Article:
Fu, Xingyi, Vanek, Norbert and Roberts, Leah (2023) Matched or moved? Asymmetry in high- and low-level visual processing of motion events. Language and Cognition. pp. 283-306. ISSN: 1866-9808

https://doi.org/10.1017/langcog.2023.37

Language and Cognition (2024), 16: 2, 283–306
doi:10.1017/langcog.2023.37

ARTICLE

Matched or moved? Asymmetry in high- and low-level visual processing of motion events

Xingyi Fu[1], Norbert Vanek[2,3] and Leah Roberts[1]

[1]University of York, Department of Education, York, UK; [2]School of Cultures, Languages and Linguistics, University of Auckland, Auckland, New Zealand; [3]Experimental Research on Central European Languages Lab, Charles University, Prague, Czechia
Corresponding author: Norbert Vanek; Email: norbert.vanek@auckland.ac.nz

(Received 27 January 2023; Revised 08 July 2023; Accepted 19 July 2023)

Abstract
Consensus on the extent to which cross-linguistic differences affect event cognition is currently absent. This is partly because cognitive influences of language have rarely been examined within speakers of different languages in tasks that manipulate the level of visual processing. This study presents a novel combination of a high-level approach upregulating the involvement of language, namely self-paced sentence-video verification, and a low-level visual detection method without language use, namely breaking continuous flash suppression (b-CFS) (Yang et al., 2014). The results point to cross-linguistic effects on event cognition by revealing variations in visual processing patterns of manner and path by English versus Mandarin Chinese speakers. Language specificity was found on both levels of processing. An asymmetry in response speed across tasks highlights an important difference between facilitation of detecting contrasts when recruitment of verbal labels is automatic, versus facilitation of verifying correspondences when labels are overt.

Keywords: motion event cognition; manner and path; continuous flash suppression; sentence-video verification; low-level visual processing; linguistic relativity

1. Introduction

Much uncertainty still surrounds the relationship between language and motion event cognition. One account of this link comes from linguistic relativity (Whorf, 1956), with the reasoning that the way in which a language typically expresses motion events can influence how speakers of that language perceive and think about motion. The domain of motion events has been intensively investigated in search of a gradually more refined understanding of the correlations between different encodings of motion across languages and their speakers' cognitive processes (e.g., Athanasopoulos & Bylund, 2013; Feist, 2016; Papafragou et al., 2008). Recent extensions also explored the impact of linguistic expression of motion on perceptual processes

(e.g., Flecken et al., 2015; Slivac et al., 2021). However, the contexts and extent to which languages affect event cognition and perception are still debatable and often fiercely contested. One reason might be that previous studies examined only high-level visual processing, or only low-level visual processing in separation, which limits direct comparability and wider generalizations. *Low-level visual processing* refers to an early perceptual stage of visual stimulus detection and recognition (e.g., find the odd-one-out shape) and *high-level visual processing* denotes a later, post-perceptual stage of interpretation and semantic analysis (e.g., click on all the cars shown in the photo) (Lupyan et al., 2020; Lupyan & Ward, 2013).

A prominent framework to examine variation in the linguistic encoding of motion is Talmy's (1985, 2000) typology of lexicalization and event integration, followed by numerous studies investigating cross-linguistic effects on event cognition. Talmy classified languages into two major categories, namely, verb-framed (e.g., Spanish, German) and satellite-framed (e.g., English). The main difference between them is the use of the main verb in a sentence. More specifically, V-languages use the main verb to describe the *path* component of motion, whereas S-languages use the main verb to describe the *manner* component (Feist, 2010). It is important to emphasize that both S-framed and V-framed languages contain both *path* and *manner* verbs, and their dissimilarities lie primarily in the degree of emphasis or relative weighting rather than in an absolute sense. Different linguistic representations might modulate a range of more general cognitive processes from attention allocation to memory. Evidence for such effects abounds, especially when study designs let language be consciously recruited during experimental task completion. However, linguistic modulation effects often vanish when the task does not rely on the involvement of language, either covertly or overtly (Casasanto, 2016; Feist, 2016; Gleitman & Papafragou, 2013).

One possibility for robust variation in findings across studies on motion event cognition lies in the transient nature of linguistic modulations, transpiring specifically in on-line processing experiments (Athanasopoulos et al., 2015; Flecken et al., 2014; Gennari et al., 2002). To illustrate, Athanasopoulos and Bylund (2013) conducted four experiments to investigate the effect of cross-linguistic differences between English (aspect grammaticalized, more focus on ongoingness) and Swedish (aspect not grammaticalized, more endpoint-focused) on motion event cognition. The measures were native speakers' motion event similarity judgements either from memory or while viewing the events. The results showed that Swedish and English speakers performed significantly differently when they described endpoint-biased and ongoingness-biased motion events, and they also varied in memory-based similarity judgements when the task demands were high. However, cross-linguistic differences disappeared in an on-line similarity judgement task and another memory-based similarity judgement task which had verbal interference. Similarly transient linguistic modulations were found in the experiments by Gennari et al. (2002). They compared motion event recognition memory with and without previous linguistic encoding performed by English (*manner*-focused in the main verb) and Spanish (*path*-focused in the main verb) speakers. The aim was to test the effect of language processing on subsequent motion event recognition and similarity judgements. Cross-linguistic effects emerged only with previous verbal encoding. In this study, we present an alternative method to investigate the transient nature of linguistic modulations through examining different levels of visual processing.

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

Another possibility for the transience of (cross-)linguistic effects is that they may have arisen but were abolished due to methodological issues. Verbal interference, used with the aim to prevent subvocal rehearsal of participants during task performance, is popular in studies on event cognition. However, verbal interference varies across studies in type and complexity, ranging from number repetition (Athanasopoulos & Bylund, 2013; Ji, 2017), nonsense syllable repetition (Gennari et al., 2002), to detecting or counting ocean waves (Flecken et al., 2014). Such variation might impact the degrees to which participants are distracted from thinking in their languages during the experiment, and these degrees are difficult to measure and control (Perry & Lupyan, 2013). For example, asking bilinguals to switch between two languages to shadow numbers showed that reliance on a specific language during categorization varies as a function of whether access to a given language is available for task facilitation or if it is kept busy otherwise (in this case via number verbalization). Athanasopoulos et al. (2015) found support for this idea through testing English–German bilinguals in a categorization task with verbal interference shifting between English and German. Categorization preferences varied depending on the language verbal interference; they were more German-like with number distracters in English, and more English-like with number distracters in German. A dual task with number repetition presents a higher cognitive load than a nonverbal distracter such as ocean waves. This may explain why Flecken et al. (2014) is one of few studies reporting cross-linguistic effects in a behavioural task with obstructed language access. Ocean waves and other verbal distracters showed that the degree to which processing of motion events is modulated by the native language varies with task demand and the nature of distracters. This study offers an alternative to linguistic interference in the form of a pair of controlled high-level and low-level processing tasks to test when and how language structures influence motion event perception.

Returning to the *satellite-framed* (S-language) versus *verb-framed* (V-language) distinction advocated by Talmy (1985), Slobin (2004) proposed an additional category of *equipollently framed* E-languages (e.g., Mandarin Chinese) into the taxonomy. The major syntactic difference between English and Mandarin Chinese regarding motion events is the relative linguistic weight of *manner* and *path* components. In example (1), *carry* is the 'heavy-weight' main verb that expresses the *manner* of the motion event, and the satellite *into* expresses the *path*, whereas as an equipollently framed language, Mandarin Chinese conveys the information about *manner* and *path* by verbs with equal linguistic status. This is exemplified in (2), where both *push* and *into* are expressed with verbs by means of a serial verb construction (SVC). SVCs are commonly used for the expression of motion in Mandarin Chinese (Chen & Guo, 2009).

(1) A man carried a suitcase into a room.

(2) Yi2 ge4 nan2 ren2 ba3 yi2 ge4 xing2 li3 xiang1
A man BA a suitcase
ban1 jin4 (Manner+Cause+Path) le wu1 zi.
push-enter ASP_{perf} a room.
V_1 V_2
'A man carried a suitcase into a room.'

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

While the difference between Mandarin and English might seem trivial syntactically, in that the former encodes both semantic components in verbal units (i.e., SVC), whereas the latter only encodes *manner* in the verb and links *path* as a prepositional phrase to that verb, the significance of this difference shows when we take a closer look beyond verbalization. The cognitive implications of the differences that *manner* is relatively more prominent in English but more equal with *path* in Mandarin Chinese were tested in a series of experiments (e.g., Ji, 2017; Ji et al., 2011b, 2011a; Ji & Hohenstein, 2014a, 2014b). The main findings are that Mandarin and English speakers not only followed different patterns to describe the same motion events, but the time it took to process *manner*-salient and *path*-salient stimuli varied as a function of their linguistic encoding. With immediate relevance for this study, Ji (2017) asked Chinese and English speakers to view triads of motion event videos and decide whether scenes that shared the same *manner* or scenes that shared the same *path* were more similar to model scenes. While both groups preferred stimulus pairs which shared the same *path* information, English speakers were found to take less time to decide about motion event pairs when the *manner* was the same compared to matches when the *path* was the same. However, Mandarin speakers took a similar amount of time to make decisions about *manner*-based and *path*-based matches. One potential explanation can be the *manner salience* account (Slobin, 2006), which suggests that in the case of English, *manner* salience is high due to a distinct emphasis on the *manner* of motion in expression. The *manner* processing advantage remained even when language use through subvocal rehearsal was excluded by means of number repetition. The study did not specify the language of the number distracters it used, which does not help to resolve the uncertainty about the depth of linguistic effects on motion event cognition. Still, these findings suggest the possibility that linguistic influence during similarity judgements of motion events may be automatic/unconscious rather than strategic.

Automaticity of linguistic influence on event perception needs to be evidenced by effects in early time-windows before conscious access to language labels can occur. Some evidence in this direction already exists. Language specificity in motion event perception was reported for instance in Flecken et al. (2015), who measured brain responses of English and German speakers while they were watching series of motion events. German speakers' brain responses showed higher sensitivity to endpoint information, while English speakers' brain responses showed higher sensitivity to trajectories. If early/prelinguistic sensitivity can differ for endpoints versus trajectories depending on the participant's native language, it is reasonable to assume that cross-linguistic effects might extend to low-level visual detection of other features of motion events too, such as *manner* and *path*. The causal link to test is whether linguistic systems act as magnifiers of different visual representations as a result of long-term co-activation between specific motion components and their verbal encodings. Namely, if *manner* is habitually given more prominence in English than in Mandarin in verbal encodings of motion, *manner* information may be pushed more strongly on its way up the visual hierarchy for English speakers than for Mandarin speakers. There are different ways to check if language specificity plays a role already at the point when *manner* and *path* enter conscious analysis. One possibility is by means of measuring early brain responses, or alternatively, through a technique known as breaking continuous flash suppression (b-CFS).

Breaking continuous flash suppression (Jiang et al., 2007) is a useful paradigm able to monitor low-level *preverbal* processing and reveal when a visual stimulus enters

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

awareness. This is done by initially suppressing the stimulus through a flickering mask shown to the dominant eye, so the target stimulus starts off as invisible. This situation can compare to a key search in a mess. At first, the eyes scan all the mess, including the key. However, the key blends in with the background so it is hard to tell where it is right away. After some searching, the key becomes visible. Now imagine that the degree to which the mess distracts the key search is regulated through contrast changes. The initially low-contrast picture of a key gradually gains contrast up to the moment when its presence gets detected. That point in time is taken as the moment when the target, in this case the key, emerges into awareness. This paradigm integrates binocular rivalry and flash suppression (Stein 2019). The participant sees different images per eye, normally one eye will be presented a low-contrast target stimulus (e.g., a key) and the other/the dominant eye will be presented a high-contrast dynamic mask flashing continuously at 10 Hz (i.e., the mess). One major advantage of the b-CFS is that the target stimulus can be blocked from awareness by the mask for a relatively long time. The time needed for a stimulus to break through suppression depends on multiple factors including familiarity (more familiar objects get detected faster, e.g., Jiang et al., 2007) or whether the gradually appearing picture has a relevant verbal label (Lupyan & Ward, 2013). We chose the b-CFS paradigm because of its potential contributions to theories relevant to event cognition research. Our work is framed within the *predictive processing* account (Lupyan & Clark, 2015), which assumes that mental representations are shaped through a dynamic interaction between top-down predictions and bottom-up sensory signals. We use b-CFS to track how cross-linguistic variation in the encoding of motion can shape predictions against which visual input is assessed.

To date, there is some indication that *manner*-salient motion events break into awareness faster for English speakers than for Mandarin speakers. In a recent study, Vanek and Fu (2023) used the b-CFS paradigm to examine how English and Mandarin speakers perceive caused motion events. A language group effect emerged in the predicted direction. Unlike *path*-salient videos, *manner*-salient videos broke through suppression faster for English speakers than for Mandarin speakers. Nevertheless, the scale of the study was modest (N = 24/group), which left the need to test whether the pattern replicates. The main aim of this study, besides partial replication, was to examine low- and high-level visual processing together. To our knowledge, it is the first experimental check of how the *bottom-up* sensory signal detection compares with the *top-down* flow of language-modulated predictions about *manner* and *path*. Co-examining the top-down flow adds value to the study because it allows us to test the predictive processing account (Lupyan & Clark, 2015) in a new way. This test is at the high level of linguistically primed motion event processing. Linguistic priming through sentences is motivated by the *situation model theory* (Zwaan & Radvansky, 1998), that is, the idea that comprehenders form mental representations of a described event. A sequential presentation of descriptions and videos can tap into how motion event components, made more or less salient in language, help to form expectations during quick sentence-video mismatch recognition.

2. The present study

In response to the identified research gap, this study attempts to bring new insights into motion event research by manipulating the levels of visual processing. Two

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

experiments were designed to examine cross-linguistic effects on motion event processing in native Mandarin Chinese (an equipollently framed language) and English (a satellite-framed language) speakers. The breaking continuous flash suppression (b-CFS) (Yang et al., 2014) paradigm was used to target low-level visual processing and measure detection speed when *manner* and *path* information were manipulated. Additionally, a self-paced sentence-video verification task (inspired by Zwaan et al., 2002) was administered to examine the effects of the same type of *manner* and *path* manipulation on the high level of visual processing and measure verification speed. Detection speed and sentence-video verification speed represent two complementary on-line measures, which in a single study add to the current combinations of experimental research on linguistic relativity (Sato & Vanek, 2023).

We tested two research questions in separate experiments. On the low level of visual processing, our interest was to see if differences in the linguistic expression of motion events (*manner*-prominent in English, equipollent in Chinese) can predict differences in the speed with which motion events with *manner* and *path* manipulations get detected. If visual processing of motion events is language-entrained, English and Mandarin speakers can be expected to exhibit differences in event perception as a result of language-modulated predictions influencing sensory signal assessment (Lupyan et al., 2020). Since the *manner* is relatively more salient than *path* in English, one might expect English speakers to require a different amount of time for *manner*-salient than for *path*-salient motion to break into their awareness. With the same rationale, Mandarin Chinese speakers can be predicted to need a comparable amount of time to detect *manner*-salient and *path*-salient motion. If this cross-linguistic difference emerges on the low level of visual feature detection, one can expect amplification of this difference on the high level of sentence-video verification when the involvement of language is upregulated (Lupyan, 2012).

3. Experiment 1. Low-level visual processing of motion event through b-CFS

To test whether cross-linguistic differences in the encoding of motion events (Manner-focused English versus equipollent Mandarin) predictably influence low-level processing of motion, we asked English and Mandarin speakers to complete a motion detection task using the b-CFS paradigm, and we examined their response speed and accuracy. Our prediction was that pre-attentional biases to *manner* and *path* would differ for English and Mandarin speakers. We based our predictions on related motion event processing research in a b-CFS feasibility study (Vanek & Fu, 2023), and hypothesized that *manner*, compared to *path*, would break into awareness faster for speakers of English (*manner*-dominant) than for speakers of Mandarin (equipollent). In other words, the English group were predicted to exhibit a reaction time advantage in processing *manner*-salient compared to *path*-salient motion events, while no such difference was predicted for the Chinese group.

3.1. Participants

A total of 107 participants took part in Experiment 1, including Mandarin Chinese and English native speakers. All participants were over 18 and at the time of testing had normal or corrected-to-normal vision. A total of 55 Mandarin Chinese native speakers (21 females; mean age = 21, SD = 1) were recruited from three different

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

universities in Zhengzhou, China. Although English is compulsory for students in China, these participants were functionally monolingual in Mandarin Chinese (self-assessed). A total of 52 English native speakers (28 females, mean age = 20, SD = 2.02) were recruited in York, UK. The inclusion criterion was to be functionally monolingual in English (self-assessed).

3.2. Materials

The stimuli consisted of 48 animations in total. Each was 4.5 seconds long. The animations show 12 event scenes (example in Figure 1). The animations are co-presented with a photo sequence in four configurations or conditions, manipulated in terms of *manner, path* or *ground.* We labeled the four conditions *full match*, *manner mismatch*, *path mismatch*, and *full mismatch*, as illustrated in Figure 1. To minimize noise, each event scene within a quadruplet shares the same agent (a man) and background, the stimulus differences are constrained to path/manner/ground manipulations. Each participants viewed 36 trials (36 picture primes, 12 critical videos, and 24 filler videos). Four trials were used for practice. The total number of trials per participant is 36, rather than the full 48, to avoid repetition effects (the four conditions share one identical scene, we let each participant see one condition in each scene only). From the 36 trials, half were matches and the other half were mismatches.

The *manners* of motion included six types, namely *pulling, pushing, rolling, dragging, carrying in front,* and *carrying on the back.* The *paths* of motion also included six types, *into, out of, across, around, up*, and *down.* Apart from *manner* salience increased through *manner mismatch*, and *path* salience increased through *path mismatch*, we also included a *full match* and a *full mismatch* condition. The *full match* condition was the reference for comparisons. The *full mismatch* condition was an added control designed to check if a visual oddity, such as a dinosaur, breaks through suppression relatively faster than the stimuli in the other three conditions. The time it took to detect a stimulus was the measure of the extent to which high-level

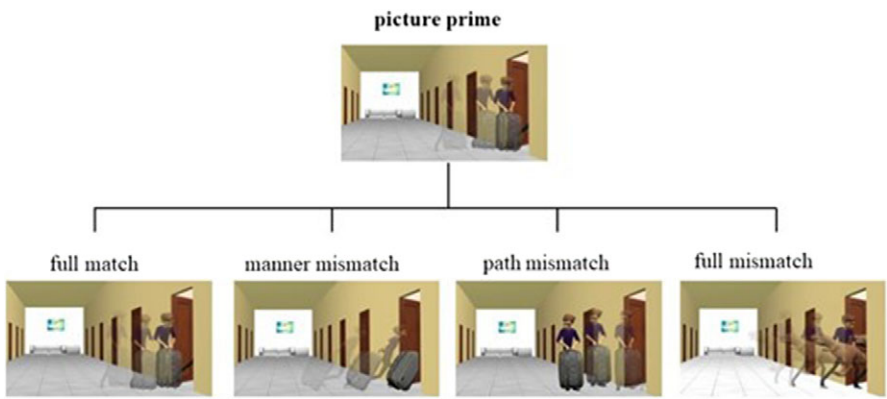

Figure 1. An example combination of a picture prime with a video quadruplet used in the b-CFS experiment. The sequence photos show (left to right) a full match between the video and the prime (a sequence picture of a man carrying a suitcase into the room), a manner mismatch (a video of a man pulling a suitcase into the room), a path mismatch (a video of a man carrying a suitcase out of the room), and a full mismatch with a visual oddity (a man carrying a dinosaur).

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

properties, including habitual linguistic encoding and familiarity, influenced perceptual suppression (Stein et al., 2011). The video stimuli in this study identical with those used in Vanek and Fu (2023), which is also a b-CFS study but without sequence pictures as primes. Adding primes in the form of sequence pictures improved the design in this study by sharpening the contrast for both *manner*-based and *path*-based mismatches. The full list of stimuli is available on the project page https://osf.io/54gse/.

The main equipment for the b-CFS experiment was a pair of mirror stereoscope goggles with four mirrors inside. The angle of the mirrors was adjustable so that the participants' vision for their left and right eye could be separated. The experiment was run in Matlab (MathWorks, Natick, MA) using the Psychophysics toolbox (Brainard, 1997).

3.3. Procedure

Before the participants arrived in the laboratory, they were asked to watch a video on YouTube (AllAboutVisionVideo, 2018), which demonstrated how to find their dominant eye. This procedure was re-checked by the researcher in the laboratory, after the participant signed the consent forms.

In the experimental process (Figure 2), each participant was seated in front of a mirror stereoscope and placed their head on a chin rest. The mirror stereoscope stood on a fixed tripod in front of a computer screen. The distance between the screen and the mirror stereoscope was 45 cm, and the distance between the participants' eyes and the mirror stereoscope was 2 cm (with slight variation across participants). The dominant eye (e.g., the left eye in Figure 2) was shown a dynamic Mondrian-like mask, which flickered at the frequency of 10 Hz. The non-dominant eye (e.g., the right eye in Figure 2) was presented with a video clip (4.5 seconds) placed randomly in one of the four corners of the screen. The contrast of the target stimulus, which was initially invisible, increased from 0% to 100% within 2.5 seconds. Initial invisibility served as an eraser of visual awareness to enable gradual perceptual enhancement as the contrast increased (Jiang et al., 2007). One participant saw 36 trials, 4 for practice and 32 as part of the main experiment. In each trial, there were two steps, and the whole b-CFS procedure took up to 20 mins. As the first step, each trial started with two white frames shown on the screen, with the aim of helping the participant adjust the mirror stereoscope goggle provided. Once they only saw one white frame on the screen, they proceeded to the b-CFS mode by pressing a key on the keyboard. Then, in the second step, a flickering screen was shown, while an animation gradually appeared in one of the four corners of the screen. The participant's task was to identify the location of the animation by pressing one of four buttons (top-left, top-right, bottom-left, or and bottom-right) corresponding to the location of the stimulus video. Stimulus locations were counterbalanced, and their order of presentation was randomized. The times taken to detect the stimuli were analyzed as indicators of suppression strength.

3.4. Results

The first point is on accuracy rates. Overall, participants identified the correct locations of the motion events with high accuracy, M = 92%, *SD* = 0.09 in the Chinese

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

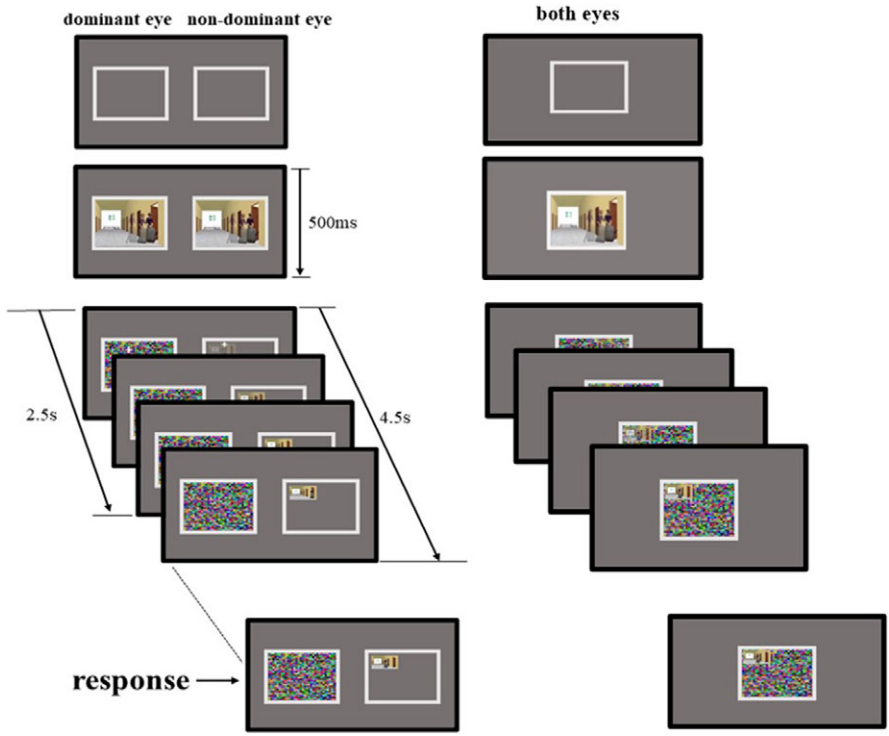

Figure 2. Illustration of the b-CFS procedure. The sequence of screens on the left demonstrates what was shown on the screen, and the one on the right shows the image that participants actually saw through the mirror stereoscope.

group, and M = 93%, *SD* = 0.06 in the English group. Data from three Chinese speakers and one English speaker were excluded due to low response accuracy (≤75%, set based on Francken et al., 2011, and Slivac et al., 2021). Further four English native speakers' data were removed list-wise because in the debrief it transpired that they also had communicative knowledge of another language. Only correct responses were included in the data analysis. Outliers, namely responses greater than ±2.5 standard deviations away from the group mean RT, were also excluded from analyses. Due to large variation, any data longer than 35 seconds were regarded as extreme values and were discounted from analyses. The 35-second cut-off was based on the scatterplot of the correct responses. In total, 119 data points were removed as outliers in the b-CFS experiment.

Reaction times were calculated from stimulus onset until button press. The average reaction times needed to correctly detect the stimulus in each condition are shown in Table 1 separately per group. The critical comparison goes in the expected direction, with the mean RT differences between path and *manner* mismatches smaller in the Chinese group (589 ms) than in the English group (1655 ms).

Figure 3 shows in greater detail that Chinese speakers needed a similar amount of time for the stimuli to break through suppression in *manner* and *path mismatch* conditions. In contrast, English speakers' data exhibit a greater difference in these two

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

Table 1. Means and SDs of the reaction times (ms) taken to detect videos by Chinese and English monolinguals across four conditions

	Mean (ms)	SD
Chinese		
Full match	8501	4469
Manner mismatch	8809	4741
Path mismatch	8220	3803
Full mismatch	8536	4791
English		
Full match	5815	4389
Manner mismatch	4707	2325
Path mismatch	6362	5039
Full mismatch	5880	4046

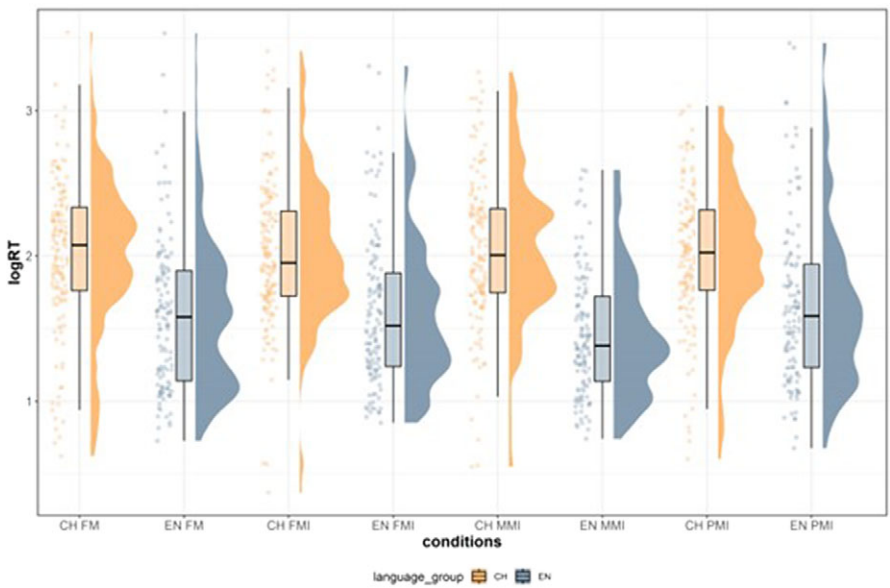

Figure 3. Log-transformed stimulus detection times in Experiment 1 shown per group (CH = Chinese, EN = English) and condition (FM = full match, FMI = full mismatch, MMI = manner mismatch, PMI = path mismatch). Box-plots show the medians and 50% of the log RTs within the boxes. Violin and raincloud plots complement visualization by showing the data distribution pattern in each condition and group.

conditions. To statistically test the effects of Condition and Language groups, as well as their interaction, we used linear mixed-effects regression through the lme4 package (Bates et al., 2015) in R (R Core Team, 2020) using the full match condition as reference. The fixed effects were Language group (2 levels) and Condition group (4 levels), both dummy-coded, the outcome variable was the log-transformed reaction times, and the random effect factors were Subject and Item. The maximal random effect structure that converged included by-subject random intercepts and by-item random intercepts. We proceeded with a forward variable selection and compared a full model with *Language group* (*lmer(logRT ~ 1 + condition* *

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

language_group + (1|subject) + (1|item)) with a reduced model excluding *Language group* (*lmer(logRT ~ 1 + condition + (1|subject) + (1|item)*). This comparison returned a significant effect of *Language group* ($\chi 2$ (4) = 46.67, $p < 0.001$), showing that the Mandarin group was overall slower. Also, a significant interaction emerged between Condition (*manner* mismatch *x* full match) and Language groups (English x Chinese) in the full model ($\beta = -0.18$, $SE = 0.06$, $t = -2.854$, $p = 0.004$; raw RT $\beta = -835$ ms).

To further explore the differences between conditions within each language group, we ran post-hoc tests using the *emmeans* package in R (Lenth, 2021). All possible comparisons were run for both groups. In the English group, RTs in the *manner mismatch* (MMI) condition were significantly shorter ($\beta = -0.19$, SE = 0.07, $t = -2.82$, $p = 0.006$; raw RT $\beta = -827$ ms) than in the *path mismatch* (PMI) condition. Also, RTs in the *manner mismatch* condition were also significantly shorter compared to those in the *full match* (FM) ($\beta = -0.16$, SE = 0.07, $t = 2.38$, $p = 0.020$; raw RT $\beta = -852$ ms) and *full mismatch* (FMI) ($\beta = -0.16$, SE = 0.07, $t = 2.30$, $p = 0.024$; raw RT $\beta = -825$ ms) conditions. However, no significant difference was found between the *manner* and *path mismatch* conditions ($\beta = 0.03$, $SE = 0.07$, $t = 0.48$, $p = 0.63$; raw RT $\beta = 1030$ ms) in the Mandarin group. No significant difference also characterized the other conditions in the Mandarin group, including *full match* versus *full mismatch* ($\beta = -0.004$ $SE = 0.07$, $t = -0.06$, $p = 0.97$; raw RT $\beta = -996$ ms), *manner mismatch* versus *full match* ($\beta = 0.02$, $SE = 0.07$, $t = -0.33$, $p = 0.75$; raw RT $\beta = 980$ ms) and also *manner mismatch* versus *full mismatch* ($\beta = 0.02$, $SE = 0.07$, $t = -0.27$, $p = 0.79$; raw RT $\beta = 980$ ms).

3.5. Discussion

Reaction time data show that *manner* salience, compared to *path* salience, provided a detection speed advantage for English speakers but not for Mandarin speakers. This finding aligns with the predicted direction of a processing speed difference reported previously for *manner* in English and Mandarin native speakers (Vanek & Fu, 2023). This study extends recent work on the effects of language on visual perception (Lupyan et al., 2020) to motion event processing. Information about *manner*, made typically prominent in English motion descriptions, was found to exert influence on a low-level process of visual feature detection during binocular rivalry (Pasley et al., 2004). This effect can be explained as the automatic recruitment of *manner* labels that increase the salience of *manner* information in visual input and thus facilitate its detection (Perry & Lupyan, 2013). In the b-CFS experiment, the time it takes to detect the change in the kind of motion functions as a low-level psychophysiological correlate of the high-level representation of *manner*. In English, where *manner* encoding is comparatively more prominent, the assessment of continuously flashing sensory signals with a gradually appearing motion event was more *manner*-based. However, in Chinese, less weight is given to *manner* information in verbal encoding, and this difference in linguistic modulation was likely to be mirrored when Mandarin speakers were detecting the emergence of motion events from visual noise. We were curious to see whether similarly predictable patterns hold on a higher level of visual processing, so we designed a separate experiment with an upregulated involvement of language.

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

4. Experiment 2. High-level processing of motion events in sentence-video verification

To further test whether differences in the linguistic encoding of *manner* (more prominent in English than in Mandarin) predictably influence the processing of motion when language is actively involved, we asked a new cohort of English and Mandarin speakers to complete a sentence-video verification task while measuring their response speed and accuracy. We predicted to observe similar patterns to those in the b-CFS experiment. Namely, we expected comparatively faster reaction times for verifications of *manner*-salient motion in English speakers than in Chinese speakers. The rationale behind the sentence-video verification experiment was to track high-level visual processing of motion with the salience of *manner* or *path* component upregulated through overt use of language. This approach has practical significance for the predictive processing account (Lupyan & Clark, 2015) as it can show how different components of motion events, emphasized or downplayed in language, contribute to the formation of expectations when quickly recognizing sentence-video (in)consistencies. Our hypothesis for the English group was derived from previous related findings (Ji, 2017); namely that English speakers should recognize sentence-video correspondences in *manner* relatively faster than sentence-video correspondences in *path*. For the Chinese group no such difference was hypothesized for *manner*- and *path*-based sentence-video correspondences. We also included *full match* and *full mismatch* conditions, in which the task was the same, to decide if the video matched the sentence they had read earlier. *Full mismatch* was expected to be the control condition for which both English and Chinese speakers would react the fastest. This idea builds on work on the influence of categories on visual processing (Lupyan et al., 2010), in which the reaction times should be shortest when the stimulus was from a different category.

4.1. Participants

One hundred and two participants were recruited in Experiment 2 in two new cohorts of Mandarin Chinese and English native speakers. The same inclusion criteria applied, to be over 18 and at the time of testing have normal or corrected-to-normal vision. A total of 51 Mandarin Chinese native speakers (28 females; mean age = 21.30, SD = 2.90) were recruited in Zhengzhou, China. All were functionally monolingual in Mandarin Chinese (self-assessed). A total of 51 English native speakers (34 females, mean age = 19.49, SD = 1.23) were recruited in York, UK. In this group, all were functionally monolingual in English (self-assessed).

4.2. Materials

The video materials were identical with those used in the b-CFS experiment. However, instead of having to match sequential pictures with the target videos, the SV experiment used sentences as primes. An example sentence quadruplet in English/Mandarin is, '*A man is rolling a log towards a cabin*'/'*Yi1 ge2 nan2 ren2 zheng4 zai4 ba3 yi2 ge4 mu4 zhuang1 gun3 xiang4 yi1ge4 xiao3 mu4 wu1*' used for the full match condition, '*A man is carrying a log towards a cabin*'/'*Yi1 ge2 nan2 ren2 zheng4 zai4 ba3 yi2 ge4 mu4 zhuang1 ban1 xiang4 yi1ge4 xiao3 mu4 wu1*' for *manner* mismatch, '*A man is rolling a log away from a cabin*'/'*Yi1 ge2 nan2 ren2 zheng4 zai4*

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

ba3 yi2 ge4 mu4 zhuang1 gun3 li2 yi1ge4 xiao3 mu4 wu1' for *path* mismatch, and '*A man rolled a dinosaur towards a cabin*'/*Yi1 ge2 nan2 ren2 zheng4 zai4 ba3 yi2 ge4 kong3 long2 gun3 xiang4 yi1ge4 xiao3 mu4 wu1'* for full mismatch. The role of the prime sentences was to enable the participants to create mental situation models (Zwaan & Radvansky, 1998) against which to compare the motion event presented in the subsequent video. In Experiment 2, each participant was presented 36 trials (36 sentence primes, 12 critical videos and 24 filler videos) out of which four trials were used for practice. For the ease of comparability across the study, we kept the condition labels identical with those in Experiment 1. As the critical dimension in Experiment 2 is sentence-video correspondence, it is important to note that in the condition labeled *manner mismatch* the sentence and the video corresponded in *path*, and the condition labeled *path mismatch* equals sentence-video correspondence in *manner*. In all videos, *manner* and *path* information was available simultaneously from the very beginning of the action. This is a methodological strength ruling out the possibility of response times having been influenced by motion component sequencing in the visual input.

The self-paced sentence-video (SV) verification experiment was run in Matlab (MathWorks, Natick, MA) using the Psychophysics toolbox (Brainard, 1997) to collect data from the Mandarin Chinese speakers. We used the Gorilla Experiment Builder (www.gorilla.sc) to create and host our experiment (Anwyl-Irvine et al., 2018) to collect data from the English speakers. The collected RTs were log-transformed to reduce skewness in a non-normal distribution. We followed this step as a standard statistical procedure used in related studies measuring RTs to track low-level motion perception (e.g., Slivac et al., 2021) as well as high-level visual processing of motion events (e.g., Sakarias, 2019).

4.3. Procedure

Two main steps characterize the SV experiment (Figure 4). After a fixation cross, participants saw one sentence presented in their native language on the computer screen, describing a motion event. After reading the sentence, they pressed the space bar to continue. Upon button press, their task was to carefully watch a video which either matched or did not match the sentence they had read before. Their task was to press the left arrow (labeled 'YES') to indicate a match, or right arrow (labeled 'NO') to indicate a mismatch. The video played in a loop until they made their decision. Stimulus presentation followed a randomized order.

4.4. Results

We first examined accuracy rates. In general, participants verified sentence-video matches and mismatches with high accuracy, M = 88.13%, SD = 0.05 in the Chinese group, and M = 90.36%, SD = 0.08 in the English group. Data from three Chinese speakers and one English speaker were excluded due to low response accuracy (≤ 75%). Other English native speakers' data were removed because it came to light during the debrief that they were also fluent in another language. Only correct responses fitting within ±2.5 standard deviations from the group mean RT were included in the data analysis. Due to large variation, any data longer than 15 seconds were regarded as extreme values and were discounted from analyses. The 15-second

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

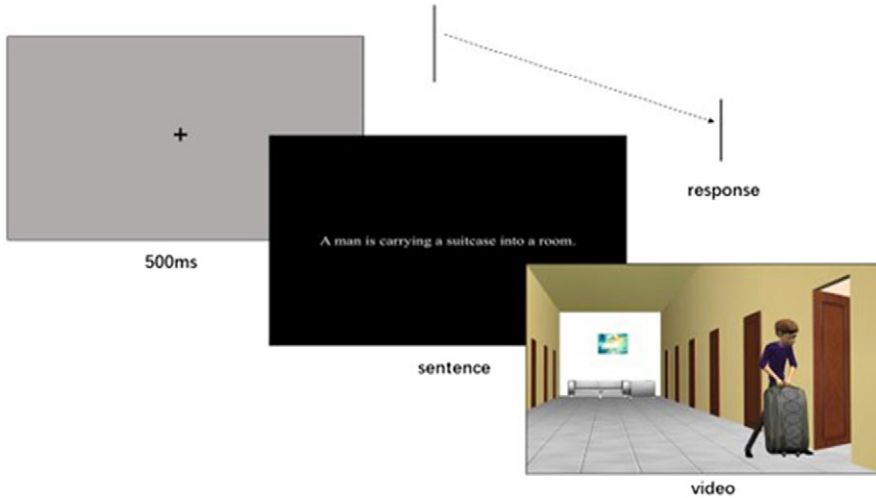

Figure 4. Illustration of the sentence-video verification procedure. A fixation cross is followed by a sentence prime (self-paced) and subsequently by a video.

Table 2. Means and SDs of the reaction times (ms) taken to recognize (mis)matches by Chinese and English monolinguals across the four conditions

	Mean (ms)	SD
Chinese		
Full match	3341	1606
Manner mismatch	3575	1530
Path mismatch	3425	1718
Full mismatch	2891	1714
English		
Full match	2280	1205
Manner mismatch	2770	1452
Path mismatch	2039	1305
Full mismatch	1633	834

cut-off was based on the scatterplot of the correct responses. In total, 78 data points were removed as outliers in the SV experiment.

Reaction times were calculated from video onset until button press. The average RTs needed to correctly verify sentence-video (mis)matches in each condition are shown in Table 2. The full mismatches were recognized fastest in each group as predicted. The key between-group difference concerns the mean RT differences between *path* and *manner* mismatches, smaller in the Chinese group (150 ms) than in the English group (731 ms), but in the opposite direction than predicted. Unexpectedly, *manner* mismatches took *longer* to recognize than *path* mismatches in the English group.

Figure 5 presents the overlaps and differences between the two groups' sentence-video verification times in finer detail. It shows that Chinese speakers took roughly the same to correctly verify mismatching sentence-video pairs in *manner* and *path mismatch* conditions. Greater differences emerged for the English speakers between these two conditions, albeit the directionality of this difference was unexpected. To

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

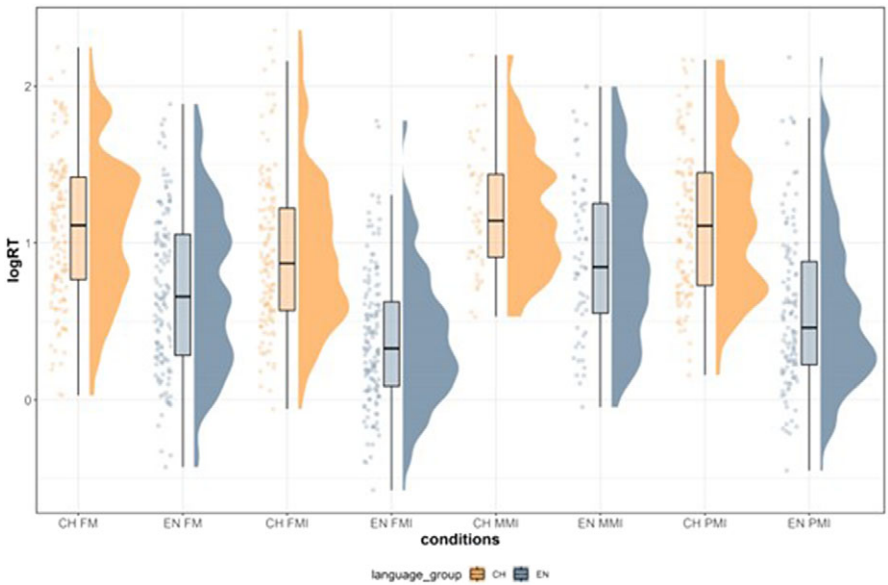

Figure 5. Log-transformed sentence-video verification times for the correct responses in Experiment 2 shown per group (CH = Chinese, EN = English) and condition (FM = full match, FMI = full mismatch, MMI = manner mismatch, PMI = path mismatch). Box-plots show the medians and 50% of the log RTs within the boxes. Violin and raincloud plots complement visualization by showing the data distribution pattern in each condition and group.

statistically test the effects of Condition and Language groups, and their interaction, we built linear mixed-effects regression models following the process in Experiment 1. Here too, we report the maximal random effect structure that converged, that is with random slopes over condition by subject and random slopes over group by item. We compared a full model with *Language group* (*lmer(log(RT) ~ 1 + condition * language_group + (1 + condition|subject) + (1 + language_group|item)*) with a reduced model excluding group *lmer(log(RT) ~ 1 + condition + (1 + condition|subject) + (1 + language_group|item)*). This comparison returned a significant effect of *Language group* ($\chi 2$ (4) = 44.89, $p < 0.001$), which shows that, on the whole, Mandarin speakers took longer to respond.

To further investigate the differences within each language group, post-hoc analyses were conducted using the emmeans package in R (Lenth, 2021). The English group took significantly longer to verify *manner* mismatches than *path* mismatches ($\beta = -0.26$, $SE = 0.13$, $t = 2.08$, $p = 0.043$; raw RT $\beta = -771$ ms).

However, no significant difference was found between *manner* (MMI) and *path mismatch* (PMI) conditions in the Chinese group. Also, RTs of both language groups in the *full mismatch* (FMI) condition were significantly shorter compared to those in the *full match* (FM) condition ($\beta = -0.18$, SE = 0.08, $t = -2.27$, $p = 0.028$; raw RT $\beta = -835$ ms) When checked separately, English speakers were significantly faster in the *full mismatch* condition than in the *full match* ($\beta = -0.32$, $SE = 0.11$, $t = -2.91$, $p = 0.005$; raw RT $\beta = -762$ ms), *manner mismatch* ($\beta = -0.49$, SE = 0.12, $t = -4.00$, $p < 0.001$; raw RT $\beta = -613$ ms), as well as the *path mismatch* condition ($\beta = -0.23$, $SE = 0.11$, $t = -2.13$, $p = 0.039$; raw RT $\beta = -795$ ms). The same holds for the Chinese

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

group, who were significantly faster in the *full mismatch* condition than in the *full match* ($\beta = -0.18$, $SE = 0.08$, $t = -2.19$, $p = 0.033$; raw RT $\beta = -835$ ms), *manner mismatch* ($\beta = -0.28$, $SE = 0.10$, $t = -2.81$, $p = 0.007$; raw RT $\beta = -756$ ms), and *path mismatch* condition ($\beta = -0.21$, $SE = 0.08$, $t = -2.66$, $p = 0.011$; raw RT $\beta = 811$ ms). However, in the Chinese group there was no significant difference between *manner mismatch* and *path mismatch* ($\beta = 0.064$, $SE = 0.1$, $t = 0.64$, $p = 0.524$; raw RT $\beta = 1066$ ms).

4.5. Discussion

Sentence-video verification times showed that *manner* prominence typical of English motion verbalization differentially affected the processing of *manner*-based versus *path*-based mismatches. However, *manner-path* equipollence, characteristic of motion encodings in Mandarin Chinese, did not lead to pronounced processing differences between *manner*-based versus *path*-based mismatches. The finding that English speakers took longer to decide about *manner*-based than *path*-based mismatches may seem puzzling, but only at the first glimpse. Lower-level processes like stimulus detection rest on a contrast spotting ability, where mismatches may matter more. High-level processes like sentence-video verification involve verbally primed semantic analysis, for which *correspondence* between the key motion component facilitates high-level decisions (Ji, 2017). In other words, in English, where *manner* encoding is comparatively more prominent, the upregulation of *manner* through a linguistic prime led to a faster assessment of *path* mismatches, which through a lens of a sameness check were essentially *manner*-matches. Cross-linguistic difference in the speed of *manner*-matches faster in English than in Mandarin speakers is in line with earlier findings testing high-level processing of caused motion (Ji, 2017). This effect can be explained as rapid, linguistically induced upregulation of *manner* labels (Lupyan, 2012) that facilitate the high-level cognitive process of a sentence-video match verification. However, this account can just partly explain the results since being able to quickly establish a *manner* match is insufficient information when the task is to press 'NO' for mismatches. *Path* must have been processed before button presses too. In the present design, the temporal distance between *path* information presented first linguistically and then visually was shorter than that between the two sources of *manner* information, so it is possible that *path* mismatch verifications were (co-)supported by a recency effect. This temporal distance account may be intuitively appealing, but it cannot fully embrace the observed cross-linguistic differences. Instead, what we find more powerful in this respect is the predictive processing account, which involves language-specific expectation optimization to quickly recognize sentence-video mismatches. We elaborate on this point in the General Discussion.

5. General discussion

5.1. On how detecting motion and judging sentence-video matches become more manner-sensitive

Two main innovative contributions emerged from this study to inform motion event cognition research. First, cross-linguistic effects of differential encoding of *manner* were observed in a prelinguistic time-window of visual feature detection. Differences

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

in detection speed when *manner*-salient motion break through suppression into awareness aligned with language-driven hypotheses based on encoding *manner* and *path* in Mandarin Chinese and English. This finding confirms that cross-linguistic effects on motion event processing are nontrivial, resilient, and exerting influence already when perceptual signals are detected. In the words of Casasanto et al. (2004), 'language can shape even primitive, low-level mental processes' (p. 575). We interpret this influence within the framework of *predictive processing* (Lupyan & Clark, 2015). The underlying mechanism of the observed cross-linguistic effect on the speed of *manner* detection within this account is explained as a *downward flow* of language-modulated predictions about visual signals that emerge through suppression. Greater weight of *manner* in the linguistic encoding of motion events in English contributes to the formation of English speakers' predictions to be relatively more *manner*-based than is the Chinese speakers' predictions. Without recruiting language consciously, when visual percepts mismatch in *manner*, English speakers may detect it faster because of increased relevance of *manner* information involved in their expectations about the incoming sensory data. These downward predictions are likely to differ in Chinese speakers, whose anticipatory mechanism is more likely to combine emerging visual stimuli with equally weighed *manner*-based and *path*-based expectations. In sum, variation in the ways of talking about motion in Chinese and English are built into predictions which can influence how motion events are perceived.

Second, findings from the sentence-video verification task were also in line with the related prediction, and confirmed that English and Chinese speakers differed in their processing of *manner* and *path* information at the high level, when verifying sentence-video correspondences. Although the Mandarin speakers were overall slower than the English speakers (and that holds for both experiments), the critical comparison to monitor processing patterns of *manner* and *path* was *within* language group, and only the differential scores were looked at cross-linguistically. Greater difference in processing *manner*-based versus *path*-based stimulus pairs in the English group than in the Mandarin group indicates that language cues, in this case the relatively greater prominence of *manner* in English motion event expressions, acted as important building blocks of situation models (Zwaan & Radvansky, 1998) against which the subsequent motion videos were compared. Considering the Talmy's typology, an E-language like Mandarin and an S-language like English might not be the ideal examples of a contrastive encoding of *manner*, but they could still predict the observed between group differences. In extension of the present study, one could expect more pronounced between-group differences with a design that compares speakers of a V-language like Spanish or Japanese with speakers of a S-language like English.

5.2. On the asymmetry in reactions when manner is suppressed or verbally highlighted

The asymmetry in processing speed in response to stimuli manipulated for *manner* and *path* information, which emerged in the English group across the two experiments, is perhaps the most intriguing result. More specifically, while *manner* mismatches took English speakers less time than *path* mismatches to detect in the b-CFS experiment, they took longer to judge in the self-paced sentence-video verification experiment. This asymmetric pattern points to the involvement of very different mechanisms guiding English speakers' low-level visual feature detection process and

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

a high-level semantic verification process. Within the framework of predictive processing, the manipulation of *manner* information was detected more quickly than the manipulation of *path* information because of a sharper contrast with default *manner*-based predictions when incoming sensory signals were breaking through suppression. More contrastive signals are less predictive input that travels up the visual hierarchy and assists learning as it triggers an update of subsequent predictions. This way, less expected input helps to refine further precision estimates. However, in sentence-video verification, the situation differs in that the sentences mentally activated the matching upcoming videos, and information that was linguistically primed got recognized as a corresponding visual match faster.

Another surprising result in the present study was that in the b-CFS experiment, the time English speakers needed to detect stimuli in the *manner* mismatch condition was the shortest compared to all the other conditions. A possible reason can be found in *salience theory* (Slobin, 2006), which advocates that on top of differentiating languages based on the main verb that expresses a motion event, languages can be further divided into how they capture the levels of salience in *manner* information. *Manner* salience in English expressions of motion is high, making *manner* stand out. *Manner*-based top-down predictions about what kind of motion to expect to show through suppression quickly alerted the participants and as soon as this expectation broke, allowing English participants to react to *manner*-based mismatches faster. The relevance of *manner* could then have carried over to other trials of a similar type. In Feist and Férez's (2013) words, 'higher codability of manner of motion correlates with improved memory for manner' (p. 398). However, no such *manner* mismatch detection advantage was found in the Chinese group, as they processed all types of mismatches with similar speed. In sum, these b-CFS results suggest that low-level visual processing of caused motion is also connected to the levels of salience in their corresponding linguistic expressions.

On another level of comparison between conditions and experiments, both the English and the Mandarin group took similar amounts of time to detect the full matches (prime picture identical with the video breaking through suppression) and the full mismatches (picture and video with an distinctively different object) in the b-CFS experiment. This was not the case in the sentence-video verification experiment, where both the English and the Mandarin speakers needed less time to verify the *full mismatches* where a conspicuously different object (a dinosaur) was moved. This variation in results of the two tasks further demonstrates that it was not the conspicuous object that led to between-group differences in breakthrough times in the b-CFS experiment. Instead, it was the mismatches in the *manner* of motion driving the detection time differences, while the object caused to move was relatively more suppressed. To confirm the functionality of the flickering mask for stimulus suppression, a small-scale validity check was run in the preparation stage. For the validity check prior to Expt. 1 and Expt. 2, ten participants (mean age = 19.1, SD = 0.74; 6 females) other than those taking part in the main experiments were asked to complete a detection task. The purpose of the validity check, a routine exercise for b-CFS experiments to test mask functionality, was to ensure that through the changes in contrast shown to the participants we were able to regulate the time delay when motion events became visible. The same stimuli as in the b-CFS experiment were used, and each participant viewed half of the stimuli with a flickering mask and the other half without a mask, with random order and stimulus counterbalancing in place. The average accuracy rate in the validity check was 0.88

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

(SD = 0.10). The average suppression time with a mask was 8113 ms (SD = 4390 ms), and without a mask, it was 5730 ms (SD = 1351 ms). A significant main effect of mask ($\chi 2$ (1) = 9.31, p = 0.02) was found when we compared a mixed-effects model with a mask included and a reduced model with a mask excluded. These results showed that the mask worked, in other words that it was sufficiently powerful to suppress the stimuli during the b-CFS experiment.

To contextualize our findings, the roughly equal processing pattern and the *manner*-biased processing pattern found in Mandarin and English speakers, respectively, were shared across the two levels tested in this study. The observed patterns align with previous work examining how *manner* and *path* processing differs between Mandarin/Cantonese and English speakers (Ji, 2017; Wang & Wei, 2021). These two studies converged in showing that *manner* and *path* information in similarity judgement contexts were processed in group-specific ways. English speakers took longer to make judgements when *manner* differed compared to when *path* differed. The explanation provided was that English speakers processed the *manner* and *path* information in a sequence, whereas Mandarin speakers processed both motion components in parallel (Ji, 2017). While the validity of this account would need to stand the test of sequential and simultaneous presentation of *manner* and *path* information in a fully crossed design, it holds that the present study contributes to the extant body of literature confirming that caused motion event processing in English speakers is more *manner*-based than in Mandarin speakers. The specific contribution comes from two levels of processing. First, applying the b-CFS paradigm to motion event research proved advantageous to document that language-specific encodings of motion are automatically (on the low level) utilized to make predictions about the incoming sensory input. Second, applying the sentence-picture verification paradigm helped to clarify that language-specific encodings of motion operate differently when they are consciously (on the high level) employed during sematic analyses that are primed by verbal cues.

At the high level of linguistically primed motion event processing, English speakers found it harder to make verifications when *manner* differed compared to when *path* differed. If the relativistic view holds, faster verifications of *path* mismatches may appear counter-intuitive at first because English typically encodes *manner* (Talmy, 2000) rather than *path* in the main verb. In the wider context, the critical reader may find the *path* mismatch advantage for English speakers evocative of what could be dubbed 'a reversed Whorfian effect' observed in Papafragou et al. (2008). Let us recall that Papafragou et al. (2008) found English speakers' late eye-fixations focusing on the *path*, which reflects the component of motion *not* typically expressed in the main verb. English speakers' increased attention allocated to *path* is a signal of greater cognitive effort not assisted by (overt or covert use of) the native language when visually inspecting scenes in preparation for a memory task. The relative increase in attention to *path*, when compared to a group of Greek speakers, resulted in English speakers' more accurate recognition of the *path*-endpoint object. Not quite convinced by the universalist interpretation advocated in the original article, one could see this 'reversed Whorfian effect' as a signal of language-specific differences in predictive processing (Lupyan & Clark, 2015), which aligns with the relativistic view. More specifically, English speakers' late fixations launched to an event component that the main verbs do not readily encode could work as optimization of dissimilar weights when visually presented *manner* and *path* details are integrated to tune predictions for maximized success in an upcoming memory task.

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

What the motion event recall in Papafragou et al.'s (2008) shares with this study's sentence-video mismatch verification is the relatively generous time-window during which the motion component *not* made salient via verbal encoding gained a processing advantage. English speakers' shorter reaction times in *path* mismatches can also be explained as an optimizer of weights for the visually salient but linguistically underfocused *path* information that needs to be considered for successful task completion. Such optimization rests on the idea that the predictive mind estimates the uncertainty of the incoming visual data and tunes predictions based on both the upregulated *and* the underfocused information in the linguistic prime.

5.3. On event integration theory, limits, and future directions

The findings of both experiments are of immediate relevance to Talmy's typology of event integration, distinguishing S-languages and V-languages depending on which sentence constituents encode semantic notions such as *manner* and *path* (2000:117), and also to the expanded typology by Slobin (2004) adding E-languages such as Mandarin, Jaminjung (Schultze-Berndt, 2000) and Thai (Zlatev & Peerapat, 2004) to the taxonomy. The present study brings new evidence that the inequivalence of encoding *manner* in English and Mandarin predictably influences how top-down information affects not only high-level semantic but also low-level perceptual processing of motion, with speaker variation groupable by the type of *manner* encoding in their native language. *Manner* plays a more prominent role than *path* in motion event processing of English speakers, whether language during the task had a chance to be recruited consciously or not. For Chinese speakers, however, *manner* and *path* are of relatively more equal weight, both in forming predictions about sensory input and in building situation models about caused motion. These explanations combine well with the predictive processing account (Lupyan & Clark, 2015), and they also complement the broader context of extant event cognition research with Mandarin versus English speakers (e.g., Tang et al., 2021; Zhang & Vanek, 2021).

Regarding motion types, this study could choose between voluntary or caused motion (Talmy, 1985). One option was to focus on motion events involving an agent that moves along a *path* in a specific *manner,* known as spontaneous or *voluntary motion* (as in *the cat is crawling under the sofa*). The other option was to choose motion events in which an agent displaces an object in a specific *manner* so that the object moves along a *path,* known as cause motion. We chose caused motion for this study to increase its relevance for, and connectivity, with a growing body of experimental work using this motion event type (Ji et al., 2011b; Ji & Hohenstein, 2014a, 2014b; Montero-Melis & Bylund, 2017; Tusun, 2023; Tusun & Hendriks, 2022; Wang & Wei, 2021). Another motivation was to partially replicate and enhance recent experimental work on the early stages of caused motion processing across Mandarin versus English speakers (Vanek & Fu, 2023), verifying the suitability of b-CFS as a method that can probe into the automatic processing of dynamic stimuli. Much of previous research using the b-CFS paradigm used static pictures as stimuli (e.g., photos of simple objects like a pumpkin in Lupyan & Ward, or photos of faces in different positions in Jiang et al., 2007). However, to examine motion event processing with static images would tap into inference rather than perception, and for this study the latter was key. Also, filtering out and manipulating the target components of motion events, more specifically, trying to increase and decrease the salience of

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

manner and *path* information across static stimuli, would have been problematic. Was the effectiveness of b-CFS compromised by fine distinctions in dynamic stimuli (Pournaghdali & Schwartz, 2020)? Between-group differences at a low level of visual detection aligning with the inequivalence of *manner* in its linguistic encoding in English and Mandarin showed that b-CFS is applicable as a method to test preverbal processing of motion events. In the future, studies might benefit from an even tighter control over non-motion related aspects, such as colour (as, e.g., in Slivac et al., 2021).

Another improvement in future work could be through an increase in the number of the items, perhaps even with an added measure of graded contrast in *manner* and *path* mismatches. Among other potential limitations is the comparability of response times between groups. Given that the Mandarin speakers took longer than English speakers to respond across all conditions in both experiments (as also found in Vanek & Fu, 2023; Zhang & Vanek, 2021), it is possible that variations in cultural background or other non-linguistic factors could have influenced the results. For example, it is possible that Mandarin speakers, influenced by cultural factors, may have prioritized confidence in their responses over speed. Emphasis on precision could potentially result in null effects across some conditions. For this reason, it was reassuring to see that full mismatches generated the fastest responses in both experiments as expected, a result which aids the validity of the theoretically motivated argument about language specificity in visual processing. Furthermore, to fully appreciate the contribution of b-CFS to motion event processing research, methodological triangulation via EEG, an approach that does not rely on button presses, could serve as an independent validity check of automaticity in motion event processing.

6. Conclusion

This study investigated the interplay between cross-linguistic differences in how *manner* and *path* are expressed and how motion events are processed, comparing two levels of processing. We targeted low-level visual processing through blind continuous flash suppression to examine detection speed, and high-level visual processing with linguistic involvement to look at sentence-video verification speed. The results confirmed the hypothesis that Mandarin and English native speakers differ in the processing of *manner*-salient motion events in language-specific ways, both when language is recruited automatically for predictions about upcoming sensory input, and also when it is used consciously during verbally primed semantic analyses.

Data availability statement. The data and code used in the analyses, together with the model outputs, are available at https://osf.io/54gse/.

Competing interest. The authors declare none.

References

AllAboutVisionVideo. (2018, October 24). *How to Determine Your Dominant Eye with Our Dominant Eye Test* [Video]. YouTube. https://www.youtube.com/watch?v=4Gbkca4RM-4

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407. https://doi.org/10.3758/s13428-019-01237-x

https://doi.org/10.1017/langcog.2023.37 Published online by Cambridge University Press

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish Speakers. *Cognitive Science*, 37(2), 286–309. https://doi.org/10.1111/cogs.12006

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, 26(4), 518–526. https://doi.org/10.1177/0956797614567509

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.

Casasanto, D. (2016). Linguistic relativity. In N. Riemer (Ed.), *Routledge handbook of semantics* (pp. 158–174). Routledge.

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., Santiago-Diaz, I., Fotokopoulu, O., Pita, R., & Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 26(26), 575–580.

Chen, L., & Guo, J. (2009). Motion events in Chinese novels: Evidence for an equipollently-framed language. *Journal of Pragmatics*, 41(9), 1749–1766. https://doi.org/10.1016/j.pragma.2008.10.015

Feist, M. I. (2010). Motion through syntactic frames. *Cognition*, 115(1), 192–196. https://doi.org/10.1016/j.cognition.2009.11.011

Feist, M. I. (2016). Minding your manners: Linguistic relativity in motion. *Linguagem Em (Dis)Curso*, 16, 591–602. https://doi.org/10.1590/1982-4017-160305-0916D

Feist, M. I., & Férez, P. C. (2013). Remembering how: Language, memory, and the salience of manner. *Journal of Cognitive Science*, 14(4), 379–398.

Flecken, M., Athanasopoulos, P., Kuipers, J. R., & Thierry, G. (2015). On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition*, 141, 41–51. https://doi.org/10.1016/j.cognition.2015.04.006

Flecken, M., Stutterheim, C. V., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross- linguistic non-verbal recognition task. *Language and Cognition*, 6 (1), 45–78. https://doi.org/10.1017/langcog.2013.2

Francken, J. C., Kok, P., Hagoort, P., & De Lange, F. P.. (2011). The behavioral and neural effects of language on motion perception. *Journal of Cognitive Neuroscience*, 27, 175–184. https://doi.org/10.1162/jocn

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83(1), 49–79. https://doi.org/10.1016/S0010-0277(01)00166-4

Gleitman, L., & Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.), *Handbook of cognitive psychology* (pp. 633–661). Cambridge University Press.

Ji, Y. (2017). Motion event similarity judgments in one or two languages: An exploration of monolingual speakers of English and Chinese vs. L2 learners of English. *Frontiers in Psychology*, 8, 909. https://doi.org/10.3389/fpsyg.2017.00909

Ji, Y., Hendriks, H., & Hickmann, M. (2011a). How children express caused motion events in Chinese and English: Universal and language-specific influences. *Lingua*, 121(12), 1796–1819. https://doi.org/10.1016/j.lingua.2011.07.001

Ji, Y., Hendriks, H., & Hickmann, M. (2011b). The expression of caused motion events in Chinese and in English. *Some Typological Issues*, 49(5), 1041–1077. https://doi.org/10.1515/ling.2011.029

Ji, Y., & Hohenstein, J. (2014a). The syntactic packaging of caused motion components in a second language: English learners of Chinese. *Lingua*, 140, 100–116. https://doi.org/10.1016/j.lingua.2013.11.009

Ji, Y., & Hohenstein, J. (2014b). The expression of caused motion by adult Chinese learners of English. *Language and Cognition*, 6(4), 427–461. https://doi.org/10.1017/langcog.2014.4

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science*, 18(4), 349–355. https://doi.org/10.1111/j.1467-9280.2007.01902.x

Lenth, R. (2021). emmeans: Estimated marginal means, aka least-squares means. R Package Version 1.7.1-1. https://CRAN.R-project.org/package=emmeans

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, 3, 54. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, 24(4), 279–284. https://doi.org/10.1177/0963721415570732

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, 24(11), 930–944. https://doi.org/10.1016/j.tics.2020.08.005

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological science*, 21(5), 682–691. https://doi.org/10.1177/0956797610366099

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, 110(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

Montero-Melis, G., & Bylund, E. (2017). Getting the ball rolling: The cross-linguistic conceptualization of caused motion. *Language and Cognition*, 9(3), 446–472. https://doi.org/10.1017/langcog.2016.22

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements *Cognition*, 108(1), 155–184. https://doi.org/10.1016/j.cognition.2008.02.007

Pasley, B. N., Mayes, L. C., & Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron*, 42(1), 163–172. https://doi.org/10.1016/S0896-6273(04)00155-2

Perry, L. K., & Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience*, 7, 122. https://doi.org/10.3389/fnbeh.2013.00122

Pournaghdali, A., & Schwartz, B. L. (2020). Continuous flash suppression: Known and unknowns. *Psychonomic Bulletin & Review*, 27(6), 1071–1103. https://doi.org/10.3758/s13423-020-01771-2

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Sakarias, 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(1), e12708. https://doi.org/10.1111/cogs.12708

Sato, S., & Vanek, N. (2023). Contrasting online and offline measures: Examples from experimental research on linguistic relativity. In S. Zufferey & P. Gygax (Eds.), *The Routledge handbook of experimental linguistics* (pp. 217–234). Routledge. https://doi.org/10.4324/9781003392972-17

Schultze-Berndt, E. (2000). *Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language*. MPI Series in Psycholinguistics, vol. 14. Ponsen and Looijen.

Slivac, K., Hervais-Adelman, A., Hagoort, P., & Flecken, M. (2021). Linguistic labels cue biological motion perception and misperception. *Scientific Reports*, 11, 17239. https://doi.org/10.1038/s41598-021-96649-1

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In L. Verhoeven & S. Stromqvist (Eds.), *Typological and contextual perspectives* (pp. 219–257). Psychology Press.

Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Roberts (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59–82). Benjamins.

Stein, T., Hebart, M. N., & Sterzer, P. (2011). Breaking continuous flash suppression: A new measure of unconscious processing during interocular suppression? *Frontiers in Human Neuroscience*, 5(167), 1–17. https://doi.org/10.3389/fnhum.2011.00167

Stein, T. (2019). The breaking continuous flash suppression paradigm: Review, evaluation, and outlook. In G. Hesselmann (Ed.), *Transitions between consciousness and unconsciousness*, (pp. 1–38). London/New York: Routledge.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Grammatical categories and the lexicon. Language typology syntactic description* (Vol. 3, pp. 57–149). Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics*. MIT Press.

Tang, M., Vanek, N., & Roberts, L. (2021). Crosslinguistic influence on English and Chinese L2 speakers' conceptualization of event series. *International Journal of Bilingualism*, 25(1), 205–223. https://doi.org/10.1177/1367006920947174

Tusun, A. (2023). Uyghur-Chinese early successive adult bilinguals' construal of caused motion events. *Language and Cognition*, 13, 1–26. https://doi.org/10.1017/langcog.2023.7

Tusun, A., & Hendriks, H. (2022). Caused motion events in Modern Uyghur: A typological perspective. *Linguistics*, 60(5), 1663–1705. https://doi.org/10.1515/ling-2020-0098

Vanek, N., & Fu, X. (2023). Low-level visual processing of motion events as a window into language-specific effects on perception. *International Review of Applied Linguistics in Language Teaching*, 61, 61–78. https://doi.org/10.1515/iral-2022-0048

Wang, Y., & Wei, L. (2021). Cognitive restructuring in the multilingual mind: Language-specific effects on processing efficiency of caused motion events in Cantonese–English–Japanese speakers. *Bilingualism: Language and Cognition*, 24(4), 730–745. https://doi.org/10.1017/S1366728921000018

Whorf, B. (1956). *Language, thought, and reality*. MIT Press.

Yang, E., Brascamp, J., Kang, M.-S., & Blake, R. (2014). On the use of continuous flash suppression for the study of visual processing outside of awareness. *Frontiers in Psychology*, 5, 724. https://doi.org/10.3389/fpsyg.2014.00724

Zhang, H., & Vanek, N. (2021). From "No, she does" to "Yes, she does": Negation processing in negative yes–no questions by Mandarin speakers of English. *Applied Psycholinguistics*, 42(4), 937–967. https://doi.org/10.1017/S0142716421000175

Zlatev, J., & Peerapat, Y. (2004). A third way to travel. The place of Thai in motion-event typology. In S. Strömqvist & L. Verhoeven (Eds.), *Relating events in narrative. Typological and contextual perspectives* (pp. 159–190). Lawrence Erlbaum.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168–171. https://doi.org/10.1111/1467-9280.0043

Cite this article: Fu, X., Vanek, N., & Roberts, L. (2024). Matched or moved? Asymmetry in high- and low-level visual processing of motion events, *Language and Cognition* 16: 283–306. https://doi.org/10.1017/langcog.2023.37