Research Article

Norbert Vanek* and Xingyi Fu

Low-level visual processing of motion events as a window into language-specific effects on perception

https://doi.org/10.1515/iral-2022-0048
Received February 28, 2022; accepted December 7, 2022; published online December 20, 2022

Abstract: This article brings a new perspective to the currently burgeoning interest in the power of language to influence how speakers from different linguistic backgrounds process motion events. While many studies have targeted high-level decision-based processes, such as Manner-based versus Path-based categorisation or motion event similarity judgments from memory, far less is known about the role of various language systems on low-level automatic processing. The goal of this article is to present an experimental method called *breaking continuous flash suppression* (b-CFS), critically assess its potential to capture language-induced biases when processing motion through a small-scale feasibility study with English native speakers versus Mandarin native speakers, and to provide practical recommendations with examples of how motion event research can respond to the epistemological challenges that this emerging data elicitation method faces.

Keywords: breaking continuous flash suppression; linguistic influence on perception; low-level visual processing; manner and path; motion events

The rationale behind b-CFS is that continuous flashes suppress visual input and thus disrupt high-level (semantic or conceptual) processing. The time it takes to detect a stimulus indicates its level of difficulty to break into awareness (Gayet et al. 2014). Perceptually more salient stimuli break through suppression faster than less salient ones. Crucially, if crosslinguistic differences in motion event encoding give rise to variations in Manner and Path saliency during visual perception, one might expect

***Corresponding author: Norbert Vanek**, School of Cultures, Languages and Linguistics, The University of Auckland, 18 Symonds Street, Auckland, 1142, New Zealand; and Experimental Research on Central European Languages Lab, Charles University, Prague, Czech Republic,
E-mail: norbert.vanek@auckland.ac.nz. https://orcid.org/0000-0002-7805-184X
Xingyi Fu, Centre for Research in Language Learning and Use, University of York, York, North Yorkshire, UK

predictable language-specific influence on the time in which specific motion components (Manner/Path) enter awareness. For example, if satellite-framed languages like English provide a boost to the perceptual strength of Manner information compared to verb-framed languages like Spanish, variations in Manner can be predicted to emerge into awareness faster for English than for Spanish native speakers. If breakthrough time differences are linguistically modulated, b-CFS presents an optimal enrichment to current measures of automatic processing in motion event research on crosslinguistic typological contrasts.

1 Introduction

Imagine you are looking at a monitor with continuously flashing high-contrast white noise. This white noise supresses a motion event, say a jumping frog, that is initially invisible in the background, but then it gradually gains intensity up to the point when one of your eyes can see it. Your task is to press a button as soon as you notice motion in one of the corners of the monitor. The moment when the motion event breaks through the visual noise is taken as the point when it becomes consciously observable, in other words, when it breaks into awareness. One great advantage of the b-CFS is that it targets low-level visual properties, such as direction of motion (Wade and Wenderoth 1978), rather than high-level semantic or conceptual features that step in at later stages of processing (Dehaene et al. 2006). Such sensitivity to early visual processes is a quality that can be particularly useful when the research aim is to test how fast a group of different first language speakers, or second language learners, can dissociate between different types of motion (e.g., a jumping frog vs. a gliding frog) in the pre-verbal stage, based solely on the crude visual feature distinctions of the stimuli.

Numerous studies suggest that language can permeate pre-verbal cognitive processes (Gray et al. 2013; Lupyan and Ward 2013; Shore and Klein 2000). They report that target picture stimuli break through suppression early, in the low-level stage of processing simple features, rather than later when high-level semantic or conceptual information is extracted. Evidence for people's ability to faster detect the presence of an object if that object is linguistically encoded, compared to whether it is not, comes from continuous flash suppression experiments with shapes (Noorman et al. 2018) and objects (Lupyan and Ward 2013). Participants in these studies were faster to report that they saw something when the gradually appearing image had a relevant label. Such detection advantage cannot be attributed to verbalising the visual stimulus during the task because the images were suppressed by continuous flashes. Instead, the authors reasoned that linguistic labelling, be it habitual or ad hoc, equips language users with the ability to assess perceptual input faster. Designs employing b-CFS are thus in a good position to address pertinent criticism of behavioural studies arguing that language effects cannot be perceptual, but arise later, through verbal influence in a post-perceptual top-down fashion (Firestone and Scholl 2014; January and Kako 2007). In response to such criticism, we discuss b-CFS as a method that can tap into motion event perception sufficiently early not to require any semantic or verbal processing.

If language affects motion event perception, evidence will need to come from early, unconscious, pre-linguistic time windows. Languages-specific effects on motion event perception were indeed found, for instance in Flecken et al. (2015), who recorded brain activity of German and English speakers and observed higher sensitivity to the perceptual saliency of endpoints in German speakers and to the perceptual saliency of trajectories in English speakers. As an alternative or complementary measure to brain signatures, reaction times in b-CFS can be collected as a window into early perceptual processing. The b-CFS approach could be used to examine sensitivity to the perceptual saliency of not only endpoints, but also other widely researched motion event components, such as Path versus Manner (e.g., Gennari et al. 2002; Hickmann et al. 2017). The underlying logic rests on the assumption that different languages can act as attention magnets that strengthen different visual representations through a long-term coactivation of specific motion features and their corresponding verbal expressions. If, for instance, English speakers typically foreground the Manner and Spanish speakers concentrate more on the Path to express the same motion event, habitual focus on different motion components could help on their way up the visual hierarchy. This way up can be supported in language-specific ways already when various motion components enter conscious analysis.

The continuous flashing approach is applicable to resonant questions is bilingualism research as well. If b-CFS tests reveal language-specific effects in how fast different groups of monolinguals detect Manner-salient versus Path-salient motion, one further question to explore would be whether newly learned lexicalised knowledge (e.g., about different manners of motion) percolates through to the low level of processing visual feature distinctions in second language learners. And if it does, to what extent can a new language help to train greater visual expertise in detecting different motion components?

In the next sections we first outline the b-CFS approach and link it to motion event perception. We then present a feasibility study using this approach with English native speakers versus Mandarin native speakers. The following sections then critically assess the contribution of b-CFS to motion event research and discuss the implications of the feasibility study for data triangulation, stimulus types and group choices.

2 Design features in a b-CFS with motion events

To measure the processing of visual stimuli as they emerge into awareness requires a few careful steps. A necessary starting point is when the motion stimulus is invisible. In b-CFS, this is achieved by showing participants a high-contrast Mondrian-like mask or visual white noise that flashes at 10 Hz (Figure 1a). Second, a low-contrast motion event is introduced to the non-dominant eye, while the dominant eye keeps seeing the high-contrast flickering mask. The contrast of the motion event gradually increases from 0 to 100%. This procedure relies on binocular rivalry, which initially serves as an eraser of visual awareness, and then the target stimulus gets perceptually enhanced through contrast manipulation (Jiang et al. 2007). Observers are asked to indicate as fast as possible on which side or in which corner of the screen a motion event emerged from suppression. The time it takes observers to detect a stimulus serves as the outcome measure that signals the extent to which higher-level stimulus properties, such as their habitual linguistic encoding or their familiarity, remain effective during perceptual suppression (Stein et al. 2011).

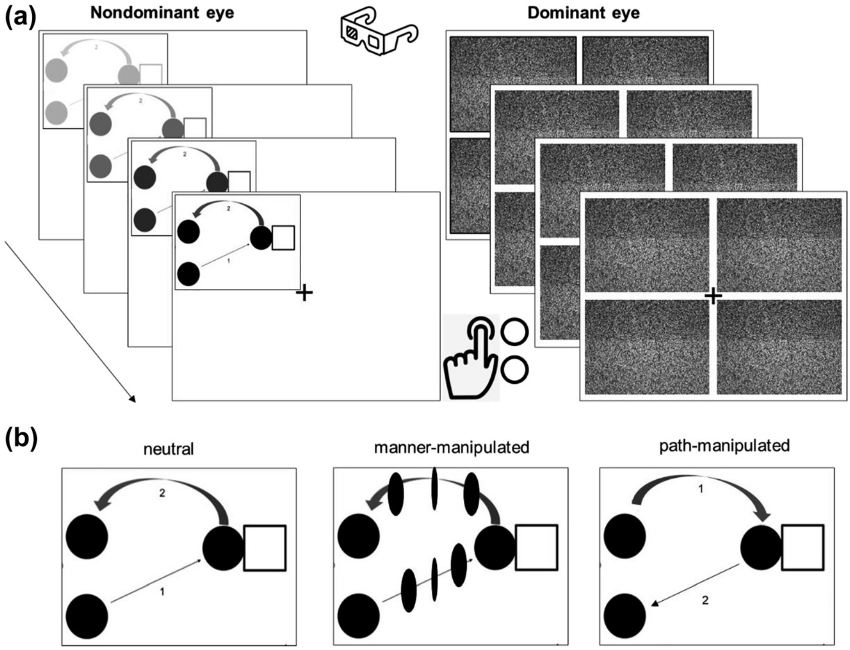

Figure 1: (a) Schematics of the experimental paradigm using b-CFS with a motion event [a circle bounces off of a square], and (b) examples of its Manner-based manipulation [a circle spins to bounce off of a square] and its Path-based manipulation [a circle descends to bounce off of a square].

Language-specific unconscious processing can be interpreted in a b-CFS scenario, when the detection speed of initially invisible stimuli varies across groups in critical items but not in control/neutral items (Figure 1b). If perceptual biases to specific motion components (Manner/Path) are predictable according to the speakers' native language, detection latencies for Manner-salient or Path-salient motion stimuli can be expected to break into awareness differently for speakers of satellite-framed languages like English and verb-framed languages like Spanish. To illustrate in more detail (Figure 1b), the reaction time differences between Manner-salient and neutral stimuli can be expected to be greater for English than Spanish, and conversely, one can expect greater RT differences between Path-salient and neutral stimuli for Spanish than for English speakers. This way, RTs varying as a function of linguistic encoding can be used as an index of automatic, low-level processing of motion in language-specific ways. An important consideration for RT calculations per group is to only include hits, i.e., correct responses, and exclude false alarm, which indicate the participant's susceptibility to guess. An alternative way is to adjust for false alarm rates by calculating *d primes* that capture the difference between the signal and the noise (Forrin et al. 2016). Also, the exclusion of reaction times in data analyses which are >2.5 SDs faster or slower than the mean RT for the given condition helps to maintain rigour and increase replicability.

Stimulus presentation can be generated with MATLAB using the Psychophysics Toolbox (Brainard 1997; Pelli 1997). Presentation of images to the two eyes needs to be fused by means of mirror stereoscope goggles that are ideally mounted on a chin rest. Some b-CFS designs keep the viewing distance at around 40 cm, use a frame that extends beyond the outer border of the stimuli, and they also add a fixation point to optimise stable image convergence (Jiang et al. 2007). One advantage of such a design is that it minimises post-perceptual high-level executive functions, such as strategic recruitment of language, because the observers are asked to make a response as soon as they detect the target (i.e., as soon as they see anything moving), be it the whole stimulus or merely some of its parts. Another advantage is that suppression time, blocking the target stimulus from awareness, can be manipulated to last from just a few seconds up to minutes (Stein 2019). Third, the compared conditions can be set to differ exclusively in the critical features (in this case Manner and Path) while keeping other potential confounds, such as Figure, Velocity, Ground and Degree of displacement, under control. If stimuli are arranged to compete against the interfering noise pattern in this way, it is possible to establish whether reduced response times in conditions with Manner or Path manipulations do or do not align with patterns one would expect based on their typical linguistic encoding.

No methodological approach is free of potential confounds, and b-CFS is no exception. Here we dissociate three factors that may affect the timing of when a motion stimulus breaks through suppression, and we discuss tips on how to minimise related risks. First, data analyses using b-CFS reaction times need to consider a phenomenon known as *lateralisation*, a left-sided hemispheric dominance in language processing shown in 95% of right-handed and in 70% of left-handed people (Lurito and Dzemidzic 2001). Interactions between language and perception were found stronger in the language-dominant left hemisphere (Francken et al. 2015), so it may be that motion events presented to the right visual field break into awareness generally faster than those presented to the left eye. A suitable response to reduce related threats, particularly to reliability, is to counterbalance left/right locations of emerging stimuli, and to add lateralisation as a random factor into statistical models. Another potential issue is if b-CFS studies do not control for detection time variation due to stimuli with different ecological relevance. Studies show that stimuli that are emotional break through suppression faster than the ones that are non-emotional (Stein et al. 2014), or that scenes that are more congruent with real-world knowledge (e.g., a basketball player holding a basketball) elicit faster reactions than less congruent stimuli (e.g., a basketball player holding a watermelon) (Mudrik et al. 2013). It is thus beneficial to consider ecological relevance in stimulus construction, for instance by excluding Path or Manner manipulations unlikely to be seen outside the lab. And third, b-CFS functionality is known for static pictures, less so for dynamic videos. There may be various ways to use motion events if videos turn out problematic, for instance by converting videos to sequence photos, or adjusting the flickering rate.

3 Proof of concept: a feasibility study of motion events via b-CFS

We designed a feasibility study with functionally monolingual English native speakers ($N = 24$) and Mandarin native speakers ($N = 24$) to check whether the b-CFS method can detect between-group and within-group differences that one would predict based on language-specific encoding of Manner and Path. Mandarin Chinese is known in motion event cognition literature as an equipollently-framed language, while English is known as a satellite-framed language (Chen and Guo 2009; Talmy 2000). To illustrate the key linguistic contrast, English typically uses the main verb and its subordinate satellite component (e.g., a verb particle) to express the Manner and Path information of the motion event respectively. In example (1), *push* is the main verb which indicates the Manner of the motion event, and the satellite (*into*) denotes the Path of the motion event.

(1) I push a box into a house.

(2) 我 把 一个 箱子 推 进 一个 房 子 。
wo3 ba3 yi1ge4 xiang1zi0 tui1 jin4 yi1ge4 fang2 zi0
I a box push enter a house
'*I push a box into a house.*'

However, Mandarin is an equipollently-framed language, which conveys the information about Manner and Path by verbs with equal linguistic weight. In example (2), *push* and *into* are both expressed with verbs, known as a serial verb construction, which is a common way to encode motion in Mandarin (Chen and Guo 2009). To examine the cognitive implications of the crosslinguistic difference that Manner is relatively more prominent in English than in Mandarin, for example Ji (2017) measured reaction times that participants needed in a triads-matching task to decide whether Path-matched scenes or Manner-matched scenes were more similar to model scenes. Although both English and Mandarin speakers preferred Path-matched scenes, their reaction times were not the same, namely, English speakers needed less time to decide about Manner-matched scenes, whereas Mandarin speakers took a comparable amount of time to decide about Path-matched and Manner-matched scenes.

Our research question was: Do differences in the linguistic encoding of motion events (Manner-focused English vs. equipollent Mandarin) affect automatic/low-level processing of motion in language-specific ways in English versus Mandarin speakers? If (pre-)attentional biases to specific motion components (Manner/Path) are predictable according to the speakers' native language, we expected Manner and Path for English speakers to emerge into awareness differently compared to Mandarin speakers. We used related motion event processing research (Ji 2017) with English (a Manner-dominant language) and Mandarin (an equipollent language) to formulate hypotheses. Specifically, English speakers were expected to show an RT advantage for Manner-salient compared to Path-salient stimuli, while Mandarin-dominant speakers' RT differences were not predicted to copy this direction. Crosslinguistically, we predicted Manner-salient stimuli to break through suppression faster for English than for Mandarin speakers, but no such advantage was expected for Path-salient stimuli.

The materials consisted of 48 animated cartoons and 24 fillers in total. Each animation, i.e., stimulus, displayed the same protagonist controlled for size and colour (Figure 2). The 48 animations were divided into 12 quadruplets. One quadruplet consisted of a motion-neutral clip (e.g., *a man carrying a suitcase into the room*), a Manner-salient clip (e.g., *a man pulling a suitcase into the room*), a

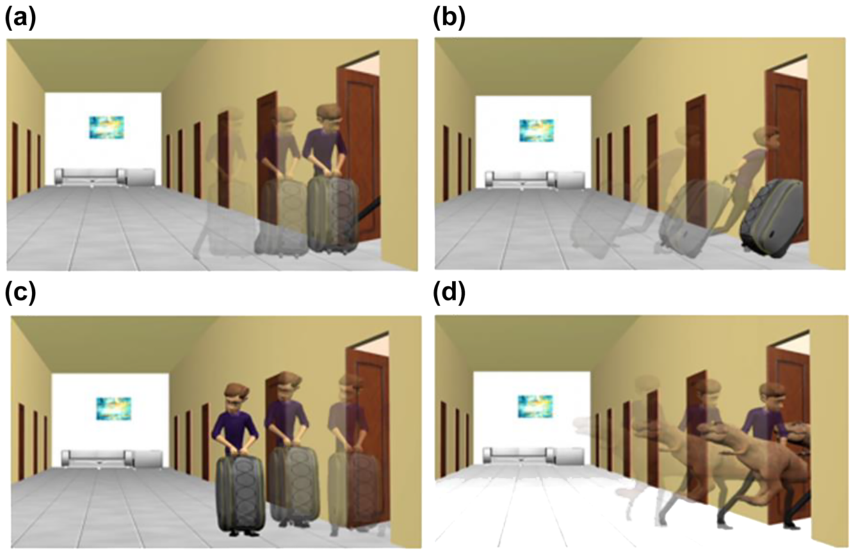

Figure 2: An example stimulus quadruplet used in the feasibility study showing photo sequences of (a) a neutral/reference video (*a man carrying a suitcase into the room*), (b) a Manner-salient video manipulated for the kind of motion (*a man pulling a suitcase into the room*), (c) a Path-salient video manipulated for the direction of motion (*a man carrying a suitcase out of the room*), and (d) a control video with a visual oddity (e.g., a man carrying a dinosaur into the room).

Path-salient clip (e.g., *a man carrying a suitcase out of the room*), a conspicuously odd video clip (e.g., *a man carrying a dinosaur into the room*) to test differences in visual feature detection. The Manner alternates were *pushing, rolling, carrying,* and *dragging*. The Path alternates were *along/across the street, towards/away from a cabin, around/across a puddle, up/down the hill, into/out of a truck, towards/away from a house, into/out of a room, away from/towards a skate park, into/out of a tunnel, across/around a garden, across/around a playground, along/across a river*. Difference RTs were compared only within quadruplets to minimise potential effects of Manner/Path-unrelated stimulus features on visual processing differences. That is, Manner saliency in this design meant the RT difference calculated by deducting the average RT for the motion-neutral clip from the raw RT for the corresponding Manner-salient clip. Likewise, Path saliency meant the RT difference calculated by deducting the average RT for the motion-neutral clip from the raw RT for the corresponding Path-salient clip. Each participant also saw 24 fillers that were randomly mixed with the clips from the quadruplets. The filler items showed animated motion events not related to the quadruplets (e.g., *a man cycling on the pavement*).

As for the procedure, participants were asked to wear a stereoscope goggle and follow the instructions. They were instructed to press a button (top-left, bottom-left, top-right, or bottom-right) as soon as they detected the location of a stimulus in one of the four corners of the rectangle they saw through the goggle. Their reaction times were recorded as an indicator of suppression durations. The animations were shown one by one. Two levels of randomisation and counterbalancing were employed. First, each participant was presented with the stimuli in a pseudo-randomized order, i.e., no two videos of the same condition could immediately follow each other. Each participant was tested on the full set of stimuli to avoid potential item effects. Second, the position of the emerging target stimulus was counterbalanced. For each participant, 25% of the stimuli appeared in the top-left, 25% in the top-right, 25% in the bottom-left, and 25% in the bottom-right corner of the screen.

Native speakers of English or Mandarin Chinese were recruited. The inclusion criteria were normal or corrected-to-normal vision, clear dominance of English or Mandarin as the native language, and no history of neurological and/or language impairments. The sample size in this concept-check is modest, corresponding to a pilot study. Both English speakers and Mandarin speakers (N = 24/group) were recruited at the University of York, UK. The English speakers were aged between 18 and 22 years (M = 19.25, SD = 0.99), 12 in the group were females. The Mandarin speakers were aged between 18 and 33 years (M = 21.96, SD = 2.99), 16 in the group were females. Although the Mandarin speakers also had some knowledge of English (score 6 or below in an IELTS test), they all self-identified as functionally monolingual Mandarin-dominant speakers who only arrived in the UK close to the time of testing.

The manipulated variables were Condition (Manner-salient, Path-salient, neutral, control) and Group (Chinese, English). The measured variables of interest were response times from the stimulus onset (ms), and accuracy of identifying the correct location on the screen (%). Participants in both groups identified the correct corner where the videos appeared with high accuracy (M = 94.06%, SD = 6.99 in the Chinese group; and M = 95.59%, SD = 4.83 in the English group). One combined measure was used, reaction time difference scores, calculated for each participant for the Manner-salient condition (RT Manner alternate minus RT neutral alternate) and the Path-salient condition (RT Path alternate minus RT neutral alternate). To test the effects of Condition and Group, and their interaction on the speed of detecting stimuli as they break through suppression, we built a linear mixed-effect regression model using the lme4 package (Baayen et al. 2008) in the R software. The fixed effect factors were Group and Condition, the outcome variable was the RT difference and the random effect factors were Participant and Item. The model included all possible random effects (Barr et al. 2013), with random slopes over condition by participant and random slopes over condition and group by item as follows:

$$\text{difference RT} \sim 1 + \text{condition} \times \text{group} + (1 + \text{condition} \mid \text{participant}) + (1 + \text{condition} \times \text{group} \mid \text{item})$$

To further explore the nature of the interaction, if significant, we planned to build a reduced model using a forward variable selection and zoom in on the Manner-salient items, comparing a model including Group with a reduced model without Group in the data for Manner-salient items only. And as the final statistical step we ran Tukey-adjusted pairwise comparisons between groups to examine how group reaction times for specific types of motion events differ.

The results showed that English speakers detected Manner-salient motion stimuli faster (M = 5.03, SD = 3.33, calculated from target stimulus onset)) than Path-salient stimuli (M = 5.63, SD = 4.35). This pattern was not mirrored by Mandarin speakers, whose detection time was faster for the Path-salient (M = 6.67, SD = 4.11) than for the Manner-salient stimuli (M = 7.41, SD = 5.85). This asymmetry in the directionality of the reaction time advantage aligns with the prediction that salient Manner broke into awareness faster than Path in English speakers, unlike in Mandarin speakers. When looking at the standardised RTs (Path/Manner-salient minus neutral), shown in Figure 3, the lowest negative average RT difference was in the English group for Manner-salient stimuli (M = –1.75, SD = 3.60), indicating that when Manner was salient, English speakers detected it faster on average than the neutral stimuli. The detection advantage was smaller for the Path-salient compared

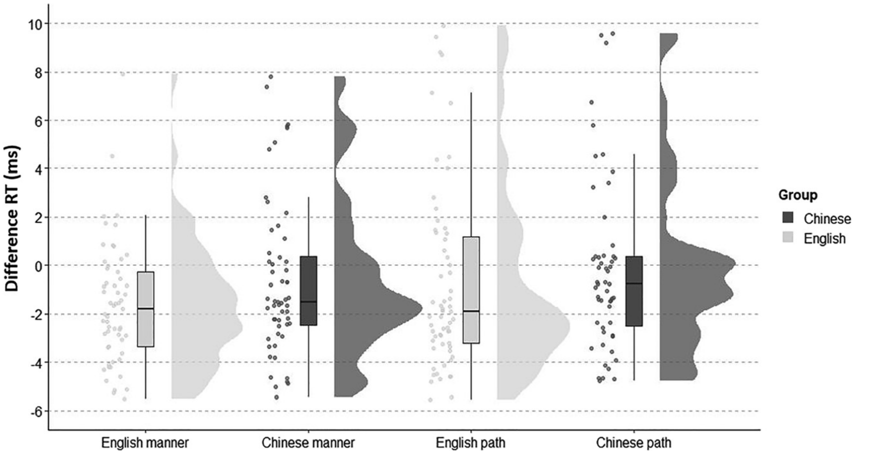

Figure 3: Motion event detection times expressed as difference RTs (RT Manner minus RT neutral; RT Path minus RT neutral) per condition and group. Box plots show the medians and the 50% of the transformed motion event detection times within the boxes. Raincloud and violin plots were added to aid visualisation of the data distribution patterns across conditions and groups.

to the neutral stimuli (RT difference M = –0.39, SD = 4.71). For Mandarin speakers, detection time was slower for both Manner-salient and Path-salient than for neutral stimuli (RT difference M = 1.04, SD = 6.17; M = 1.05, SD = 5.29 respectively). This within-group difference in the Mandarin speakers was substantially smaller than in English speakers. Whether the results can be interpreted as language specificity in motion event detection times was further tested through statistical modelling.

To statistically check the effect of language group on reaction speed when participants detect motion events, we built a mixed-effects model. We followed the analysis plan and specified Group and Condition as fixed effect factors, and Participant and Item as random effect factors. We used the RT differences as the outcome variable for greater between-group comparability. The RT difference distribution is shown in Figure 3. Interaction between the fixed effects factors was also tested. We built a maximal model that converged: differenceRT ~ 1 + condition × group + (1 + condition | participant) + (1 + condition + group | item). This model returned a significant effect of Group (estimate = –2.75, SE = 1.02, t ratio = –2.68, $p < 0.05$). To confirm whether this effect was driven by the differences between detection times for Manner-salient motion events, we next built a model zooming in on the Manner-salient events only. We found that detection times for Manner were significantly shorter in the English than in the Chinese group (difference RT, estimate = –2.83, SE = 1.02, t ratio = –2.77, $p < 0.05$). The between-group difference was not significant for the subset of reaction times for Path (difference RT, estimate = –1.45, SE = 0.91, t ratio = –1.59, $p > 0.05$), showing that the main crosslinguistic difference was indeed Manner-based as predicted.

In sum, perceptual processing of Manner in English speakers was faster than that of Mandarin speakers, which is what one would expect with an equipollent language (Ji 2017). We interpret these results as an indicator that b-CFS is a sensitive methodological approach able to show within and between-group differences that are based on language-specific encoding of Manner and Path. These data distributions, although in the predicted direction, need to be interpreted with caution, as they only come from a pilot-sized sample.

4 Critical assessment of the contribution of b-CFS to motion event research

A key point to clarify when considering the usefulness of the proposed technique is the link between language and perception. How exactly do low-level perceptual processing signals detected by the b-CFS relate to the high-level linguistic representations of Manner and Path? We adopt the *predictive processing* framework

(Lupyan and Clark 2015) to explain this relationship. Within this framework, language-modulated predictions *flow downward* and influence sensory signals that *flow upward*. In the context of the current study, the prominence of encoding Manner to express motion events in English supports high-level predictions of English speakers to expect Manner information to be particularly relevant for low-level perceptual processing of clips with motion events. However, for speakers of languages where the prominence of encoding Manner is comparatively lower, such as Mandarin or Spanish, predictions against which sensory signals are assessed would be less Manner-based. A timed motion detection task fuels the formation of probabilistic predictions that integrate any source of knowledge available to reduce prediction error. When a high-level linguistic representation, such as the encoding of Manner, turns out to be relevant for the task of detecting motion through continuous flashes, Manner information becomes more readily available for predictions. But if a language does not typically make Manner information salient, its speakers are likely to give less weight to Manner information in making predictions about the gradually appearing motion. This way, the b-CFS technique allows us to test whether different ways of talking about motion across languages influence how motion is perceived. In a wider context, this b-CFS affordance is in a good position to inform the resonant debate on whether linguistic encoding can shape perceptual processing at the very basic levels (Boroditsky 2010; Casasanto 2008; Thierry 2016) or if the two levels are not interconnected at all (Gleitman and Papafragrou 2005; Pinker 1994). Our experiment suggests that verbal cues, in this case the foregrounding of Manner when talking about motion in English, can act as a source of predictions that modifies what kind of top-down information influences perceptual processing. We propose that language-specific encoding of Manner gains a prominent role in motion perception because it can adjust the weight of Manner information when perceivers form top-down expectations about sensory input as motion events unfold.

One might wonder whether the b-CFS technique, shown to work for static visual displays (e.g., shapes in Noorman et al. 2018; simple objects in Lupyan and Ward 2013; and faces in Jiang et al. 2007), extends its functionality to contexts with more complex dynamic motion stimuli. Currently there is some empirical support for this idea. Paffen et al. (2008) combined b-CFS with motion events in a speed discrimination task to investigate whether objects that moved in a matching direction broke through suppression faster than objects that moved in mismatching directions. The speed of discrimination showed a steady improvement across training, and b-CFS times for directions that were *not* matching increased. One interpretation was that mismatching directions became more strongly suppressed because of attentional inhibition. In this case, b-CFS demonstrated that it is suitable for tracking sensitivity to changes in the direction of motion. This is of immediate relevance to the present study because the time to detect (changes in) the direction of motion is the low-level psychophysiological correlate of the high-level semantic representation of Path. Analogously, we consider the time to detect (changes in) the kind of movement to be the low-level psychophysiological correlate of the high-level semantic representation of Manner. Differences in direction (such as upward vs. downward) may strike as more salient than differences in the kind of movement (such as *pushing* vs. *rolling*). Following the rationale from the predictive processing framework, small but relevant digressions form predictions receive more weight because increased attention to them can reduce prediction error (Den Ouden et al. 2012). This way, greater subtlety that characterises Manner-based differences can be advantageous for tuning sensory processing. Sceptics may argue that finer motion-based stimulus distinctions can jeopardise the effectiveness of continuous flash suppression (Pournaghdali and Schwartz 2020). Afterall, the stimuli in Paffen et al. (2008), and numerous earlier b-CFS studies, were moving dots or similar simple objects, valid as tools for tapping into the low-level visual processing of *crude* visual features rather than finer motion distinctions. However, the between-group differences in the predicted direction reported here show that b-CFS can also render reliable results with dynamic stimuli, making the method suitable for probing into the unconscious preverbal processing of motion events.

Factors other than language-modulated Path and Manner saliency could have contributed to the pattern of b-CFS results in our feasibility study. Individual difference effects, linked for instance to entrained vision, cannot be ruled out. It is possible that frequent computer gamers outperformed other participants in the speed of detecting motion because of more frequent exposure to, and interaction with, animated motion events shown on a computer screen compared to non-gamers. Collecting relevant background information of this type and incorporating it into the random effects structure in statistical analyses will be beneficial for future b-CFS studies to tighten control over potential confounds.

5 On data triangulation, motion event types and group choices

The hypotheses of this study rest on the assumption of crosslinguistic differences in the verbalisation of motion events between Mandarin and English speakers. Production data collected from the same participants verbalising the same events that were used for the b-CFS task would be a quality booster, but it was not collected in our feasibility study. In the absence of such data, we attempt, at least partly, to rectify this limitation by revisiting earlier studies with production test results from groups comparable with those in this study. One study that highlights the differences in how native Mandarin versus English speakers express Manner and Path is Ji and Hohenstein (2014). Participants described 16 short (5 s) animated videos to an imaginary listener. Each video depicted a caused motion event with Manner (e.g., *rolling/sliding*) and Path (e.g., *towards/away from*) kept comparably salient. Results showed a contrast between how the two language groups expressed Manner (coupled with cause) and Path information. In English 98% of the descriptions followed a pattern of encoding Manner in the main verb and Path in particles, tightly packaged within one matrix clause (e.g., *He pulled the treasure into the pyramid*). In Mandarin, only about 50% of the motion events were expressed this way, while around 50% of utterances followed a looser pattern separating Manner and Path in a matrix clause and a subordinate clause/ZHE construction, thus pushing the Manner information to the periphery of the utterance (e.g., *Nan hai tui zhe mu chai zou-xiang huo dui* [The boy, pushing logs, walked towards the fire]). This pattern of results pointing out the same crosslinguistic variation in how the Manner component is distributed in English versus Chinese utterances was also reported in Ji et al. (2011). Our rationale for the b-CFS predictions was built on this crosslinguistic difference in motion event encoding, following the idea that differential linguistic encoding of Manner, more peripheral in Mandarin than English, would give rise to between-group variation in how quickly Manner-salient motion gets detected.

Regarding motion event types, the options for this study were to select either simpler, spontaneous (voluntary) motion events or more complex caused motion events (Talmy 1985). Spontaneous motion (as shown in Figure 1) comprises fewer semantic components (the circle in Figure 1 is the *Figure*, the square is the *Ground* or the reference frame in relation to the circle's motion, the upward-to-the-right trajectory is the *Path*, the spinning is the optional *Manner*, and the actual *Motion* of the circle is the fifth component). Simpler stimuli, such as those in Paffen et al. (2008), are advantageous for the internal validity of b-CFS designs as they allow researchers to control for as many visual features as possible to manipulate Path and Manner in the most stringent way. In caused motion events, there is an additional component, *Cause*, which is coupled with *Manner* (e.g., *pushing, rolling, carrying, dragging*). This study opted for the more complex caused motion for two reasons. The first motivation came from reports on the difference between how Mandarin versus English encode caused motion. Ji et al. (2011) and Ji and Hohenstein (2014) observed that while English speakers habitually foregrounded Manner in the main clauses, a large proportion of Chinese speakers backgrounded Manner in the periphery of their event descriptions through subordination. This variation observed in the two language groups served as a springboard to test if differences in the expression of Manner are predictive of inequivalence in its perceptual saliency at a low level, when a motion event breaks into awareness. Second, experimental designs with caused motion are gaining momentum in studies on the link between first or second language and cognition (e.g., Engemann et al. 2012; Montero-Melis and Bylund 2017; Wang and Wei 2021), making b-CFS with the same event type more directly relatable.

In terms of group choices, testing Mandarin versus English speakers might not, at first, seem ideal for a crosslinguistic comparison of how quickly Manner-salient motion gets detected. An equipollently-framed language like Mandarin and a satellite-framed language like English do not exhibit as clear-cut a contrast in the salience of how Path versus Manner are typically expressed as for instance a verb-framed language like Spanish or Japanese would. One alternative for a b-CFS feasibility study was to follow perhaps an intuitively more straightforward approach and build on pronounced contrasts reported for the expression of Manner in English versus a verb-framed language where the expression of Manner is less frequent, such as Spanish (e.g., Cadierno and Ruiz 2006; Duncan 2001; Naigles et al. 1998) or French (e.g., Hickmann and Hendriks 2010; Treffers-Daller and Tidball 2015). However, from the perspective of precision weighting integral to predictive processing, a smaller between-group difference in Manner encoding could be advantageous because, in a difficult task like the b-CFS, a relatively smaller difference may require more attention to minimise subsequent prediction error. To firmly establish if this is indeed the case, future studies will find it beneficial to include speakers of satellite-framed versus verb-framed languages, or even better, use a fully crossed design (satellite-framed vs. verb-framed; satellite-framed vs. equipollent; equipollent vs. verb-framed).

The b-CFS approach could be a particularly welcome addition to the existing tool set in studies testing low-level language effects on motion perception since the currently used inventory, particularly in second language (L2) research, is often limited to skin conductance responses (e.g., Vanek and Tovalovich 2022) or early brain responses in the pre-attentive time window (e.g., Boutonnet et al. 2012). One area where the b-CFS approach could shine is research on low-level perceptual effects of learning a new language. This could be achieved, for instance, through exploring whether learners' verbal responses and detection speed change in tandem when one learns an L2-based motion distinction. As an illustration from existing research on motion, Kersten et al. (2010) trained Spanish learners of L2 English to categorise unfamiliar moving objects, and found that, in a Spanish context, bilinguals' categorisation resembled that of L1 Spanish speakers, and in an English context the same participants' categorisation was more like the pattern observed in L1 English speakers. As a form of data triangulation, an added b-CFS experiment could employ a similar perceptual training paradigm with Spanish learners of L2 English, in which relatively faster RTs to Path manipulations in a Spanish context and gradually faster RTs to Manner manipulations in an English context could be interpreted as a manifestation of L2-driven perceptual changes. Another area for b-CFS use would be to test whether different types of training with variations in L2 involvement (see e.g., Vanek 2020), boost response speed in detection times of specific motion event features in different ways. These are stimulating avenues for future inquiry, and there are some results from related research testing the effects of L2 feedback (e.g., on risk taking, Gao et al. 2015) signalling that greater crosslinguistic difference in verbally modulated behavioural responses can be expected to correlate with greater difference in automatic measures as well. Transforming perception of motion features through a second language is an exciting research prospect possible to carry out through a combination of b-CFS used for perceptual learning with linguistic feedback. Such a combination could reveal how perceptual and linguistic cues are gradually co-integrated while participants are learning new ways of encoding motion events in the target language.

References

Brainard, David. 1997. The psychophysics toolbox. *Spatial Vision* 10. 433–436.

Baayen, Harald, Douglas Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Boroditsky, Lera. 2010. How the languages we speak shape the ways we think: The FAQs. In Michael Spivey, Marc Joanisse & Ken McRae (eds.), *The Cambridge handbook of psycholinguistics*, 615–632. Cambridge: Cambridge University Press.

Boutonnet, Bastien, Panos Athanasopoulos & Guillaume Thierry. 2012. Unconscious effects of grammatical gender during object categorisation. *Brain Research* 1479. 72–79.

Cadierno, Teresa & Lucas Ruiz. 2006. Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics* 4(1). 183–216.

Casasanto, Daniel. 2008. Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language and learning* 58. 63–79.

Chen, Liang & Jiansheng Guo. 2009. Motion events in Chinese novels: Evidence for an equipollently-framed language. *Journal of Pragmatics* 41(9). 1749–1766.

Dehaene, Stanislas, Jean-Pierre Changeux, Lionel Naccache, Jérôme Sackur & Claire Sergent. 2006. Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences* 10. 204–211.

Den Ouden, Hanneke, Peter Kok & Floris De Lange. 2012. How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology* 3. 548.

Duncan, Susan. 2001. Co-expressivity of speech and gesture: Manner of motion in Spanish, English, and Chinese. *Annual Meeting of the Berkeley Linguistics Society* 27(1). 353–370.

Engemann, Helen, Anne-Katharina Harr & Maya Hickmann. 2012. Caused motion events across languages and learner types: A comparison of bilingual first and adult second language acquisition. In Luna Filipović & Kasia Jaszczolt (eds.), *Space and time in languages and cultures*, 263–288. Amsterdam, Netherlands: John Benjamins.

Firestone, Chaz & Brian Scholl. 2014. "Top-down" effects where none should be found: The El Greco fallacy in perception research. *Psychological Science* 25. 38–46.

Flecken, Monique, Panos Athanasopoulos, Jan Rouke Kuipers & Guillaume Thierry. 2015. On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition* 141. 41–51.
Forrin, Noah, Brianna Groot & Colin MacLeod. 2016. The d-Prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(7). 1090–1111.
Francken, Jolien, Peter Kok, Peter Hagoort & Floris De Lange. 2015. The behavioral and neural effects of language on motion perception. *Journal of Cognitive Neuroscience* 27(1). 175–184.
Gao, Shan, Ondrej Zika, Robert Rogers & Guillaume Thierry. 2015. Second language feedback abolishes the "hot hand" effect during even-probability gambling. *Journal of Neuroscience* 35. 5983–5989.
Gayet, Surya, Stefan van der Stigchel & Chris Paffen. 2014. Breaking continuous flash suppression: Competing consciousness on the pre-semantic battlefield. *Frontiers in Psychology* 5(460). 1–10.
Gennari, Silvia, Steven Sloman, Barbara Malt & Tecumseh Fitch. 2002. Motion events in language and cognition. *Cognition* 83. 49–79.
Gleitman, Lila & Anna Papafragou. 2005. *Language and thought*. Cambridge: Cambridge University Press.
Gray, Katie, Wendy Adams, Nicholas Hedger, Kristiana Newton & Matthew Garner. 2013. Faces and awareness: Low-level, not emotional factors determine perceptual dominance. *Emotion* 13. 537–544.
Hickmann, Maya & Henriëtte Hendriks. 2010. Typological constraints on the acquisition of spatial language in French and English. *Cognitive Linguistics* 21(2). 189–215.
Hickmann, Maya, Helen Engemann, Efstathia Soroli, Henriëtte Hendriks & Coralie Vincent. 2017. Expressing and categorizing motion in French and English. In Iraide Ibarretxe-Antuñano (ed.), *Motion and space across languages: Theory and applications*, 61–94. Amsterdam: Benjamins.
January, David & Edward Kako. 2007. Re-evaluating evidence for linguistic relativity: Reply to Boroditsky. *Cognition* 104. 417–426.
Ji, Yinglin. 2017. Motion events 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). 1–12.
Ji, Yinglin & Jill Hohenstein. 2014. The syntactic packaging of caused motion components in a second language: English learners of Chinese. *Lingua* 140. 100–116.
Ji, Yinglin, Henriëtte Hendriks & Maya Hickman. 2011. The expression of caused motion events in Chinese and in English: Some typological issues. *Linguistics* 45(5). 1041–1077.
Jiang, Yi, Patricia Costello & Sheng He. 2007. Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science* 18. 349–355.
Kersten, Alan, Christian Meissner, Julia Lechuga, Bennett Schwartz, Justin Albrechtsen & Adam Iglesias. 2010. English speakers attend more strongly than Spanish speakers to Manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General* 139(4). 638–653.
Lupyan, Gary & Emily Ward. 2013. Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences of the United States of America* 110. 14196–14201.
Lupyan, Gary & Andy Clark. 2015. Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science* 24(4). 279–284.
Lurito, Joseph & Mario Dzemidzic. 2001. Determination of cerebral hemisphere language dominance with functional magnetic resonance imaging. *Neuroimaging Clinics of North America* 11(2). 355–363.
Montero-Melis, Guillermo & Emanuel Bylund. 2017. Getting the ball rolling: The cross-linguistic conceptualization of caused motion. *Language and Cognition* 9(3). 446–472.
Mudrik, Liad, Hagar Gelbard-Sagiv, Nathan Faivre & Christof Koch. 2013. Knowing where without knowing what: Partial awareness and high-level processing in continuous flash suppression. *Journal of Vision* 13. 1103.

Naigles, Letitia, Ann Eisenberg, Edward Kako, Melissa Higher & Nancy McGraw. 1998. Speaking of motion: Verb use in English and Spanish. *Language and Cognitive Processes* 13(5). 521–549.
Noorman, Samuel, David Neville & Irina Simanova. 2018. Words affect visual perception by activating object shape representations. *Scientific Reports* 8(1). 1–10.
Paffen, Chris, Frans Verstraten & Zoltán Vidnyánszky. 2008. Attention-based perceptual learning increases binocular rivalry suppression of irrelevant visual features. *Journal of Vision* 8(4). 25.
Pelli, Denis. 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10. 437–442.
Pinker, Steven. 1994. *The language instinct: How the mind creates language*. New York: Harper Collins.
Pournaghdali, Ali & Bennett Schwartz. 2020. Continuous flash suppression: Known and unknowns. *Psychonomic Bulletin & Review* 27(6). 1071–1103.
Shore, David & Raymond Klein. 2000. The effects of scene inversion on change blindness. *The Journal of General Psychology* 127. 27–43.
Stein, Timo. 2019. The breaking continuous suppression paradigm, review, evaluation and outlook. In Guido Hesselmann (ed.), *Transition between consciousness and unconsciousness*, 1–38. New York: Routledge.
Stein, Timo, Martin Hebart & Philipp Sterzer. 2011. Breaking continuous flash suppression: A new measure of unconscious processing during interocular suppression? *Frontiers in Human Neuroscience* 5(167). 1–17.
Stein, Timo, Kiley Seymour, Martin Hebart & Philipp Sterzer. 2014. Rapid fear detection relies on high spatial frequencies. *Psychological Science* 25(2). 566–574.
Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical form. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3, 36–149. Cambridge: Cambridge University Press.
Talmy, Leonard. 2000. *Toward a cognitive semantics*, vol. 2. Cambridge: MIT Press.
Thierry, Guillaume. 2016. Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning* 66(3). 690–713.
Treffers-Daller, Jeanine & Françoise Tidball. 2015. Can L2 learners learn new ways to conceptualise events? Evidence from motion event construal among English-speaking learners of French. In Pedro Guijarro-Fuentes, Katrin Schmitz & Natascha Müller (eds.), *The acquisition of French in multilingual contexts*, 145–184. Bristol: Multilingual Matters.
Vanek, Norbert. 2020. Changing event categorization in second language users through perceptual learning. *Language Learning* 70(2). 309–348.
Vanek, Norbert & Artem Tovalovich. 2022. Emotionality ratings and electrodermal responses to university-related expressions in a native and a non-native language. *International Journal of Bilingual Education and Bilingualism* 25(8). 2817–2833.
Wade, Nicholas & Peter Wenderoth. 1978. The influence of colour and contour rivalry on the magnitude of the tilt after-effect. *Vision Research* 18. 827–835.
Wang, Yi & Li Wei. 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.