.1177/0956797609357327>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511807572>
- Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi block-tapping task: Standardization and normative data. *Applied Neuropsychology* 7, 252–258. https://doi.org/10.1207/S15324826AN0704_8
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48, 16–32. [https://doi.org/10.1016/S0749-596X\(02\)00505-3](https://doi.org/10.1016/S0749-596X(02)00505-3)
- Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes* 22, 1212–1236. <https://doi.org/10.1080/01690960701461426>
- Koranda, M., & MacDonald, M. (2015). Language and gesture descriptions affect memory: A nonverbal overshadowing effect. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 1183–1188.

- 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. <https://doi.org/10.1111/j.1551-6709.2011.01220.x>
- Landau, B., Dessalegn, B., & Goldberg, A. M. (2010). Language and space: Momentary interactions. In P. Chilton & V. Evans (eds), *Language, cognition and space: The state of the art and new directions*, 51–78. Equinox Publishing.
- Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods* 41, 841–849. <https://doi.org/10.3758/BRM.41.3.841>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press. <https://doi.org/10.7551/mitpress/6393.001.0001>
- Lupyan, G. (2008). From chair to “chair”: A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology. General* 137, 348–369. <https://doi.org/10.1037/0096-3445.137.2.348>
- Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, 3, Article 54. <https://doi.org/10.3389/fpsyg.2012.00054>
- Mamus, E., Speed, L. J., Özyürek, A., & Majid, A. (2021). Sensory modality of input influences the encoding of motion events in speech but not co-speech gestures. In T. Fitch, C. Lamm, H. Leder, & K. Tessmar-Raible (eds), *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*, 376–382. Cognitive Science Society.
- Marsh, E. J., Tversky, B., & Hutson, M. (2005). How eyewitnesses talk about events: Implications for memory. *Applied Cognitive Psychology* 19, 531–544. <https://doi.org/10.1002/acp.1095>
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press. doi:10.7208/chicago/9780226514642.001.0001
- McNeill, D. & Duncan, S.D. (2000). Growth points in thinking-for-speaking. In D. McNeill (ed), *Language and Gesture*, 141–161. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620850>
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition* 148, 10–18. <https://doi.org/10.1016/j.cognition.2015.12.001>
- Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology* 44, 1040–1054. <https://doi.org/10.1037/0012-1649.44.4.1040>
- Özyürek (2017). Function and processing of gesture in the context of language. In R. Breckinridge Church, M. W. Alibali & S.D. Kelly (eds), *Why Gesture?: How the hands function in speaking, thinking and communicating*, 39–58. John Benjamins Publishing Company. <https://doi.org/10.1075/gs.7Mc>
- Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from cross-linguistic variations and similarities. *Gesture* 5, 219–240. <https://doi.org/10.1075/bct.10.15ozy>
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195066661.001.0001>
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition* 108, 155–184. <https://doi.org/10.1016/j.cognition.2008.02.007>
- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, ‘n’ roll: The representation of motion in language and cognition. *Cognition* 84, 189–219. [https://doi.org/10.1016/S0010-0277\(02\)00046-X](https://doi.org/10.1016/S0010-0277(02)00046-X)
- Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition* 98, 75–87. <https://doi.org/10.1016/j.cognition.2005.05.005>
- Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology* 1, Article 227. <https://doi.org/10.3389/fpsyg.2010.00227>
- Pourcel, S. (2004). What makes path of motion salient? *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society* 30, 505–516. <https://doi.org/10.3765/bls.v30i1.963>
- Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. In *Technical Report DAMTP 2009/NA06*. Centre for Mathematical Sciences, University of Cambridge. http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf

- Pruden, S. M., Göksun, T., Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2012). Find your manners: How do infants detect the invariant manner of motion in dynamic events? *Child Development* 83, 977–991. <https://doi.org/10.1111/j.1467-8624.2012.01737.x>
- Pruden, S. M., Roseberry, S., Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Infant categorization of path relations during dynamic events. *Child Development* 84, 331–345. <https://doi.org/10.1111/j.1467-8624.2012.01843.x>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Core Team. <https://www.R-project.org/>
- Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199898138.001.0001>
- Sakarias, M., & Flecken, M. (2019). Keeping the result in sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science* 43, e12708. <https://doi.org/10.1111/cogs.12708>
- Skordos, D., Bunger, A., Richards, C., Selimis, S., Trueswell, J., & Papafragou, A. (2020). Motion verbs and memory for motion events. *Cognitive Neuropsychology* 37, 1–17. <https://doi.org/10.1080/02643294.2019.1685480>
- Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind*, 157–191. MIT Press. <https://doi.org/10.7551/mitpress/4117.001.0001>
- Talmy, L. (2000). *Toward a cognitive semantics*. MIT Press. <https://doi.org/10.7551/mitpress/6848.001.0001>
- Ünal, E., Manhardt, F., & Özyürek, A. (under review). Speaking and gesturing guide event perception during message conceptualization: Evidence from eye movements.
- Ünal, E., Pinto, A., Bunger, A., & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language* 87, 157–176. <https://doi.org/10.1016/j.jml.2015.10.009>
- Wechsler, D. (1944). *The measurement of adult intelligence*. The Williams & Wilkins Company.
- Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *WIREs Cognitive Science* 2, 253–265. <https://doi.org/10.1002/wcs.104>
- Yeo, A., & Alibali, M. W. (2018). Does visual salience of action affect gesture production? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44, 826–832. <https://doi.org/10.1037/xlm0000458>
- Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019). The production effect and the generation effect improve memory in picture naming. *Memory* 27, 340–352. <https://doi.org/10.1080/09658211.2018.1510966>

A. Motion Event Stimuli

	Event	Memory type	Changed memory item
1	A woman hops to a cactus	path-change	hop from
2	A woman tiptoes out of a garden tent	path-change	tiptoe into
3	A woman twirls from a drawers	path-change	twirl to
4	A woman runs into a gazebo	path-change	run out of
5	A woman hops from an armchair	manner-change	tiptoe from
6	A woman tiptoes into a phone booth	manner-change	twirl into
7	A woman twirls to a pile of rocks	manner-change	run to
8	A woman runs out of a greenhouse	manner-change	hop out of
9	A woman hops into a market stand	no-change	
10	A woman hops out of a beach umbrella	no-change	
11	A woman tiptoes to a street lamp	no-change	
12	A woman tiptoes from a traffic sign	no-change	
13	A woman twirls into a beach bar	no-change	
14	A woman twirls out of a bus stop	no-change	
15	A woman runs to a pine tree	no-change	
16	A woman runs from a palm tree	no-change	

B. Filler Event Stimuli

	Event	Memory type	Changed memory item
1	A woman puts on a scarf	object-change	putting on a wool scarf
2	A woman puts on reading glasses	object-change	putting on sunglasses
3	A woman cuts an apple	object-change	cutting a lemon
4	A woman tears a piece of white paper	object-change	tearing a piece of red paper
5	A woman turns pages of a newspaper	object-change	turning pages of a book
6	A woman puts on a hat	object-change	putting on a different hat
7	A woman eats a piece of a muffin	object-change	eating a piece of chocolate
8	A woman bites an apple	object-change	biting a banana
9	A woman puts tape on a piece of paper	no-change	
10	A woman combs her hair	no-change	
11	A woman closes a box	no-change	
12	A woman rolls dice	no-change	
13	A woman tears a piece of paper towel	no-change	
14	A woman puts paper clips on paper	no-change	
15	A woman opens the cover of a book	no-change	
16	A woman inflates a balloon	no-change	

Cite this article: ter Bekke, M., Özyürek, A. & Ünal, E. (2022). Speaking but not gesturing predicts event memory: a cross-linguistic comparison. *Language and Cognition* 14: 362–384. <https://doi.org/10.1017/langcog.2022.3>