Investigating the Linguistic Relativity in a breaking Continuous Flash Suppression study: language-specific evidence in motion event cognition among Mandarin Chinese speakers, English speakers, and Chinese L2 English learners

Xingyi Fu
付星一

Doctor of Philosophy
University of York
Education

May 2024

Abstract

The current study explores the language-specific effects predicted by the linguistic relativity (LRH, Whorf, 1956) in motion event cognition among monolingual Chinese and English speakers, and Chinese learners of English. Three non-verbal experiments were conducted: a breaking Continuous Flash suppression (b-CFS) paradigm, a self-paced video-video (VV) verification, and a self-paced sentence-video (SV) verification task. The major linguistic difference between Chinese (equipollently-framed language) and English (satellite-framed language) in expressing motion event is the *manner* and *path*. For example, in '*man carrying suitcase into room*', English uses the main verb (*carry*) and its subordinate satellite element (*into*) to express those two components respectively (Talmy, 2000), whereas Mandarin Chinese conveys them by verbs with equal linguistic terms: *carry* and *into* bear the same linguistic feature (a 'serial verb construction', Chen & Guo, 2009). English speakers are expected to take different amounts of time when processing *manner* and *path* in all three experiments, while Chinese speakers should process them similarly, since both components carry equal weight. The results supported the majority of the predictions: in b-CFS, English speakers were significantly faster to process stimuli with *manner* manipulation, while Chinese speakers showed no such difference; in SV, similar results were obtained but the English speakers spent more time to process stimuli in the same condition (i.e., with *manner* manipulation); in VV, both language group exhibit distinct language-specific patterns, but only the Chinese group showed statistically significant results. Cognitive transfer (Jarvis & Pavlenko, 2010) was evident in the L2 learners examined in all three experiments. In sum, the current study supports the LRH by providing consistent cross-linguistic evidence of the conceptualisation of motion events in nonverbal experiments, even when conscious linguistic recruitment is not allowed.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All resources are acknowledged as References. Part of this thesis has been published in the following journal articles:

Fu, X., Vanek, N., & Roberts, L. (2023). Matched or moved? Asymmetry in high- and low-level visual processing of motion events. *Language and Cognition*, 1–24. https://doi.org/10.1017/langcog.2023.37

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*(1), 61–78. https://doi.org/10.1515/iral-2022-0048

Acknowledgements

I often consider myself as a lucky person who has encountered great teachers throughout my life, and this is also the story of how my Ph.D. journey began. I'd like to express my appreciation to my previous supervisor, Norbert Vanek, who brought me into the intriguing field of studying language and cognition. I also want to express my gratitude to my current supervisor, Leah Roberts, who is both wise and kind, supporting me like a fairy godmother. They have shown me the best way of becoming an independent researcher.

I would also like to thank my family for everything they did to support me in finishing my work, including my husband, Wankang Zhang, who encourages me and supports me with endless love. My son, the cutest human being in the whole universe, who slept early so I could find time to work. My parents, my little brother, my parents-in-law, and my friends who have supported me throughout my doctoral studies.

Lastly, I want to thank myself for not giving up and for finishing the work that I thought I would never be able to accomplish many times in the past six years. The journal has just begun, and I will try my best to thrive.

Contents

List of figures

List of tables

Chapter 1: Introduction

This thesis investigates the effects of language on thoughts, according to the Linguistic Relativity Hypothesis (LRH, Whorf, 1956), people speak different languages perceive the world in distinct ways. With the accumulation of cross-linguistic evidence (Boroditsky, 2001; Casasanto *et al*., 2004; Vanek & Zhang, 2023) substantiating the existence of the language-specific effects across various domains, the discussion regarding the effects of language on thoughts has shifted from the mere inquiry of 'whether languages influence thoughts' to an exploration of the underlying mechanisms that govern the manifestation of such effects. Specifically, are these language-specific effects pervasive enough to guide our cognitive processes even when the language is not used consciously? Consequently, examining the cross-linguistic differences in non-verbal experiments becomes the core mission in the field of language and cognition (Lucy, 1997; Athanasopoulos, 2012; Bylund & Athanasopoulos, 2014; Levinson, 2012). Furthermore, this exploration now encompasses not only the investigation of cross-linguistic differences between speakers with different language backgrounds across various semantic domains (Winawer *et al*., 2007; Gilbert *et al*., 2006; Levinson & Majid, 2014), but also the cognitive shifts observed in second language learners as they navigate between their source and target language-based cognitive processes (Aveledo & Athanasopoulos, 2023; Bylund & Athanasopoulos, 2014b; Athanasopoulos & Aveledo, 2012; Bylund & Jarvis, 2011; Bylund, 2011a).

However, only a few studies successfully discovered language-specific effects of motion expression on motion perception without the mediation of conscious linguistic involvement (Kerstan *et al*., 2010; Flecken, Athanasopoulos *et al*., 2015; Kamenetski *et al*., 2022). Particularly in studies focussing on the motion lexicalisation, that is, language-specific effects were found to be task-specific (Gennari *et al*., 2002; Athanasopoulos & Albright, 2016; Athanasopoulos & Bylund, 2013a), for example, cross-linguistic differences of encodings *manner* and *path* are only discovered in non-verbal experiments with a proceeding verbal encoding task (Wang & Li, 2021a), or when the instruction directing attention to motion details (Montero-Melis & Bylund, 2017). Furthermore, in studies concerning cognitive transfer induced by motion lexicalisation in L2 learners, the converging conclusion mirrored those found in cross-linguistic studies, that is, cognitive transfer in L2 learners is observed only when language use was possible during the nonverbal experiment (Athanasopoulos, Bylund *et al.*, 2015).

To further explore the reasons of missing evidence for the LRH in examining the motion lexicalisation, the current study focusses on different stages of the visual processing of motion events with *manner*/*path* manipulations by applying three novel non-verbal experiments: breaking continuous flash suppression (b-CFS, Jiang *et al*., 2007), sentence-video (SV) verification, and video-video (VV) verification experiments (modified based on Zwaan *et al*., 2002). Specifically, the b-CFS examines the low-level processing patterns through detecting *manner* and *path* manipulated motion signals among monolingual Chinese, English, and English L2 (native language is Chinese) speakers, thus maximally excluding the possibility of consciously recruiting language during the non-verbal experiment. The two verification experiments investigate the encoding of identical motion components at a higher cognitive level involving semantic analysis with (i.e., sentence-video) and without (i.e., video-video) the assistance of overt linguistic labels.

The current study focusses on the language-specific effects induced by Mandarin Chinese (an equipollently-framed language) and English (a satellite-framed language) motion expressions, which varies at the degrees of *manner*/*path* saliency (Slobin, 2006) rather than an absent motion component (e.g., *manner* can be optional for verb-framed Spanish). Specifically, the major cross-linguistic difference between Chinese and English is the semantic features used to describe *manner* and *path* components in motion events. For example, in '*A man is carrying a suitcase into a room*', English uses the main verb (*carry*) and its subordinate satellite element (e.g., a verb particle: *into*) to express *manner* and *path* respectively (Talmy, 2000), whereas Chinese uses two verbs to illustrate those two semantic components (a 'serial verb construction', Chen & Guo, 2009; Ji, 2017). Therefore, English speakers were expected to use different amounts of time when processing *manner* and *path* salient stimuli in experiments conducted in the present study, in contrast, Chinese speakers should process them similarly (Ji, 2017). Furthermore, partially overlapped motion expressions between Chinese and English might provide fresh evidence for cognitive transfer in L2 learners when online linguistic recruitment is maximally blocked from consciousness. Since previous literature focussed more on the L2 learners whose source and target language share contrasting motion lexicalisation (Aveledo & Athanasopoulos, 2016; Filipovic, 2018, cf. Wang & Li, 2021a), and converging findings suggest that cognitive restructuring or shift are evident when language use was available for participants during non-verbal experiments.

The present study aimed to examine the extent to which motion expression in Chinese and English affects motion perception between monolingual Chinese and English speakers. Specifically, will such language-specific effects emerge even when sub-vocalisation is prevented, such as in the b-CFS experiment? Will the same effects be consistently observed at a higher-level of visual processing, and are they mediated by overt linguistic labels, such as in the VV and SV verification experiment? If so, do L2 learners achieve cognitive transfer in relation to motion events observed in all three non-verbal experiments? Do they exhibit similar cognitive processes as L2 learners with less contrasting motion expressions between their source (i.e., the equipollently-framed Chinese) and target (i.e., the satellite-framed English) languages? Do these specific cognitive processes vary at different levels of visual perception, that is, do they exhibit different processing patterns of *manner* and *path* among the b-CFS, SV, and VV verification experiments? What factors might potentially affect these specific cognitive outcomes in the three experiments, respectively?

This thesis comprises six chapters. In Chapter 2, detailed background information related to the effects of language on thought are introduced, including the fundamental theorical account: the LRH and two alternative frameworks (*thinking-for-speaking* and *label-feedback*), followed by summarise of key empirical studies investigating the LRH across various semantic domains. More importantly, key findings in the domain of motion are discussed in terms of the LRH. In the second half of the Chapter 2, cognitive transfer in L2 learners/ bilingual speakers predicted by the LRH are illustrated with evidence across various semantic domains, followed by potential predictors that may impact these specific cognitive processes. Then key evidence regarding cognitive transfer in L2 learners/ bilingual speakers in the domain of motion is provided and discussed, as well as the related predictors. Following this, key findings of cross-linguistic differences of motion cognition between Chinese and English, and the research gaps are introduced and discussed. In Chapter 3, the research methodology of the current study is introduced, including the general rationale for the current study and the detailed methodology in each experiment, including the research questions, predictions, participants, and procedures. In Chapter 4, the results in each experiment are presented and briefly discussed. In Chapter 5, detailed discussion regarding the results of each experiment, potential contributions to the field and limitations of the study are set out. In Chapter 6, the conclusions of the current study are presented.

Chapter 2: Literature review

This chapter illustrates and discusses previous research associated with the investigation of language-specific effects on thought from two main perspectives: language-specific differences among monolingual speakers and potential cognitive/ conceptual transfer in L2/ bilingual speakers. In the first section of the current Chapter, the effects of language on cognition are discussed by looking at the evidence from studies of monolingual speakers across various semantic domains. Following this, key findings of the language-specific effects on the domain of motion events are viewed in detail. The second section extends to the observation of language-specific effects in the field of second language acquisition, that is, the potential cognitive/conceptual transfer that L2/ bilingual speakers might experience after acquiring a second language. The last section introduces the current study by discussing research on how Chinese and English descriptions shape motion event cognition for Chinese, English, and L2 English speakers, as well as the methodological rationale behind the three non-verbal experiments presented in this thesis. More details in each section are presented in the following.

2.1 The effects of language on cognition

The debates regarding the effects of language on thought have been ongoing for many years, and *linguistic relativity* hypothesis (LRH), also known as the ‘Sapir-Whorf hypothesis’ (Whorf, 1956), might be the most influential theory on the topic. It asserts that people who speak different languages perceive the world in distinct ways. Specifically, the absence of linguistic representation in one language may result in a conceptual gap compared to a language requiring linguistic representation of the same reality. For example, Boroditsky (2001) discovered that Mandarin Chinese learners of L2 English tended to perceive time vertically even when required to describe time in English, which frequently utilizes horizontal terms to express the same temporal concepts.

Earlier in the last century, the LRH was criticised in a number of quarters. For instance, one notable categorization, proposed by Brown (1976), divides LRH into two versions: the ‘strong’ and ‘weak’. The former (or ‘determinism’) contends that language determines thought, while the latter, also known as "neo-Whorfianism" (Lucy, 1992a, 1992b), acknowledges the impact of language on thoughts and underscores the relative role of linguistic effects on thought. Specifically, the determinism claim of the 'strong' version has been rejected since people convey same thought in distinct ways might be due to the difficulty of expressing certain ideas in language (Pinker, 1995), that is, the determining role of language on thought seems to be impossible to be proved since not all thoughts can be verbalised by languages.

In line with the 'weak' version of the LRH, alternative frameworks predicting specific effects of language on thoughts have been brought to attention, such as the *thinking-for-speaking* (TFS, Slobin, 1987; 1996; 2003) which emphasizes the language-specific effects discovered in verbalisation. For instance, English speakers tended to be more attentive to the specific details of the actions since English in an aspect language, reflected in their motion description: "the boy *fell* from the tree and the dog was *running* away". On the contrary, as a non-aspect language, German speakers used past tense for both actions illustrated in the same motion scenario, that is, "the boy *fell* from the tree and the dog *ran* away". Furthermore, following the LRH, *label-feedback* hypothesis (Lupyan, 2012) also acknowledges the effects of language on thought but highlights the dynamic modulation of linguistic labels on the high-level semantic analysis of incoming visual stimuli, such as, Winawer *et al* (2007) discovered that the language-specific effects induced by distinct colour labels between Russian and English disappeared when verbal interference was introduced during the categorization procedure. To sum up, these two frameworks predicts that the effects of language on thought are limited to certain conditions, whereas the LRH believes the same linguistic influence should be exert beyond these limitations, that is, when language is in use (i.e., TFS) or sensitive to linguistic labels (i.e., the *label-feedback*).

However, this dichotomy is criticized as an oversimplification and misunderstanding of the original hypothesis (Everett, 2013; Bohenemeyer, 2020), since Whorf never proposed the absolute *manner* of language effects on thoughts. Instead, he did indicate that the effects of language on thought should be evident cognitively, as clarified by Athanasopoulos (2011), 'it is clear from Whorf's original formulation of the linguistic relativity principle that he viewed language as an attention-directing mechanism, which influences higher-level cognitive processes, namely, conscious evaluation of the experienced world' (p.30). That is, the evolutionary dynamics of language demonstrates that the biological genes remain stable while language undergoes dynamic evolution through experiential variations across diverse populations (Christiansen & Chater, 2008). For instance, adult English speakers prioritize the sorting of malleable objects based on their shape, in contrast to Yucatec Maya adult speakers. This difference is attributed to the grammatical number distinctions present in English, which distinguish between countable and mass nouns. In terms of the Maya and English children, who were examined at age of seven and nine, respectively, for categorizing same objects, the disparity in preference patterns found in the adult groups began to emerge at the age of nine in the children's groups, owing to the accumulation of language experience since the age of seven (Lucy & Gaskins, 2001; 2003).

Another critique of the LRH lies in apparent circularity due to failure to operationalize *thought* in a number of studies (Casasanto, 2016). In essence, the acknowledgement of language as a representation of thought creates an intricate interconnection between language and thought, thus, examining thoughts based on linguistic description might devolve into an analysis of the relationship between language and language, resulting potential failures. Therefore, instead of focussing on cross-linguistic differences reflected in verbal descriptions, investigations into the adequacy of the LRH are encouraged to provide evidence demonstrating high-level cognitive processes induced by language-specific effects, such as categorization and recognition (Lucy, 1997; Athanasopoulos, 2012; Levinson, 2012; Casasanto, 2016). More importantly, evidence observed in non-verbal approaches might resolve the earlier concerns brought by Pinker (1995) and examine the effects of language on thought when the same reality is difficult to be described by certain languages. Thus, providing evidence for examining the relatively strong version (cf. determinism or neo-Whorfianism) of the LRH, which indicates a stable and pervasive nature of the language effects on thought. Specifically, such effects should be evident regardless of covert or overt linguistic involvement (Gumperz & Levinson, 1996).

That is, to demonstrate that language can influence thought, predicted by the LRH, one must use a method that does not employ language yet shows evidence of its impact on thought. This is a challenging task to achieve. Some studies have attempted to do so, however. For instance, to avoid the pitfall of verbalizing languages silently in non-verbal experiments, specific measurements have been employed in various studies, such as manipulating the experimental stimuli: replacing linguistic labels by schematic symbols (Vanek & Zhang, 2023), increasing the cognitive load of the participants: introducing concurrent task during non-verbal categorisation (Winawer *et al*., 2007), or utilizing novel approaches that 'explore lower-level cognitive processes, that is, processes that are automatic and unconscious (e.g., oddball detection, visual search)' (Bylund & Athanasopoulos, 2014, p.958).

In this main section of Chapter Two, key studies across various semantic domains (e.g., *time*, see Boroditsky, 2001; *negation*, see Vanek & Zhang, 2023; *colour*, see Winawer *et al*., 2007; *object*, see Gilbert *et al*., 2008) have claimed to provide non-verbal evidence to support the LRH will be illustrated first, followed by discussion of specific measurements employed in these studies. Furthermore, two alternative claims (i.e., the *thinking-for-speaking* and *label-feedback* hypotheses) for specific effects of language on thought will also be briefly discussed., After this, empirical studies investigating the effects of language on thought specifically in the domain of motion events will be introduced and discussed before introducing the discussion regarding the effects of language on thoughts in L2 learners/ bilingual speakers. Lastly, the rationale of the current study is presented, covering the specific linguistic representations used to convey motion events in Chinese and English, as well as the non-verbal experiments employed to investigate the LRH, particularly its relatively stronger version.

2.1.1 Evidence from various semantic domains

2.1.1.1 Time-space

One classic example that has been claimed to support the weak version of LRH is part of the work by Boroditsky (2001), which examined how English and Mandarin Chinese speakers encode time metaphorically by associating it with spatial metaphor. Real-time encoding patterns were recorded and analysed based on reaction times. The cross-linguistic difference is that Chinese habitually describes time more in vertical terms: '*shang*' (up) and '*xia*' (down) than the English speakers (e.g., earlier/ later), even though they both can use horizontal terms when discussing time, such as, '*before*/*after* the year of tiger' in English, and '*hu nian zhi qian/hou* (in front of/behind the year of tiger)' in Chinese. In Boroditsky's first experiment, participants performed two sequential tasks, a spatial task serving as a prime and a subsequent time task as a target. In the spatial task, participants decided if the spatial relationship (horizontal/vertical) between two objects in the picture matched the linguistic description below it. The results aligned with language-specific features, showing that English speakers responded faster when the primes were associated with horizontal relationships compared to those with vertical relationships. Chinese speakers exhibited the opposite pattern. The author concluded that these findings are in line with the weak version of the LRH: that thought is shaped by the language, but not completely determined by it.

Similarly, Casasanto *et al*. (2004) also investigated the LRH in the domain of time and space.. However, unlike the Boroditsky (2001), this study did not include overt linguistic labels. Instead, they examined the influence of language on the low-level estimation of temporal representations. In this study, English, Indonesian, Greek, and Spanish speakers were examined in non-linguistic psychophysical experiments. Cross-linguistic differences were observed in the first experiment through corpus analysis: English and Indonesian speakers tended to use linear distance (e.g., a long time) to express duration, while Greek and Spanish speakers use quantity (e.g., much time). Subsequently, participants were examined in two non-verbal experiments in sequence, which examined the influence of estimating distance and quantity on the estimation of time, respectively. The rationale is that English and Indonesian speakers should be more affected by the perception of linear distance than the quantity when estimating time, while Greek and Spanish speakers should exhibit an opposite pattern, that is, they should more readily be affected by a proceeding estimation of quantity rather than linear distance.

Specifically, in the first non-verbal experiment, participants viewed a growing line and indicated its placement and duration after the line disappeared by clicking the mouse. That is, participants started by clicking the mouse at the centre of the screen (marked as ‘X’), then moved to the right and clicked at a position to indicate the distance the line grew. Afterwards, they clicked at a ‘hourglass’ icon at the centre of the screen and waited before clicking it again to indicate the time duration of the growing lines. In the second one, the procedure was identical to the first, except that the growing lines were replaced by filling containers. That is, participants started by clicking at the bottom of the container, then clicked at the location where the water reached to indicate the volume of the water. Following this, they clicked on the hourglass icon twice to indicate the time taken to fill the container. Both experiments were conducted in a self-paced manner. The results confirmed the authors’ predictions, which were based on the metaphorical features of space inherent in each language for conveying time. Specifically, English, and Indonesian speakers struggled to suppress the influence of linear distance when estimating time, while Greek and Spanish speakers were strongly affected by volume. The authors concluded that their results against the weak version of LRH, but not in line with the strong version either. Instead, they labelled their results as evidence for a *deep* version of the LRH, and specifically pointed out the influence of language on thoughts are evident even when the language is represented by non-linguistic representations.

Indeed, similar issue might be raised in this study for supporting the strong version of the LRH, which requires non-verbal context. Specifically, the non-linguistic representations were not controlled for leading a potential internal verbalisation. Even though participants were examined in a non-verbal context only involving low-level estimation of length, volume, and time, it is difficult to be sure that participants did not use language to guide their estimation behaviours. For instance, English speaker might describe a growing line as ‘a long line’, which is the same linguistic label that can be used to describe time (e.g., a long time).

2.1.1.2 Negation

A recent study by Vanek and Zhang (2023) also claimed to discover language-specific effects in area of negation. In this study, the process of negation was transformed into simple non-linguistic shapes (e.g., ▲ and ■) and equation symbols (e.g., positive “=” and negative “≠” equations) to study negative yes-no question on negation cognition in English and Mandarin Chinese. Specifically, English speakers tend to use ‘*Yes, he does/ No, he doesn’t’* answer a yes-no question (e.g., ‘*Doesn’t he love dogs?’*) due the polarity-based system in English. On the other hand, Mandarin Chinese speakers would answer as ‘*No, he does/ Yes, he doesn’t*,’ reflecting the truth-based system in Chinese. Using the example ▲≠■, participants were asked to agree or disagree with the equation. The polarity- based system in English allows speakers to focus solely on the equation symbol, while the truth-based system in Chinese necessitates the processing of both symbols and equations before making decisions. Specifically, Chinese speakers should take relatively longer time to verify stimuli involving negative equations compared to English speakers, since they habitually process both the truth of the equation and the non-linguistic shapes, while the English speakers made similar decisions as soon as they identify the truth of the equation. The results confirmed the predictions and discovered that English speakers indeed required less time to process the negative equation symbol with congruent shape compared to the Chinese speakers, reflecting a pervasive language-specific effect which supports the LRH.

By manipulating the stimuli, the authors separated the procedures of negation into two verification processes represented by equations and symbols, thus comparing the non-verbal performance between Chinese and English speakers. However, similar to previous studies examining the LRH through abstract linguistic representations, this study did not control for the potential bias of using language sub-vocally when judging whether the two shapes are the same or not. For instance, to judge based this stimuli: ▲≠■, English speakers might sub-vocally say “no, triangle does not equal to square”, and Chinese speakers might sub-vocally say “yes, it is true that triangle does not equal to square”. Thus, invalidating the non-verbal context created in this study for examining the LRH.

2.1.1.3 Gender

In linguistics, grammatical gender is assumed to be arbitrary, and thus not related to biological gender. However, Phillips & Boroditsky (2003) investigated whether the grammatical genders assigned to the object names influence the categorization of objects when they are compared to human genders. Spanish-English and German-English bilinguals were recruited for a series of experiments, since the same object can be assigned with distinct grammatical gender in Spanish and German, respectively, such as the name of *toaster* is feminine in Spanish but masculine in German. Such cross-linguistic differences might direct participants' categorical behaviour when needed to rate the similarity between object with human genders (e.g., human male and female).

In the first experiment, participants were required to compare objects with human males and females by rating their similarities based on the grammatical gender features assigned to these objects and the biological gender of human males and females, and the instructions were in English. The results confirmed that participants tended to rate the similarity higher for object-human pairs with matching genders (biological vs. grammatical) assigned to the objects in their native language, suggesting that despite its arbitrary nature, some link between grammatical and biological gender is made in the minds of speakers of such languages. In the second experiment, to test whether language experience plays a role in exerting linguistic effects, as predicted by the LRH, bilingual participants with various proficiency levels in L2 were recruited for an identical experiment, and the results suggested that bilinguals were influenced by their proficiency levels in the additional languages they speak. Specifically, they rated similarity higher for pairs in which the biological gender matched with the grammatical gender assigned by the language in which they were more proficient. The third experiment examined the influence of introducing a verbal interference (repeating English letters displayed to them outload) task during the similarity judgement task. The results showed that participants were not disrupted by the concurrent verbal shadowing and exhibited consistent patterns akin to those observed in the first experiment. In the fourth and fifth experiments, an effort was made to eliminate cultural bias by instructing to assign novel distinct labels ('*soupative*' vs. '*oosative*') to objects using a fictional language. For instance, pans, forks, and girls were designated as '*soupative*', while pots, spoons, and boys were labelled as '*oosative*'. Subsequently, participants engaged in a similarity rating task, with or without a verbal interference task. The results revealed that participants rated pairs with matching gender as having higher similarity, and this was not affected by the presence of a verbal interference task. Therefore, the authors concluded that the influence of grammatical gender on object perception is profound and was evident even when the inner speech was prevented by a verbal interference task.

2.1.1.4 Colour

Unlike those that utilize cross-modal connections employed in earlier abstract domains for examining the LRH, the domain of colour does not have such an advantage. For instance, Winawer *et al*. (2007) explored the cross-linguistic differences of colour encoding in nonverbal experiments between Russian and English speakers. In Russian, '*goluboy*' is for lighter blue, and '*siniy*' is for darker blue, a distinction that is not present in English. In the experiment, participants viewed one sample at the top and two alternates (one is from the same category as the sample, and the other not) at the bottom in each trial. They were asked to decide which of the two at the bottom was identical to the one at the top, while also performing a dual task under three conditions: viewing without interference, viewing with verbal interference (silently repeating heard digits), and viewing with spatial interference. The results showed that Russian speakers spent less time discriminating colours from different linguistic categories when there was no interference or spatial interference. However, such effects were not observed in the English group under all three conditions. The author argued that their results also answered the LRH, but in a different way, that is the linguistic representation can affect the effects of language on thoughts even when the stimuli is as simple as colour patches. This conclusion indirectly aligns with the weak version of the LRH, that is, the colour labels affect the categorization of non-linguistic colour patches, but such effects emerge when the inner speech was possible and easily to be wiped out when such inner speech was prevented by a verbal interference task. Thus, this study seems to align with the thinking-for-speaking hypothesis, rather than the LRH.

Similarly, Gilbert *et al*. (2006) also claimed to support the LRH owing to the observation of language-specific effects on colour perception through a lateralized colour discrimination task. The underlying premise is that the left hemisphere (LH) of the brain is primarily responsible for language processing, and the observation of categorical discrimination predominantly in the Right visual field (RVF) rather than Left visual field (LVF) serves as evidence supporting the effects of language on thoughts (Athanasopoulos & Casaponsa, 2020), thus a language-specific effects discovered in non-verbal experiments examining visual stimuli presented in the RVF will be in line with the LRH. In this study, three experiments were conducted. Through these experiments, participants were tasked with identifying which side of the ring stimuli contained a boundary between ‘blue’ and ‘green’ by pressing keyboard keys. The ring stimuli comprised of 12 squares displaying gradation of blue and green. To create a non-verbal context, a verbal interference task was concurrently introduced during the visual search, requiring participants to silently rehearse the displayed eight-digit numbers and then verbally repeat them at the end of the searching block. Reaction times for decision-making were recorded and analysed. The results found that participants exhibited faster identification of the blue-green boundary when it was presented in the RVF, indicative of a language effect on perceptual processing due to the predominant language operations in the LH. However, such effects were absent when the verbal interference task was introduced.

To ascertain whether the absence of language-specific effects was indeed attributable to linguistic involvement, a second experiment with a similar procedure was conducted. This experiment introduced two concurrent tasks with varying degrees of linguistic involvement. Specifically, participants engaged in a verbal interference task where they viewed linguistic labels for colour (e.g., red) and had to judge if the colour matched the one, they had previously viewed. In contrast, the nonverbal interference task required participants to make judgments between spatial grids of squares. The results of this experiment were consistent with those of the first, with language-specific effects observed only when the concurrent interference task was non-verbal. This indicates that the earlier language-specific effects observed in the first experiment might indeed be due to an advantage of inner speech. Furthermore, to further validate that the LH processed target presented in the RVF, a callosotomy patient with LH as the dominant language hemisphere was examined in the visual search task. The results showed quicker differentiation for RVF targets, affirming a language effect on colour perception.

To utilise the lateralisation design of examining the effects of language on thoughts, an extra layer was added to confirm the linguistic influence reflected by the cross-linguistic differences in non-verbal experiments. However, it does not solve the issue of potential language use during non-verbal performance.

2.1.1.5 Objects

Further evidence claimed to support the LRH has also been observed in Gilbert *et al*. (2008) which investigated the language-specific effects on perceiving animal shapes through lateralization. In this study, Gilbert and her colleagues conducted three discrimination tasks that required participants to differentiate between different lexical categories (e.g., ‘dog’ and ‘cat’), same lexical categories (e.g., ‘dog 1’ and ‘dog 2’) and perform similar tasks with verbal/ nonverbal concurrent tasks, respectively. The experimental design closely resembles those utilized in Gilbert *et al*. (2006), with the exception that the coloured squares were replaced by animal shapes for examining the effects of language on object perception. The prediction is that if language influences the perception of animal shapes, participants should react faster when the ‘cat’ was displayed in the RVF compared to LVF when distinguishing them between lexical categories. In terms of responses in the situation required to differentiate within lexical categories, participants should be slower when identifying the difference between two dogs present in the RVF compared to LVF. Such reaction patterns should disappear under verbal interference task, but not the non-verbal interference task. The results confirmed the predictions, and confirmed the lateralisation is not limited to examinations of language influence on thought in the domain of colour, and it also function well in the domain of objects. However, since the earlier conclusion of supporting the LRH in Gilbert *et al*. (2006) is not valid due to the possibility of using language silently, the conclusion of the present study of supporting the LRH remains to be questionable.

In sum, although the results of all these studies described above claim to support the LRH, these conclusions need to be interpreted with caution. For example, the LRH should be stable and consistent across different experimental tasks, while the language-specific effects observed in Winawer *et al*., (2007), Gilbert *et al*. (2006), Gilbert *et al*. (2008) are mediated by the concurrent tasks employed in the non-verbal experiments. In contrast, the cross-linguistic differences observed in a non-verbal similarity rating task in Phillips & Boroditsky (2003) were consistent regardless of the concurrent tasks. One main reason for such mixed findings across distinct semantic domains might be due to the specific linguistic representations involved in these studies. Language-specific effects are more readily observed in domains with cross-modal connections due to a greater degree of experience with language in these domains (Boroditsky, 2001), such as, the 'spatiotemporal metaphors' in Casasanto *et al*. (2004), '*long*' can be used to describe distance, and also time in English. That is, this linguistic experience enhances the formation of the underlying concepts, thereby reinforcing the connection and making it more evident in non-verbal experiments. Moreover, it appears that these domains with more intense implicit concepts have an ideal non-verbal scenario for examining the LRH since these internal concepts can be activated without overt linguistic labels (Vanek & Zhang, 2023).

However, these advantages might become disadvantages in a reversed way. Specifically, shared linguistic representations for multiple metaphorical expressions might lead to failure of preventing participants from using linguistic labels in silence. For instance, displaying sentences describing spatial relationship between two objects in the stimuli (e.g., *The black worm is ahead the white worm*) in Boroditsky (2001) might directly activate the concept of time (i.e., *before*) due to the same Chinese label even though they convey distinct concepts in individual context. Thus, speeding up reactions in the subsequent time matching tasks. Furthermore, since the internal concepts are stronger and can be activated without overt linguistic labels, such as in Vanek & Zhang (2022), they utilized equation symbols to represent the negation process. Without other measurements to prevent participants from using language sub-vocally, would using other linguistic labels to describe the negation process also be considered as a bias of using languages during non-verbal experiments? Therefore, the mission of creating pure non-verbal experiments in domains concerning more abstract concepts might be more challenging. Regarding the Phillips & Boroditsky (2003), the concurrent task utilized in their study might not be as effective as those examining concrete and singular lexical labels, such as Winawer *et al*., (2007), since there might be more possible ways to verbally express the gender concepts. Measurements taken to prevent participants from using language in silent should be employed according to specific linguistic representations involved in the investigation of the LRH.

The current study aims to examine the LRH in the domain of motion, which contains abstract linguistic representation, such as grammatical aspect features, and includes typological expressions that vary at lexical levels, for instance, equipollently-framed Chinese uses two verbs to describe *manner* and *path* (e.g., *man walking entering room), while satellite-framed English only uses verb to describe *manner* (e.g., man walking into room). Through three novel non-verbal experiments (e.g., b-CFS, sentence-video verification, and video-video verification), the influence of motion lexicalisation on motion perception between Chinese and English speakers was examined when linguistic involvement was prevented, encouraged, and suppressed, respectively. Thus, investigating the language-specific effects predicted by the LRH, which should not be task-specific and are expected to be observed across different experimental designs.

2.1.2 Attempts to create non-verbal experiments

2.1.2.1 Novel stimuli

As discussed above, semantic domains with cross-modal connections seem to provide an ideal foundation to reduce the possibility of verbalizing internally, for instance, Phillips & Boroditsky (2003) examined effects of grammatical gender through categorizing objects with human genders, and Casasanto *et al*. (2004) investigates the temporal perception through spatial concepts. To resemble this non-verbal scenario, studies investigating lexical items, such as *colour* and *object* often utilise novel stimuli to distract direct activation of linguistic labels (Winawer *et al*., 2007; Gilbert *et al*., 2006; Gilbert *et al*., 2008). However, the effectiveness of this practice remains questionable. Specifically, the stimuli used in Winawer *et al*. (2007) comprised various shades of blue, with the aim of eliciting novel conceptualizations of blue and reducing the potential reliance on linguistic cues in non-verbal experiments. However, relying solely on shades of blue may not be entirely reliable, as participants might still associate these shades with the linguistic label 'blue' due to their visual similarity, potentially introducing a bias influenced by indirect linguistic cues into their non-verbal performance. Similarly, Gilbert *et al*. (2006) had two variations of green and blue respectively as stimuli, and Gilbert *et al*. (2008) utilized animal shapes instead of pictures of actual animals to avoid specific lexical activation. However, the limited variations in colour and object shapes (two dogs and two cats with different shapes) might slightly reduce, though not eliminate, lexical activation. Therefore, the current study uses various motion animations reflecting the real-life scenarios, thus enhancing the activation of motion concepts. To avoid the bias of lexical activation in non-verbal contexts, these motion animations are processed through low-level detection paradigm: the b-CFS experiment. Moreover, although two additional verification experiments were conducted to confirm the language-specific effects, unlike those non-verbal experiments in previous studies, no such tasks were employed in the present study due to concerns about their low reliability in creating non-verbal context.

2.1.2.2 Concurrent tasks

Employing concurrent tasks is a typical way to deter participants from using language internally, which increases the cognitive load, so participants might not be able to use language sub-vocally during non-verbal experiments. However, different concurrent tasks function in distinct ways. Specifically, Winawer *et al*., (2007) discovered that language-specific effects disappeared only with concurrent tasks involving verbal interference, not with non-verbal ones, such as spatial concurrent tasks. Cross-linguistic differences vanished in Gilbert *et al*. (2006) and Gilbert *et al*. (2008) as well when the nonverbal task was performed simultaneously with an extra verbal interference task.

The rationale behind using verbal interference task was based on the *phonological loop* theory proposed by Baddeley (1992). According to this theory, speech is stored in memory temporarily, typically lasting less than 2 seconds. To retain this information, it needs to be repeated silently in a recycling process. Verbal interference disrupts this process by impeding and overloading the phonological loop, thus creating an appropriate context for investigating the LRH (Lucy, 2014; Gleitman & Papafragou, 2012; Perry & Lupyan, 2013). As explained by Baddeley (1992), 'the requirement for the subject to utter some irrelevant sound. This prevents the material from being rehearsed, and also interferes with any attempt to encode visual material by subvocalization. Articulatory suppression thus forces the subject to abandon the phonological storage of visually presented material, reducing the level of performance and abolishing any effect of phonological similarity or irrelevant speech' (p.284). Therefore, the disappearance of language-specific effects under conditions with verbal interference in above three studies may suggest that such effects are specific to the online processing of specific conceptual items (e.g., *colour*, *animal shapes*), thus contradicting the stable feature that the effects of language on thoughts asserted by the LRH (Lucy, 1992a, 1992b).

However, verbal interference tasks might function differently in various domains. The essential criteria for implementing a functional verbal interference task, as elucidated by Nedergaard *et al*. (2023), emphasizes that "covert language is likely to play a facilitative role in memory and categorization when items to be remembered or categorized have readily available labels" (p.464). It is more effective in examination associated with lexical labels, such as *colour* and *object*, due to a more readily phonological connection between the internal concepts and their corresponding lexical labels, such as the lexical label '*blue*' and the concept blue in Winawer *et al*., (2007). The effectiveness reduces in those with connections between covert linguistic labels, such as examining gender through object naming in Phillips & Boroditsky (2003). Specifically, in this study, participants rated object-human pairs based on the grammatical features present in the native language they speak, and this process was not interrupted when they were required to repeat English letters out loud in a verbal interference task.

Similarly, Speed *et al*. (2023) explored whether verbal interference affects the influence of *colour* on *odour* perception during verbal and non-verbal experiments. In this study, an online naming experiment was initially conducted to ascertain the specific colour associated with the odour-source objects, as participants tend to use object (e.g., apple) with concrete name rather than abstract (e.g., stinky) description words when naming odours. The results of this experiment were utilised as stimuli in the subsequent critical experiments. In the first critical experiment, participants engaged in an odour-colour matching task while concurrently undertaking either a verbal or non-verbal (spatial) interference task. Participants were presented with a string of digits (verbal), or a block pattern (non-verbal) for 4 seconds, followed by exposure to the provided odour stimuli for 4 seconds. They were then required to select a matching colour from 12 displayed options on the screen. Subsequently, participants viewed a pair of digits or block patterns and identified the one they had seen earlier. Following the critical experiment, participants were instructed to rate and name all the odours. Additionally, they performed a semantic fluency task by naming members from categories, such as ‘*animals*’ and ‘*professions*’.

The findings revealed that participants exhibited higher accuracy in the spatial interference compared to the verbal interference task. However, the performance in the matching task remained unaffected by concurrent verbal and nonverbal tasks. To further investigate the influences of the interference tasks, a second critical experiment was conducted, maintaining identical stimuli and procedures while introducing variations in the conditions. These conditions included odour-colour matching without interference, odour-colour matching with active verbal interference (recall digits by saying out loud), odour-colour matching with spatial interference, and odour-colour matching with rhythm interference (recall rhythm by tapping). The results showed that there were no significant differences across those four conditions, and the lowest accuracy was observed in the condition without the verbal interference task. This indicates that the colour influenced odour both linguistically and perpetually.

In sum, verbal interference tasks should be approached with caution when examining language-specific effects on abstract linguistic representations. One potential explanation for these findings proposed by the authors in Phillips & Boroditsky (2003) is that this verbal interference task employed may have been too simplistic and therefore ineffective. It is conceivable that repeating English letters displayed at a speed of one letter per second was too facile for participants, particularly given their high proficiency in English. Alternatively, it was plausible that the phonological loop, typically engaged by concrete linguistic labels such as colours or object names, was not activated by the subtle grammatical gender distinction. Consequently, the repetition of English letters may have paralleled the cognitive processing of abstract genders, resulting a failure of interruption. Thus, the utilization of a verbal interference task to establish a non-verbal context for offering corroborative evidence for the LRH appears to exhibit variability across diverse domains, including the domain of motion examined in the present study. Particularly, studies concerning motion lexicalisation often can only observe language-specific effects when details of motion are activated by linguistic representations priori (Gennari et a., 2002) or during (Papafragou *et al*., 2008) the non-verbal experiment, which contradicted to the rationale of using verbal interference tasks, that is, to block the online activation of the linguistic representation. Consequently, the current study uses b-CFS experiment to capture the cross-linguistic differences between Chinese and English in non-verbal experiments without linguistic involvement and measure the same stimuli in two high-level verification (with overt and covert linguistic labels, respectively) experiments to explore the language-specific effects under the influence of overt and covert linguistic labels.

2.1.2.3 Real-time processing

Furthermore, it remains unclear whether the absence of language-specific effects under the verbal interference task is due to the non-existence of the effects or an unexpected function of the verbal interference task. For instance, Trueswell and Papafragou (2010) also employed a verbal interference task concurrently during non-verbal recognition task in motion events. Despite participants exhibiting similar responses irrespective of their cross-linguistic differences, analysis of their eye movement patterns unveiled language-specific distinctions, that is Greek speakers tended to focus more on the endpoint due to the lack of grammatical aspect in Greek, while English has rich aspect features to express more details of the actions, thus directing the English speakers to pay more attention to the ongoingness of the same motion event. Consequently, this observation yielded affirmative evidence supporting the existence of language-specific effects. In this study, researchers recorded participants' eye movements as they observed an image extracted from the last frame of same motion scenes during a verbal interference task, followed by a recognition task. Participants were given 3s to view the image before engaging in the verbal interference task which involved repeating numbers displayed audibly. Language-specific patterns of eye movements were observed right before and during the verbal interference tasks. The authors concluded that language influenced thought even under conditions of high cognitive load.

Employing methodologies capable of detecting the underlying features of "thoughts" at a lower level (such as automatic detection) might prove to be more viable in this context (Thierry, 2016). For instance, Thierry *et al*. (2009) investigated the unconscious effects of language on colour perception, focussing on Greek and English. Methodologically, the research employed the recording of the event-related potentials (ERPs) during an oddball shape discrimination task and the visual mismatch negativity (vMMN) was recorded and analysed. In this study, the Greek and English participants were recruited for a colour discrimination task and their brain activities were recorded during performing the non-verbal experiments. The results indicates that language-specific effects are observed at early stage of the visual processing. Specifically, Greek speakers detected different shades of blue relatively quicker than those with different shades of green due to the specific linguistic labels for the light and darker blue. Whereas English did not show such distinction between discriminating different shades of blue and green.

As explained by Athanasopoulos & Casaponsa (2020), 'ERPs are measures of particular interest for linguistic relativity research since they provide an online measure of the different stages of visual and/or auditory processing without the need for overt responses. ERPs thus allow us to track temporal unfolding of automatic and unconscious cognitive process in response to particular events (e.g., presentation of words, pictures, sounds)' (p.400). The rationale behind the oddball paradigm lies in prompting participants to discern a ball with a distinct control feature, and control features aimed to distract participants. For instance, a ball with a distinct shape (e.g., square) amidst others, and they filled with green and blue colours. Instead of focussing on the colour, the target stimuli in the present study, participants were directed to pay attention to the shapes. Moreover, Visual mismatch negativity (vMMN) effects captured here started before 200ms, a timing attributed to semantic decision-making, as specific naming typically emerges after 200ms (Thierry, 2016). The findings suggest that Greek speakers exhibited enhanced differentiation between different shades of blue compared to their discernment of green, while the English showed no such distinction. Specifically, ERP results revealed different patterns of vMMN effects for Greek speakers when perceiving luminance deviants in blue as opposite to green, while English exhibited overall similar patterns. This suggests that the language-specific effects emerge at an early perceptual stage, even in the absence of conscious language use.

Similarly, Mo *et al*. (2011) also investigated the language-specific effects of colour through ERPs, however, unlike the conventional way of observing cross-linguistic effects in non-verbal experiments, this study observed the effects of language through the brain activities of detecting non-linguistic representations exclusively induced by the 'language brain' to support the LRH. In this study, they recruited Mandarin Chinese speakers to discriminate colour patches via the oddball paradigm. In Chinese, blue and green are presented with distinct linguistic labels, if language has effects on cognitive process, Chinese speakers should be quicker to discriminate the target colour from a different category, for instance, they should differentiate green and blue patches more easily than light and dark green/blue patches. More importantly, such difference should only be observed when the stimuli were present to the RVF, which indicating an effect exerted from the language brain (i.e., left). Specifically, in the oddball paradigm, stimuli comprised coloured squares and simple symbols. The squares were filled with four colours: a natural green, a bluish green, a greenish blue, and a natural blue, while the symbols consisted of plus ('+') or circle ('○') signs. Participants were presented with two squares with manipulated colours (between-category or within-category) displayed on the left and right sides of the screen respectively. Alongside the coloured squares, there was a plus or a circle sign shown in the middle of the two-coloured squares. The task required participants to press a button when the plus sign changed into a circle sign. The vMMN was recorded and analysed. The results confirmed that vMMN effects of the between-category deviants were significantly stronger than those of the within-category deviants in the RVF, and such difference was not found in the LVF. Thus, this is an indication of linguistic influences predicted by the LRH, which were not biased by potential linguistic recruitment, and such effects were pre-attentive and emerged at an early stage of the visual processing.

Therefore, online measurements like eye-tracking in Trueswell & Papafragou (2010) and ERPs in Thierry *et al*. (2009), Mo *et al*. (2011) offer more detailed insights into the process of linguistic influence on thoughts. This evidence suggests that despite interference task mediating the outcomes of non-verbal behaviour regardless of the cross-linguistic differences, the underlying mechanism indicates a systematic cognitive pattern paralleling the distinct linguistic features available in their specific languages.

In sum, the essence of supporting evidence for the LRH suggests that such effects of language on thoughts are stable and consistent, emphasizing the necessity to dissociate “thoughts” from linguistic labels to avoid the circulatory of examining the effects of language on language (Casasanto, 2006; Lucy, 1992; Bylund & Athanasopoulos, 2015). This major obstacle diverges research on the effects of language on thoughts into two directions. One involves seeking methodological solutions (such as eye-tracking, ERPs) to eliminate the linguistic involvement during non-verbal experiments and discover detailed underlying mechanism of the language-specific effects (Lucy, 1992a, 1992b; Bylund & Athanasopoulos, 2015; Feinmann, 2020; Theirry, 2016). Alternatively, researchers may choose to embrace the flexibility of linguistic influences on thoughts revealed by linguistic labels and move towards answering the question of ‘what extent and in which ways does language mediate cognition?’ (Zlatev & Blomberg, 2015, p.3).

Consequently, the present study not only employs a low-level detection paradigm (b-CFS) to examine the language-specific effects of motion event cognition among monolingual Chinese and English speakers, and Chinese learners of L2 English when linguistic recruitment is not possible, thus investigating the linguistic effects predicted by the relatively strong version of LRH. Two additional verification experiments (sentence-video verification and video-video verification) were also utilized to examine language-specific effects potentially exerted at higher cognitive levels, thus exploring whether consistent language-specific effects are discovered when the language use is encouraged and not controlled, as predicted by the LRH.

2.1.3 Alternative predictions for the effects of language on thoughts

2.1.3.1 Thinking-for-speaking

'Thinking for speaking (TFS)' (Slobin, 1996; 2003)' is another theorical framework that acknowledges the influence of language on cognition, focussing on examining such effects when language is in use. The initial interpretation of this claim focussed on linguistic influence reflected in verbal expression. For instance, the earliest evidence for the TFS was gathered by Slobin (1996), which compared narratives of various event scenes in English, German, Spanish and Hebrew. For instance, Spanish speakers used simple verbs (e.g., '*threw the boy*, *slipped on top of the deer*, *threw down*') to describe the change of location, while English speakers convey the same information in a relatively more complex structure (e.g., '*a deer that was behind the boy, where there was a river*'). Slobin argues 'these systematic contrasts between Spanish and English reflect different patterns of thinking for speaking—different on-line organization of the flow of information and attention to the particular details that receive linguistic expression' (p.78). Specifically, unlike the predictions proposed by the LRH, that is, that language-specific effects should not be mediated using linguistic labels, either overtly or covertly (Gumperz & Levinson, 1996). The TFS specifically predicts that language effects emerge only when the language is in operation.

However, as highlighted in the earlier discussion of the LRH, drawing conclusions regarding the influence of language on thought solely based on verbal narratives is not reliable due to the notorious circuitry of examining thought in utterance

(Athanasopoulos & Albright, 2016; Casasanto, 2016). According to TFS, “each language is a subject orientation to the world of human experience, and this orientation affects the ways in which we think while we are speaking” (Slobin, 1996, p.91) or while the ‘ongoing construction of mental representation’ (Slobin, 2003, p.160). Therefore, Athanasopoulos & Bylund (2013) further elucidate three temporal stages encompassing the process of “thinking” in conjunction with “speaking”: the phase involving preparation for subsequent verbalization, the moment of articulation, and the ensuing outcome arising from the immediate verbal production.

Moreover, TFS focusses on the “dynamic process of performing speech rather than static underlying competence” (Cook, 2015, p.155). Empirical evidence extends beyond linguistic narratives to co-verbal behaviours that occur during the ‘thinking’ process in TFS (Athanasopoulos and Bylund, 2013), for instance, gestures and eye movements. For instance, Stam (2006) investigated how Spanish (verb-framed) and English (satellite-framed) speakers express path component in verbal description of motion events and gestures during the description. The findings align with the TFS framework, showing that Spanish and English speakers exhibited distinct Gesture patterns when encoding motion events in verbal description. Specifically, gestures, such as ‘*stoke*’ for filled pause during speech, and ‘*hold*’ for unfilled pause, were recorded during describing motion events. Through coding the motion descriptions and the simultaneous gestures, Spanish speakers were observed to use more path verbs than English speakers, they also utilized path gestures more frequently when verbalizing verbs (66%) compared to English speakers (26%).

Similarly, von Stutterheim (2012) conducted a study investigating attention allocations among speakers of seven languages (Standard Arabic, Czech, Dutch, English, German, Russian and Spanish) when describing motion events. The hypothesis posited that participants would diverge in their tendency to view the motion scenes based on the grammatical (imperfective vs. progressive aspect) structures inherent in their respective languages. For instance, German speakers might describe a ‘*man walked into a room*’ with fewer temporal details regarding the actions than English speakers who might describe the same motion scene as ‘*man was walking into a room*’ or ‘*a man walked into a room*’ to indicate an ongoing action or finished action (more detailed explanation of aspects in motion are presented later in this Chapter in the section of *motion event cognition*). Thus, German, Dutch and Czech speakers were expected to focus more on the endpoint of the motion scenes due to the lack of rich aspect features in these languages, while speakers in other language groups were anticipated to pay greater attention to ongoing regions of the scenes owing to their rich aspect features for describing actions in motion events. The results confirmed the predictions and revealed that speakers who verbally convey motion events differently also exhibit distinctions in real time processing.

2.1.3.2 Label-feedback hypothesis

The Label feedback hypothesis predicts that ‘language produces transient modulation of ongoing perceptual (and higher-level) processing’ (Lupyan, 2012, p. 4). As explained in Montero-Melis *et al*. (2016), “verbal labels and conceptual categories are coactivated in a feedback loop whereby visual stimuli (e.g., the picture of a dog) activate linguistic labels (the word “dog”) in a bottom-up process, while at the same time, verbal labels in a top-down fashion activate perceptual features that are diagnostic of the category referred to by the label” (p.638). This modulation operates in two ways, up-regulation and down-regulation. According to Perry and Lupyan (2013), ‘if internally generated labels support some cognitive or perceptual process, then redundant externally presented labels can be thought to up-regulate the linguistic contribution and verbal interference to down-regulate it’ (p. 1).

In essence, down-regulation indicates an inhibitory influence on real-time processing. To illustrate, Lupyan (2009) discovered that verbal interference task had negative impact on categorizing specific features (e.g., size or colour) of an object compared to thematic features, a phenomenon akin to the impairment observed in aphasic patients. In this study, participants engaged in a categorization task where they selected the irrelevant object based on colour, size or thematic relationship, with or without a verbal interference task. For example, in a group consisting of a potato, a balloon, and a cake, participants were instructed to choose the potato if the selection was based on the theme since balloon and cake belong to one theme: party. The stimuli were presented in pictures or words.

The verbal interference task involved two main procedures, including rehearsing displayed digits verbally throughout the categorization task, and selecting the correct digit string. The time taken to make selections was recorded and analysed. The results showed that participants were significantly affected by the verbal interference task when selecting based on size and colour, but not theme, regardless of being presented with pictures or words. This indicated that the longer reaction times were not due to picture naming. The author conducted a second experiment with an identical procedure, except that the verbal interference task was replaced by a visuospatial interference task. The results showed no difference between selecting irrelevant objects based on colour or size and selecting based on theme, confirming that the different patterns observed in the verbal interference task was due to the linguistic labels. This suggests that categorization was affected by the phonological loop formed by linguistic labels, and such online influence disappeared when the loop is blocked by the verbal interference task.

Up-regulation, on the other hand, denotes a facilitation effect. For example, Lupyan & Thompson-Schill (2012) explored conceptual activation driven by verbal and nonverbal labels through a series of picture verification tasks. In the first experiment, participants judged whether the presented picture matched the cue (a word: e.g., ‘cow;’, or a sound: “mooing”) displayed to them earlier. Reaction times were recorded and analysed. The results indicated that participants reacted faster and more accurately after the verbal cue compared to the nonverbal cue. To further probe whether this label advantage was due to an unequal weight between a word or a referential label as phonological form is easier to be reproduced. Two additional types of cues, verbal-label and imitations, were introduced alongside the original verbal and nonverbal cues The results showed that participants reacted fastest in the condition with noun label cues, confirming that the response advantage observed was not solely due to the ease of reproducing speech sounds. A third experiment aimed to eliminate bias arising from utilizing the picture name for matching performance. Participants were presented with two identical pictures side-by-side, one of which was upside-down, and were instructed to select the upright picture after hearing three types of cues: label cue, sound cue, and uninformative cue. Results revealed that the advantage of the label cue persisted even when the participants were not required to make a matching judgment between the cue and picture. Lastly, a final experiment investigated whether the label cue advantage was driven by familiarity. Participants were trained with novel verbal and nonverbal cues from six categories before engaging in a speeded orientation judgment task. Findings showed that participants still reacted faster after the label cues, even though they were newly learned, indicating that this label activation is categorical and consistent.

To sum up, instead of classifying these two theoretical accounts as alternative predictions to understand the relationship between language and thought, it might be more suitable to classify them as consistent claims aligns with the LRH. Both TFS and label-feedback acknowledges the existence of language-specific effects, but they diverged into different directions. For instance, the former pays more attention to the real-time influences induced by languages, while the latter expands the time span but emphasizes those exerted by overt linguistic labels. In terms of the current study, these two frameworks will be introduced to interpret the results from the two verification experiments. Specifically, the TFS might be supported by positive results discovered in the video-video verification experiments while participants were allowed to use language to perform the non-verbal experiment. Whereas the label-feedback hypothesis might account for the findings obtained from the sentence-video verification experiment due to the involvement of overt linguistic labels. However, the ultimate goal for the present study is to examine the consistent language-specific effects predicted by the LRH, despite the possible variation of specific differences discovered at different stages of the cognitive process among the participants.

2.1.4 Effects of language on thought in the domain of motion

The current study focusses on motion events, which typically consist of one entity moving in relation to another. This is a fundamental concept that people experience every day. Unlike in other domains, such as object or colour, motion events contain multiple semantic components with dynamic relationships among each other.

As predicted by LRH, ‘speakers will select different details, different aspects, from their representations of each scene or event, depending on what language they are speaking’ (Clark, 2003, p.17). Consequently, research on the language-specific effects in the domain of motion is complex and results are mixed and inconsistent across studies. This section introduces the empirical studies in the domain of motion regarding the effects of language on thought from two main linguistic contrasts: lexicalisation of semantic components (e.g., *manner* and *path*) and grammatical segmentation related to aspect e.g., aspect vs. non-aspect).

2.1.4.1 Manner and path

Specific semantic components in motion events vary slightly depending on the two distinct types of motion events: voluntary and caused motion events (Tamly, 1985). For example, in [1], *boy* is the Figure who moves, *runs* is the *manner* which indicates the action, *into* is the *path* which expresses the direction or trajectory, and *house* is the Ground which refers to the reference that agent moves. Alternatively, it also might include an entity exhibits an extra force to an object which makes it move in relation to another entity. Such as, in [2], *kicks* indicates both the action of the agent and the cause of the moving for the ball. The type of motion events illustrated in example [1] is called voluntary motion, and the other one displayed in example [2] is called caused motion events.

(1) A boy runs into a house.

(2) A boy kicks a ball into a room.

In terms of the crosslinguistic differences of expressing *manner* and *path*, one of the most cited frameworks in motion is Talmy’s event integration (1985, 2000), which claims that almost all languages can be divided into two major categories based on their semantic structures of expressing the motion events, namely, verb- (e.g., Spanish, Japanese) and satellite-framed (e.g., English) languages. Specifically, a verb-framed language tends to describe the *path* in the main verb, whereas a satellite-framed language narrates it in the complementary element attached to the main verb, and the main verb is used to illustrate the *manner* instead. For example, in [3], *running* is the *manner*, *out* is the *path*. Whereas, in a Spanish example illustrated in [4], *saliendo* is the main verb indicates *path*, and *corriendo* is the *manner* which can be omitted sometimes, especially in voluntary motion events.

(3) A marmot running out of a cave
MANNER PATH

(4) Una marmota saliendo de una cueva (corriendo)
PATH MANNER
A marmot exiting from a cave running
('a mamount exits a cave [running]')

(Bylund & Athanasopoulos, 2015, p.3)

Consequently, as predicted by the LRH, speakers in verb- and satellite-framed languages should vary in their cognitive processes, such categorization preference for similarity between different motion scenarios. Specifically, speakers of verb-framed languages should be more sensitive to motion scenes with shared *path*, whereas speakers of satellite-framed should pay more attention to those with identical *manner*. For example, Spanish speakers should select more alternates with shared *path* in categorizing motion scenes, whereas English speakers would prefer motion scenes with shared *manner* in the same categorization task.

However, such dichotomy classification across languages has been challenged by Slobin (2004). He introduced an additional type of language: equipollently-framed, for languages using equal weight of linguistic terms to express *manner* and *path*: Thai (Zlatev & Peerapat, 2004), Mandarin Chinese, Jaminjung (Schultze-Berndt, 2000). For example, Mandarin Chinese has a unique syntactic structure called 'serial verb construction (Chen & Guo, 2009)', which allows Chinese speakers to convey *manner* and *path* in verbs equal linguistic terms. As shown in [5], *chasing* is the *manner* and *entering* is the *path*.

(5) 一个男孩 把 一只猫 赶 进 屋子。
A boy (ba) a cat chasing entering a house.
MANNER PATH
Yi2ge4 nan2hai2 ba3 yi1zhi1mao1 gan3 jin4 le0wu1zi0
'A boy is chasing a cat into a house'.

Therefore, the cognitive process exhibited by speakers in equipollently-framed languages should be different from those from satellite- and verb-framed languages. Specifically, instead of having distinct cognitive patterns between *manner* and *path* within and across languages, speakers of equipollently-framed languages should distribute attention evenly on *manner* and *path*. For example, unlike English and Spanish speakers with imbalanced preference for categorizing motion scenes, Mandarin Chinese speakers should exhibit similar preference between motion scenes with shared *manner* and *path*.

2.1.4.2 The saliency of manner

With emphasis on the attention of *manner* in equipollently-framed language, Slobin (2000; 2004; 2006) proposed the *manner salience* theory, which claims that linguistic constructions of motion events vary with regard to the degree of *manner* salience among languages. The degree of *manner* salience is measured from two perspectives: the frequency and density of *manner* verbs utilized in expressing motion events with and across languages. For example, English has rich lexicon indicating specific details of *manner*: hop, jump, spring, etc., whereas Spanish might use *saltar* refers to a certain range of *manner*s.

As a result, when exploring the cognitive patterns across languages driven by the variation of *manner* saliency, speakers in languages with higher degrees of *manner* salience should pay more attention to the *manner* of motion compared to those who speak languages with a lower degree of *manner* salience. For example, Spanish, Mandrian Chinese, and English speakers should display a gradually increasing preference towards a paired motion scene with same *manner* (Feist, 2016).

2.1.4.3 The mere evidence for supporting the LRH

In line with those findings illustrated in other domains, studies claiming to support the LRH in the domain of motion have also employed novel stimuli or employed low-level detection approaches to examine the language-specific effects on motion event when language use is highly impossible. Specifically, Kersten *et al*. (2010) claimed to support the LRH (possibly the stronger version) by providing non-linguistic evidence via non-verbal task. In this study, novel stimulus of motion events: a bug-like creature was designed to represent *manner* and path separately. English and Spanish speakers were asked to focus on either *manner* or *path* when learning to recognize the motion events. The results indicated that English and Spanish speakers were similar when classified the novel motion events based on *path*, in contrast, language-specific difference was observed when they categorized novel motion scenes based on *manner*. Specifically, English speakers were better than Spanish speakers at classifying novel motion events when focussing on *manner*.

However, similar to the previous inference for using novel stimuli (Winawer *et al*., 2007; Gilbert *et al*., 2006; Gilbert *et al*., 2008; Phillips & Boroditsky, 2003; Casasanto *et al*.,2004) to reduce the silent use of language, the current novel motion stimuli might not be able to avoid the activation of motion expressions as well. Even though Kersten *et al*. (2010) utilized bug-like subjects for performing motion events, motion components are essentially mapped universally, such as directions and actions (Radvansky & Zacks, 2014). In another study, which also claimed to support the LRH through examining motion events, Flecken, Athanasopoulos *et al*., (2015) recorded the event-related brain activities (ERPs) of English (an aspect language) and German (a non-aspect language) speakers while performing a matching task between a prime video and a target picture, which varied in manipulations between trajectory and endpoint. Instead of having real-life film or animation to display the motion events, this study used schematic stimuli to highlight the motion details (e.g., actions and directions), which consisted of a dot moving in relation to a geometrical shape (representing endpoint in the motion scene). The stimuli contained three conditions with distinct proportion of the stimuli: match (5%, both trajectory and endpoint), mismatch (75%, both trajectory and endpoint), trajectory partial match (10%), endpoint partial match (10%), and participants were required to only respond when there was a full match between the trajectory and endpoint. The results showed that a greater P3 amplitude was found when observing items in the endpoint partial match condition than those in the trajectory partial match condition in the German speaker's group, whereas no difference was discovered in the group of English speakers.

It appears that only approaches examining low-level cognitive processes can provide non-verbal evidence to support the LRH in the domain of motion. According to the language-specific effects observed in other domains, specific linguistic representations exert distinct influences on thoughts in different ways. Absence of cross-linguistic differences in cognitive processes between participants does not necessarily indicate a non-existence of language-specific effects. Particularly, in the domain of motion, the stimuli are dynamic rather than static (i.e., object), this might also affect the results examined for understanding the effects of language on thoughts. For instance, as indicated by Papafragou *et al*. (2002), 'the static pictorial format made the recovery of path information somewhat harder than that of *manner* information' (p.206).

Therefore, the present study employed novel non-verbal experiments to investigate the LRH, particularly, the b-CFS paradigm maximally reduces the possibility of using language intentionally, thus capturing the low-level processing of motion events without being biased by the subvocalization. However, the examination of specific language representation in motion events needs to be selected with caution, as abstract (Casasanto *et al*., 2004; Vanek & Zhang, 2023) and non-abstract (Gilbert *et al*., 2006; 2008) linguistic representations might exert language-specific effects in distinct ways (Boroditsky, 2001). In the domain of motion events, there are two typical routes for examining the influence of language on thought: motion lexicalisation, and motion aspects. In the following sections, evidence concerning the motion lexicalisation is illustrated and discussed first, and this is followed by evidence observed in studies examining motion aspect.

2.1.4.4 Evidence from motion lexicalisation

Mixed results were observed in studies investigating language-specific effects of motion events focussing on motion lexicalisation in non-verbal experiments, and these findings were found to diverge between offline (e.g., nonverbal similarity judgement task and recognition task) and online measurements (e.g., eye-tracking and reaction times) as well. For instance, Papafragou *et al*. (2002) and Gennari *et al*. (2002) failed to observe language-specific effects of motion events in their nonverbal tasks involving memory retrieval, but when memory task was conducted along with measurements recording real-time processing (e.g., ‘eye tracking’ in Papafragou *et al*., 2008), the previous absent language-specific effects were evident. To investigate the potential reasons, key studies (e.g., Papafragou *et al*., 2002; Gennari *et al*., 2002; Papafragou & Selimis, 2010, etc.) are reviewed in the following section.

2.1.4.4.1 Evidence in offline nonverbal tasks

A task-specific language effect was observed in Gennari *et al*. (2002), which supports the weak version of the LRH as those linguistic effects were only evident when language could assist performance in non-verbal experiments. This study is an iconic study that systematically examined the language-specific effects induced by motion lexicalisation between two contrasting languages: verb-framed Spanish and satellite-framed English. As illustrated in earlier section, the main difference between these two typological languages is the use of the main verb, that is, English speakers tended to use the main verb to describe *manner*, while Spanish speakers often express path by the main verb. In this study, Spanish and English speakers did exhibit differential encoding of *manner* and *path* in a similarity judgement task, but this emerged only when participants engaged in a linguistic production task beforehand. Specifically, participants performed three tasks in sequence in this study, namely, encoding task, memory task and similarity judgement task. The encoding task included three conditions aimed to manipulate the degrees of linguistic recruitment for the following two non-linguistic tasks. Specifically, the Naming First condition required participants to describe the unfolding motion scenes while viewing, the Free condition allowed them to view the motion scenes freely, and the Shadow condition enabled participants to view the motion scenes while engaging in a verbal interference task. Language-specific effects were not found in the recognition task regardless of the various conditions in the encoding task. In terms of the similarity judgment task, the Spanish and English speakers performed similarly apart from in the Naming First condition. That is, Spanish speakers selected more alternations with the same *path* compared to English speakers when they used language verbally before the non-linguistic task.

In contrast, Papafragou *et al*. (2002) failed to observe language-specific differences in any of their non-verbal tasks, thus their findings do not support the predictions of the LRH. In this study, Greek (a verb-framed language) and English (a satellite-framed language) speakers were examined in two experiments, each involving a description and a memory task. In the first experiment, participants described motion scenes from single static pictures (e.g., man running upstairs) and subsequently determined if they matched new pictures depicting *manner* (e.g., man running in the corridor)/*path* (e.g., man walking upstairs) alterations represented two days later. The second experiment utilized three pictures for each motion scene (e.g., picture 1: man running in the corridor, and the subject is close to the viewer; picture 2: man running in the corridor, and the subject is a bit far away to the viewer; picture 3: man running in the corridor, and the subject is far away to the viewer), displaying a more detailed progression of the events. Participants first viewed an example, judged whether the *manner*/*path* alternations were the same as those seen earlier, and then described each motion scene. Interestingly, Greek speakers employed *path* verbs significantly more frequently than English speakers when describing motion scenes, and this tendency was enhanced with increasing age. However, both groups performed comparably in the recognition and similarity judgement task.

A commonality observed in both studies lies in the absence of language-specific effects when participants viewed motion scenes freely. This suggests that the processing of motion events appears to be uniform regardless of the characteristics of linguistic representations in their respective languages unless the specific connection between language and thought is activated via task requirements. Specifically, *path* is compulsory across languages, which exhibits a pervasive influence when viewing motion holistically. However, such universal effects disappear when participants are allowed to focus more on the details in a motion event, for example, conducting non-verbal categorization task after naming encoding in Gennari *et al*. (2002). Conversely, in Papafragou *et al*. (2002), where the description task followed the similarity judgement task, diminishing the priority of linguistic-specific activation, thus language-specific effects were not observed.

Indeed, this matter is addressed in Papafragou & Selimis (2010), where linguistic prompt was manipulated during non-verbal experiments. This study investigated whether the habitual use of describing *manner* and *path* in motion by Greek (verb-framed) and English (satellite-framed) speakers would influence their non-verbal performance, both with and without a linguistic prompt. The first two experiments shared identical procedure but different linguistic prompts, that is, participants viewed motion video clips comprising one sample and two alternates featuring variations in *manner*/*path* in sequence. In the first experiment, two linguistic prompts: '*Look! The turtle is doing something*' and '*Do you see the turtle doing the same thing now*' were played to the participants while viewing the sample and alternative videos, respectively. In the second experiment, the linguistic prompts were replaced by '*Look*' and '*Do you see the same now*'. To further minimize the influence of linguistic prompts, a third experiment was conducted. In this experiment, participants viewed the sample and its two alternates simultaneously without linguistic prompts. The sample was played first (3s ahead) and placed in the middle screen, while the two alternates were played on the other two screens positioned on the left and right side of the sample screen. These three videos were played continuously until a decision was made. Language-specific effects were found in the first experiment but not in the other two. Specifically, Greek speakers exhibited a tendency to choose more alternates with the same *path* compared to English when the linguistic prompts clearly indicated the details (e.g., actions) in the motion video clips.

In contrast, Montero-Melis & Bylund (2017) found language-specific effects even without overt linguistic indication or prompts. In this study, Spanish (verb-framed) and Swedish (satellite-framed) speakers were examined in similarity assessment tasks with manipulation of linguistic encoding and verbal interference task. Three subsequential experiments were conducted. In the first experiment, participants verbally described all the motion animations before assessing the similarities among animations with shared *manner* of cause (pull or push), *manner* of object (roll or slide) or path (up, down, across, or into). In the second experiment, participants undertook the similarities assessment task after freely viewing the motion animations. In the last experiment, participants performed the similarities assessment task under a verbal (repeat digit numbers) interference task. Language-specific effects were observed in the first two experiments and disappeared under verbal interference. Specifically, in the first experiment, Spanish speakers were more likely to categorize motion animations with similar *path*, whereas Swedish speakers tended to assess similarities based on *manner*. Moreover, the correlation between verbal description and the following similarity preference was weak, and the significance was found only among animations with similar *manner* of object. This is consistent with the results obtained in the second experiment, which is that Swedish speakers tended to focus more on the *manner* of object compared to Spanish speakers even without overt language use.

Furthermore, the results across the three experiments were cross-analysed, and the authors found that motion components were encoded differently across the three experiments. Specifically, the effects of *path* were weaker under free coding which indicates that it is unlikely that participants employed language sub-vocally for the similarity assessment task, and both linguistic and non-linguistic representations were activated during free coding. The authors concluded that this is consistent with the thinking-with-language hypothesis (Wolff & Holmes, 2011), thus supporting the weak version of the LRH.

The language-specific effect seems to disappear in non-verbal experiments when it is explored between languages with partial similarities of *manner* and *path*, even after a verbal description task, i.e., with detailed information within the motion events activated prior to the non-verbal cognitive process. For instance, Wang & Li (2021b) found that Cantonese Chinese (equipollently-framed) and English (satellite-framed) speakers categorized motion events similarly even after linguistic encoding. In this study, Chinese and English speakers performed a verbal description task, followed by a similarity judgement task. Specifically, Chinese and English speakers viewed one target motion animation first, then made a selection based on two alternative animations with same *manner* or *path*, respectively. The results indicated that both Chinese and English speakers selected more alternates contained same *path* compared to those had same *manner*.

Similarly, cross-linguistic differences of encoding patterns in languages with distinct saliency of *manner* were also discovered in Feist & Ferez (2013). English and Spanish are defined as high-*manner*-salient and low-*manner*-salient language, respectively, according to the variation of *manner* salience prominent in the specific languages. In this study, English and Spanish speakers were examined in two phases: study and recognition. There were two conditions in the study phases: high-*manner*-salient (HMS) condition and low-*manner*-salient (LHS) condition. The HMS had seven motion videos shared the same *path* but distinct *manner*, and the LHS contained three motion videos, also varied in *manner*s but shared the same *path*. Participants were assigned to either HMS or LMS condition, followed by a recognition task, which required participants to judge whether the displayed videos were shown in the earlier study phase. The rationale is that as speakers of a high-*manner*-salient language, English speakers tend to pay more attention to *manner* details (Slobin, 2003). Overall, English speakers were more accurate than Spanish speakers in the recognition task. More importantly, errors in the recognition task were coded into two types: false alarm and miss, which refer to the errors for new items and old items, respectively. Spanish speakers made more errors than English speakers across all conditions when encountering new videos in the recognition task, that is, English speakers remember *manner* details better than the Spanish speakers owing to the higher saliency of *manner* in English. Critical cross-linguistic differences were observed in the patterns of recognizing old videos, specifically, English speakers had more errors in the HMS condition than the LHS condition, whereas Spanish speakers exhibited an opposite pattern. This cross-linguistic difference lies in the salient English *manner,* and the rationale is that English speakers tended to focus more on the *manner,* and identifying an old item needed to retrieve all the *manner* details in the seven stimuli in the HMS condition, and three stimuli in the LHS respectively. In contrast, Spanish speakers are less sensitive to the *manner* details, thus retrieving memory from seven or three motion videos did not result in different ways.

In sum, language-specific effects observed in the studies mentioned above indicate that language might indeed influence motion cognition, but such effects may only emerge under certain conditions. Specifically, participants who express *manner* and *path* in different ways exhibited distinct cognitive patterns in a similarity judgment task after describing motion scenes in Gennari *et al*. (2002); or with linguistic prompt in Papafragou & Selimis (2010); or with indication of focussing on details of the motion in the instruction (i.e., "*place videos with similar actions together*") in Montero-Melis & Bylund (2017). Moreover, similar effects were also discovered when the stimuli were designed to focus on a specific detail (e.g., *manner*) in motion events, as illustrated in Feist & Ferez (2013). Consequently, the advantage of linguistic activation only functions in experiments without involving cognitive load, such as similarity judgement tasks. In contrast, neither recognition tasks (differentiate *manner* or *path*), online similarity judgement tasks, nor similarity judgement tasks under verbal interference produced language-specific effects. That is, without language-specific activation or generous cognitive access, people encode motion events in a uniformed way. Specifically, they attend to the most salient element in dynamic motion events: spatial information (Tamly, 2000; Radvansky & Zacks, 2014).

2.1.4.4.2 Evidence in online nonverbal experiments

However, situation changes when using online measurements to record the real-time processing procedure of making non-verbal semantic decisions. For instance, despite that participants selected more path than *manner* alternates in the similarity judgment task in Wang & Li (2021b) regardless of the languages they speak, distinct patterns of encoding motion events between Chinese and English speakers were observed in their reaction times for making the non-verbal decisions, even under verbal interference. This suggests that the missing language-specific effects might be due to the nature of using an impropriate experiment (e.g., recognition task), rather than a simple non-existence. Consequently, one solution to further investigate the perplexing results obtained in the recognition task is to utilize online measurements, since they "tap into the moment-by-moment processes that occur in real time, allowing researchers to investigate the time course in which specific processes emerge" (Sato & Vanek, 2023, p.217).

To illustrate, Papagragou *et al*. (2008) examined if the habitual use of *path*/*manner* verbs influences the eye movements of Greek (verb-framed) and English (satellite-framed) speakers in motion perception. The experiment consisted of two conditions aimed at examining linguistic recruitment during non-verbal experiments by creating linguistic and non-linguistic contexts. In the linguistic condition, participants viewed a motion video clip, then performed a description task based on a still image extracted from the last frame of video clip they viewed earlier. A recognition task was given after the description task, and the stimuli were still images selected from the mid-point of each corresponding motion animation present earlier in the description task. In the non-linguistic conditions, identical motion videos and still images were used, and the only difference was that the verbal description task introduced between the viewing stage and recognition stage was removed.

The participants' eye movements during viewing the unfolding video clips and the subsequent still images were recorded and analysed, and the data were coded as path and *manner* regions. The results showed distinct patterns of eye movements during the description of the motion video clips which confirms that the linguistic differences did affect the cognitive process underneath, that is, Greek speakers viewed more path regions, while the English speakers focussed more on the *manner* region. More importantly, this language-specific difference also emerged without the description task. However, instead of viewing more about the regions that is more salient in each language as reflected in the linguistic condition, participants inspected more on the regions that is contrary to the language-specific difference. That is, after the motion animation stopped, English started to inspect more on the path regions of the still image, while Spanish speakers inspected more on the *manner* regions. The authors interpreted this as reversed Whorfian effects and concluded that the use of language did impact the results dramatically when the language can be used consciously for both preparing the verbalization and retrieving from memory, even in real time processing of motion events.

Most recently, Soroli (2024) recorded French (verb-framed) and English (satellite-framed) speakers' eye movements while performing three sequential tasks:

non-verbal categorization, verbal categorization, and verbal description tasks. In the study, stimuli were manipulated based on congruent path or *manner* between targets and alternates. In the non-verbal categorization task, participants viewed the target video first displaying voluntary motion scenes, then chose from two alternates shown in sequence with same *manner* or path, respectively, based on the degrees of similarity. A similar procedure was conducted in the verbal categorization task, except the target video was replaced by a sentence displayed auditorily. Decisions and time used to make such decisions were both recorded and analysed. In the description task, participants verbally described two sets of motion videos: one set is real-life film, and the other one is animated video clips. Eye movements were recorded when participants viewed the alternates in the first two experiments and viewed the main stimuli before describing in the third experiment. The descriptive results were consistent with the language-specific features in French and English, respectively. Specially, French speakers tended to use verbs to describe *path*, whereas English expressed *manner* in verbs. Moreover, French speakers tended to omit *manner* more often when viewing animation compared to real-life videos.

However, no language-specific patterns were observed in non-verbal similarity judgment task, more specifically, English speakers tended to choose more *manner* congruent alternates than those with the same *path*, whereas French speakers exhibited an opposite pattern after hearing the verbal target sentence. *path* alternates were favoured in both language groups in the non-verbal categorization task. Results of reaction times and eye movements displayed a similar pattern as the decisions made in those two categorization tasks: cross-linguistic differences were only found in verbal categorization task. Specifically, English speakers spent longer choosing alternates with same *path*, and paid more attention to alternates with same *manner*, whereas French used less time to make similar decisions.

To sum up, in line with the results discovered in studies using off-line nonverbal tasks, current evidence obtained in the on-line nonverbal experiments reveals that language indeed influences the perception of motion events, but it is not as stable as predicted by the LRH. However, there are two potential biases for having such a conclusion. Firstly, motion events are complex, and the cross-linguistic differences of linguistic representations across languages lie in the details of the motion events: the lexicalisation of *manner* and *path* components, which needs to be activated by linguistic labels or indication to the details. This contradicts the essential condition of examining the LRH: observing non-verbal cognitive processes without covert language use (Lucy, 1997; Athanasopoulos, 2012; Bylund & Athanasopoulos, 2014; Levinson, 2012).

Secondly, as a classic measurement of preventing participants from using language covertly, verbal interference does not function well in the domain of motion, at least in studies investigating motion lexicalisation. Even though it has been regarded as an efficient way of creating pure non-verbal context since language-specific effects were not discovered with its engagement (Gennari *et al*., 2002; Trueswell & Papafragou, 2010). However, such claims are questionable due to a different baseline compared to those discovered in other domains. Recall the rationale of verbal interference lies in the phonological loop (Baddeley, 1992), which indicates that information of a visual stimuli is normally stored in memory temporarily, and the long-term stored status is reinforced thought phonological rehearsal. Verbal interference interrupts such loop, thus participants were not able to silently use the related linguistic representation during non-verbal experiments.

Specifically, Winawer *et al*. (2007) found language-specific effects without utilizing a verbal interference task, whereas in motion events, the evidence is mixed. For instance, Feinmann (2020) incorporated a verbal interference task during a similarity rating task aimed at encoding motion in both English and Spanish. In this study, participants were tested with rating the degrees of similarity between one model (e.g., steps into the box) and one alternate (different-*path*: step out of the box/ different-*manner*: jump into the box) video clip in each trial, subsequent to engaging in a verbal interference (repeating nonsense syllables) task for 10s. The verbal interference task lasted throughout the non-verbal task. Additionally, a follow-up linguistic description task was conducted. The findings revealed no cross-linguistic differences in the rating of similarities for *manner* and path components between English and Spanish speakers. Consequently, the verbal interference tasks not only serve to obstruct the silent use of language but also sever the connection between superficial linguistic representations and the underlying non-linguistic representation. For instance, the language-specific effects persisted when employing a non-linguistic (spatial) interference task with a comparable level of cognitive load, as demonstrated in Winawer *et al*. (2007). This phenomenon, in turn, suggests the significant influence that language exerts on cognitive processes. However, the function of verbal interference tasks remains questionable in Gennari *et al*. (2002) and Ji (2017). Since language-specific effects were not found either with or without under the verbal interference task in the former, and no baseline was provided in the latter either. Furthermore, the verbal interference task functions under a primary condition of observing language-specific effects in motion lexicalisation, that is, to use language for directing participants' attention to motion details, as demonstrated in Montero-Melis & Bylund (2017) and Wang & Li (2021b). Thus, it is difficult to recognize whether verbal interference simply blocks the superficial linguistic recruitment, or completely cuts off the internal connection to the language-specific effects.

Similar language-specific effects should be more evident through cross-linguistic difference driven by linguistic representations with more language experience, such as grammar, since it is more profound and stable compared to superficial lexical linguistic labels (Langacker, 2008). Consequently, focussing on the variation of grammatical aspect used to express motion events might resolve the conflicts of examining *manner* and *path* encodings in nonverbal experiments, specifically, cross-linguistic differences require linguistic activation but creating non-verbal experiments needs to prohibit linguistic activation.

2.1.4.5 Evidence from motion aspect

In terms of the grammatical structures in motion events, verb aspect is the key to the linguistic distinctions among different languages. According to Madden & Zwaan (2003), "verb aspect is the grammatical construction that denotes an event's duration, onset, and completion status" (p.663). Specifically, aspect and non-aspect languages would describe motion events differently in terms of the state expressed within the verb

(Slobin, 2006). For example, in [6], a non-aspect language, such as Swedish, expresses the action in the simple present, whereas in [7], an aspect language, such as English, describes action in more detail. In terms of examining the effects of language on thought, cognitive grammar (Langacker, 2008) predicts that such a linguistic distinction directs participants to focus on different conceptual elements in motion events. For example, the ‘ongoing’ features of the motion events might be more attractive to speakers of aspect languages due to the availability of rich expression of action, and they have to obligatorily pay attention to the temporal/aspectual details of the event in order to choose the correct grammatical form of the verb, while the ‘endpoint’ features might be more highlighted for speakers of non-aspect languages due to less varied and detailed description of actions.

(6)	En pomeranier	springer	mot	ett thus.
	A Pomeranian	runs	towards	a house
		SIMPLE PRESENT		GOAL

(‘A Pomeranian runs towards a house’)

(7)	A Pomeranian	is running.
		PROGRESSIVE

Evidence regarding the language-specific effects in encoding motion events with distinct aspectual features is also mixed. For example, von Stutterheim and colleagues (2012) have explored how speakers from seven (standard Arabic, Czech, Dutch, English, German, Russian and Spanish) languages with variations of grammatical aspect features (imperfective vs. progressive aspect) encode motion events in linguistic production. They also employed a subsequent memory task, and their eye movements were recorded while undertaking the verbal description task. The results showed that speakers in non-aspect languages (e.g., Czech, Dutch and German), i.e., without grammaticalized aspect tended to describe more endpoint in motion events compared to verbal production obtained by speakers in aspect languages (e.g., Arabic, English, Russian and Spanish). Moreover, the eye-tracking data recorded while the participants were describing the events also showed a consistent pattern: speakers from non-aspect languages focussed more on the endpoint than those who from aspect language backgrounds. In terms of the memory task, only the Arabic and German speakers were found to be significantly different from each other.

In contrast, Liao *et al*. (2020) investigated how Chinese and Dutch speakers encoded motion events by examining their encoding patterns of *path* in motion description and memory tasks. The cross-linguistic difference lies in the comparison between Chinee and Dutch is that in Chinese, *zai* and *le* are two aspectual markers represent imperfective and perfective status, respectively; where Dutch mainly expresses progressive aspect through a lexical construction (i.e., *aan het*-construction) in motion events, which might not be acknowledged as a grammatical aspect despite the related disagreement. Thus, Dutch speakers should be more sensitive to the endpoint in motion events, compared with Chinese speakers due to the comparatively rich aspectual feature in Chinese. In this study, participants described motion videos before engaging in a memory task. The stimuli were manipulated based on whether reaching an endpoint in a motion scene, such as endpoint-oriented and endpoint-reached videos. Results revealed that even though Chinese and Dutch speakers described motion events differently regarding selection of trajectory and location information, they exhibited similar patterns in the memory task.

Athanasopoulos & Bylund (2013) explored the cross-linguistic difference regarding the encoding patterns of endpoint in motion events. In this study, Swedish (non-aspect) and English (aspect) speakers were recruited to view motion videos manipulated with degrees of orientation in the motion scenes. For example, imaging a motion scene of '*a cat is walking towards a house*.' An alternate with a low level of orientation (labelled as [-*endpoint*]) would mainly show the trajectory that the agent moves, such as, a scene of '*cat stopped in the middle on the way towards the house*', whereas an alternate with a high level of orientation (labelled as [+*endpoint*]) would clearly show that the agent reached the endpoint, such as, a scene of '*a cat walked into a house*'. Four (verbal description, triads matching with/without verbal interference; online triads matching) tasks were conducted. The results indicated that Swedish speakers not only mentioned more endpoint in the verbal description task but also selected more [+*endpoint*] alternates in the triads matching task compared to English speakers. However, such cross-linguistic differences were not found with the increase of cognitive load: when participants received a simultaneous verbal interference task during the triads matching task or required to select the preference when viewing all three (one target and two alternates) motion videos at same time.

Verbal interference was also found to wash out the language-specific effects in Athanasopoulos & Albright (2016), who explored the encoding of endpoints in motion events through a perception learning paradigm. In this study, native English speakers were recruited to undertake a triads-matching task, in which participants selected a more similar alternate with two levels of the goal-orientation: high-goal-orientation (a motion scene showing the subject reached the endpoint: e.g., '*walks through a door*') vs. low-goal-orientation (a motion scene with no obvious endpoint: e.g., *a person walkin*g) after viewing the target motion video (intermediate-goal-orientation: a motion scene reflecting a subject is moving towards an visualised endpoint and the subject reaches the endpoint at the end of the motion video). The training phase started after making the selection in the categorization task, in which participants received feedback corresponding to aspect features in English (an aspect language) or Swedish (a non-aspect language). Specifically, in the English training phase, the correct answers were low-goal-orientation alternates, whereas high-goal-orientation alternates were the correct answers in Swedish training phase. Participants received a green tick, or a red cross displayed on the screen for indicating whether they made a correct or wrong selection, respectively. The results showed successful training outcomes in both training sessions. To further explore whether the results obtained in the first experiment were due to verbal meditation, a second experiment with a similar procedure was undertaken. This time, participants viewed target videos under verbal interference (i.e., repeated displayed numbers), which ended right before they made selections. The results showed that the verbal interference task impacted training in both language contexts, indicating a potential online linguistic recruitment during receiving the training. More importantly, significant improvement was only discovered when participants were trained in a Swedish way, demonstrating that these English speakers might indeed were using language to facilitate their non-verbal performance. Such facilitation disappeared when the English phonological loop was blocked by the English verbal interference task. Whereas in the Swedish training conditions, the participants did not use language to facilitate their performance, thus improved performance was evident. This language-specific results demonstrates the effects of language in an unconventional way, which does not require comparing verbal or non-verbal performance between speakers with contrast language backgrounds.

Unlike the complex outcomes of employing interference tasks in studies focussing on *manner* and *path*, studies that require participants to encode motion events through the grammatical aspect obtained a clear distinction between verbal and nonverbal interference task. For instance, Flecken *et al*. (2014) discovered language-specific effects in a nonverbal sound cue recognition task under a non-verbal interference task. Participants, including Arabic (aspect) and German (non-aspect) speakers, were presented with seven motion videos accompanied continuously by the sound of ocean wave, with two or three additional sound cues (e.g., beeps) played randomly while viewing every seven motion videos. Following the viewing phase, participants were shown seven screenshots on a computer screen and asked to select the one that was played with the additional sound (beep sound). Eye movements of viewing the videos were recorded and analysed, revealing cross-linguistic differences in both the verbal description and non-verbal task. Specifically, the German speakers tended to describe more endpoints compared to the Arabic speakers, and they also focussed more on the regions of the endpoint (e.g., regions include ‘a house’ which is specific and identifiable) regardless of whether they verbalised the motion events or not.

In contrast, Trueswell and Papafragou (2010) also investigated how Greek (non-aspect) and English (aspect) speakers encode motion events in real-time processing by recording their eye movements during recognition tasks. Additionally, this study explored the influence of language use by introducing different interference tasks. Two separate experiments were conducted in this study. In the first, participants performed a recognition task after an encoding task which was conducted with and without interference tasks consisted of one non-linguistic (tapping) interference task and one linguistic (counting numbers verbally) interference task. Specifically, participants viewed all the target stimuli in the encoding task, during this stage, they needed to describe a motion scene while verbally repeating numbers or silently tapping on the table based on the displayed numbers and drumbeats, respectively. Subsequently, participants undertook the memory task, which required them to recognise whether the presented videos stimuli were viewed in the early encoding phrase. These video stimuli were either the same or different compared to those presented in the encoding stage, and the different ones were manipulated based on two conditions: an alternate with a different *manner* or a different *endpoint*. For example, if the target stimulus was '*a boy is skating towards a house*', the alternates would be '*a boy is sliding towards a house*', or '*a boy is skating towards a chair*'. The eye movements, particularly, the cumulative looking times at *manner* region or *endpoint* region, during the encoding stage were recorded, since these data demonstrates the how long the participants studied for each region. The results showed that no cross-linguistic differences were observed in the memory task, but a small language-specific effects were found in two conditions: when the interference task was to tap the drumbeats and when interference task was not employed, that is, the Greek speakers looked more at the *endpoint* regions, while English speakers viewed more about the *manner* regions.

To further examine whether the language-specific effects observed in the first experiment were due to the strategy of using language to facilitate the subsequent memory task (Gennari *et al*., 2002), a second experiment was conducted. Specifically, this experiment replicated the first under the condition of linguistic interference task, except the interference task was introduced 3s after freely viewing the unfolding motion scenes which gave the participants an opportunity to recruit languages for the subsequent memory task. Thus, the participants observed an image extracted from the last frame of same motion scenes during the interference task, and then received a recognition task. Cross-linguistic patterns of eye movements were found right before and during the interference tasks, regardless of whether they were related to language or not. This difference contradicted the preference associated with the linguistic representation specific to the native language. That is, Greek speakers viewed more of the *manner* regions instead of the language predicted *endpoint* region, while the English viewed more of the *endpoint* region, rather than the *manner* region predicted by the aspect feature in English. Therefore, the use of language not only appears to affect the categorization outcomes in non-verbal tasks but also may influence the encoding process of motion events in real time. Specifically, the verbal interference task and online linguistic recruitment seem to interact with each other in a dynamic way. Specifically, verbal interference can effectively prevent participants from using languages silently before the internal linguistic verbalization is in the dominant position.

To sum up, language-specific effects observed in studies focussing on the cross-linguistic differences of grammatical aspect are more consistent and stable across studies, compared to those found examining the encoding of *manner* and *path* in motion events. The main reason for this is the less variation of examining the grammatical aspect compared to lexicalisation in motion events. It seems that researchers are being cautious to probe the language-specific effects of motion events when examining grammatical aspects since grammar can be a bit more deep and difficult to examine for cross-linguistic differences compared to lexical representations, even though it might exert more stable patterns (Boroditsky, 2001). However, the relatively richer evidence in motion lexicalisation illustrates a complicate but more intriguing angle to investigate the relationship between language and thoughts. For example, why minor difference between linguistic prompt (‘*Look*’ vs. ‘*Look! The turtle is doing something*’) result in distinct observations of language-specific effects induced by *manner* and *path* encoding (Papafragou & Selimis, 2010), but sometimes the effects can be quite resilient and evident even under a verbal interference task (Wang & Li, 2021b). Therefore, to further understanding of the effects of language on thought in the domain of motion, the current study focusses on the examination of *manner* and *path* encoding between monolingual Chinese and English speakers in three novel non-verbal experiments, including a b-CFS experiment focussing on the low-level detection of motion signals which maximally reduces the possibility of using language sub-vocally during non-verbal experiments, a sentence-video verification experiment aiming to investigate motion event cognition when overt linguistic labels are available during non-verbal experiments, and a video-

video verification experiment exploring motion event cognition when covert linguistic representations are available during non-verbal experiments.

2.2 Cognitive transfer and motion events

This second main section of the Chapter Two continues to illustrate and discuss the effects of language on thought, but instead of focussing on cross-linguistic differences, the present Chapter extends the discussion to the field of second language acquisition. In line with the *linguistic relativity* hypothesis (LRH, Wholf, 1956), if speakers of different languages perceive reality in different ways, acquiring a second or additional language may entail not only a reformulation of linguistic representations but also an underlying cognitive shift from a source language based cognitive process to a target language based cognitive process (Athanasopoulos, 2011; Athanasopoulos, 2012; Bylund & Athanasopoulos, 2014b; Lucy, 2016). For instance, distinct cognitive patterns observed among native Russian speakers in distinguishing between “*goluboy* (lighter blue)” and “*siniy* (darker blue)” may diminish following full acquisition of English as a second language, reflecting the universal cognitive process of categorizing lighter and darker blues among native English speakers.

In line with the major concern reflected in the cross-linguistic evidence, circularity remains to be a threat for examining the cognitive shift experience by L2/ bilingual learners (Bylund & Athanasopoulos, 2014b). To illustrate, Sachs & Coley (2006) used text paragraphs to evoke the encoding of emotions in Russian-English bilinguals. Specifically, they sorted text paragraphs together without differentiating “jealousy” or “envy”, which are observed in the Russian monolinguals. They found that bilingual experienced a conceptual shift which mirrored the L2 English monolinguals in a categorization task. However, it is unknown whether such a change would also be observed without linguistic involvement, thus this evidence remains questionable for supporting the LRH.

The Conceptual transfer hypothesis (CTH, Jarvis, 2007; Bylund & Jarvis, 2011), which claims to be an extension of LRH and specifically focusses on the language-specific effects experienced by L2 learners or bilingual speakers (Odlin, 2008). As stated by Jarvis (2007), ‘second/foreign language learners and bilinguals from different language backgrounds often refer to the same objects and events in conceptually different ways and in ways that are specific to their language backgrounds' (p.44). However, instead of acknowledging a general effect of language on thought as in the LRH, CTH follows the *thinking-for-speaking* hypothesis (TFS, Slobin, 1996; 2003), thus specifies such effects when language is in use. For example, Brown and Gullberg (2011) observed conceptual transfer in Japanese learners of L2 English through *path* expressions in L1 Japanese and L2 English after viewing motion animations. The results indicated that L2 learners exhibited a lexical pattern that is different from both L1 Japanese (tend to convey *path* in verb types: e.g., *agaru* 'rise', and *hairu* 'enter') and L2 English (tend to convey *path* in adverbial types: 'up', 'into') monolinguals, that is, they expressed *path* in verb and adverbial types equally (e.g., *hairu* 'enter', *komu* 'into'). Thus, indicating a conceptual rather than cognitive transfer experienced by the L2 learners.

Cognitive (or conceptual) shift is often observed in bilingual speakers owing to their equivalent proficiency levels in L1 and L2, which is the vital condition to observe such relatively ultimate stage of cognitive process experienced by L2 learners (Bylund & Athanasopoulos, 2014b). However, the relationship between L1 and L2-based concepts are dynamic and evokes specific cognitive (or conceptual) processes before (e.g., *the influence of L1 on L2*, *convergence*) and after (e.g., *L2 influence on L1*, *L1 attrition*) a complete transformation (e.g., *restructuring*, *internationalization*) from a L1- to a L2-based cognition (Pavlenko, 2011; Jarvis & Pavlenko, 2010). The primary distinction between cognitive and conceptual transfer, as discussed in the present study, lies in whether the examination process involves language use. Specifically, conceptual transfer pertains to verbal or non-verbal examination where language use may be present, while cognitive transfer refers to instances where sub-vocalisation is suppressed during non-verbal experiments. The critical studies reviewed on the conceptual and cognitive transfer were not categorically separated, as the current chapter focuses on the overarching concept of 'transfer'. In other words, even studies employing overt language use to investigate conceptual transfer are included to illustrate the detailed process of distinguishing conceptual and cognitive evidence from those observed in equivalent monolingual speakers. Current evidence indicates the conceptual/ cognitive transfer is observed across different semantic domains: *object* (Barner *et al*., 2009; Ameel *et al*., 2009; Cook *et al*., 2006; Malt & Sloman, 2003), *emotion* (Sachs & Coley, 2006; Panayiotou, 2004), number (Athanasopoulos, 2006), *spatial* (Basseti *et al*., 2018; Park & Ziegler, 2014), *colour* (Athanasopoulos, 2009), and *motion* (Wolff & Ventura, 2009), and these specific processes are mediated by both internal predictors, such as language proficiency, and various extra-linguistic predictors, including age of onset of bilingualism (AOB), cultural immersion, length of stay in a second language speaking country, etc. (Pavlenko, 2011; Jarvis & Pavlenko, 2010; Bylund & Athanasopoulos, 2014b; Athanasopoulos, 2011).

In sum, it is conceivable that L2 learners with higher L2 proficiency or balanced bilinguals undergo certain cognitive transformations. However, these transformations may consist of multiple processes mediated by various predictors. The present section of the Chapter 2 illustrates specific cognitive/ conceptual processes exhibited by L2 learners/ bilingual speakers across various domains, followed by discussions of potential predictors affecting these specific processes. After this, key findings of cognitive transfer have been discovered in the domain of motion are illustrated and discussed.

2.2.1 Processes of cognitive transfer

2.2.1.1 Coexistence of L1 and L2 concepts

Coexistence is defined as “bilingual’s ability to maintain the categories and frames of references relevant to both languages and to use them in accordance with the constraints placed by particular languages” (Pavlenko, 2011, p.246). To illustrate, Barner and colleagues (2009) investigated how advanced English-Chinese bilinguals and English monolinguals categorized novel objects in a word extension judgement task, and found that bilinguals exhibited distinct pattern with different linguistic instructions. The syntactic distinction between English and Chinese lies in the tendency of English speakers classify novel object relatively more on shape (not material) compared to Chinese speakers, given that English is a mass-count language. In the experiment, English-Chinese bilinguals and English monolinguals were examined in two conditions, respectively. In one condition, the experimenter verbally instructed (e.g., "*Look, look at the blicket*" in English; "*Kan4 kan4 zhe4 fen2yan2*" in Chinese) bilinguals to view a standard novel object in ambiguous syntax, and then presented an alternate with same shape or same material. Participants were asked (e.g., "*Can you point at the blicket*" in English, and "*Qing2 ni3 zhi3 zhe fen2yan2*" in Chinese) to select one that matched the name of the standard object they have previously observed. English monolinguals underwent a similar procedure, except the standard novel object was linguistically labelled in either mass or count syntax (e.g., "*This is some/a wug. Have you ever seen any wug(s) before? This is some/a wug*"). The results showed that English-Chinese bilinguals resembled English monolinguals when tested in English, that is, they categorized novel objects based on shape, while they obtained a similar pattern compared to the Chinese monolinguals when tested in Chinese, and categorized novel objects based on material. Such distinct patterns align with the stage of coexistence in the process of conceptual transfer.

Similarly, Sachs & Coley (2006) explored whether Russian-English bilinguals experienced conceptual transfer in perceiving emotion. In English, the term "jealousy" can encompass both feelings of jealousy and envy, whereas Russian tends to differentiate between these two emotions with distinct terms, such as, '*revnuet'* for the emotion of jealousy, and '*zaviduet*' for the emotion of envy. The study commenced with an emotion rating task, wherein participants were asked to rate 10 emotions after reading a short story. Three groups of participants were recruited: English monolinguals, Russian monolinguals, and Russian-English bilinguals. All participants received identical instructions, except for the bilingual group, which was divided into two sub-groups and tested in English and Russian, respectively. Results were in line with the linguistic characteristics of using "jealousy" and "envy" in each language group. Russian-English bilinguals exhibited a shift in their preference based on the language condition they were tested in. Specifically, the bilinguals separated the rating for "jealousy" and "envy" for emotions expressed in stories are "jealousy" and "envy" when they were tested in Russian, and this resembled the preference pattern observed in the Russian monolingual group. In contrast, the bilinguals tested in English shared a similar pattern with the English monolinguals and tended to rate both “jealousy” and “envy” words in an envy story.

Consequently, if a coexistence of L1 and L2-based concepts is discovered in the current study, the L2 learners (native Chinese speakers) should mirror the English monolinguals when receiving instructions in English. That is, they should detect and verify the *manner* and *path* manipulation differently in the b-CFS and two verification experiments, respectively. Whereas the L2 results should shift to a L1-based pattern when the instruction is changed to Chinese. In other words, they should resemble the Chinese monolinguals and show similar patterns in detecting and verifying stimuli with *manner* and *path* manipulation in the b-CFS and two verification experiments, respectively.

2.2.1.2 The influence of the L1 on the L2/ transfer of L1 concepts

The concept of the influence of the L1 on the L2 refers to “cases where speakers’ L2 performance is guided by L1 linguistic categories, frames of reference or preference” (Pavlenko, 2011, p.246). For instance, Athanasopoulos (2006) found Japanese L2 learners employ the numeric feature in Japanese in a picture-naming task. Specifically, this study explored how the lack of number marking in Japanese influences Japanese speakers of English as a second language (L2) sorting objects in a picture-matching task with three types of stimuli: Animal ([+animate, +discrete], e.g., *three cat/ three cats), Implement ([-animate, +discrete], e.g., *three book/ three books), and Substance ([-animate, -discrete], e.g., *three waters/ three glasses of water). The linguistic contrast between English and Japanese is that English is a number-marking language which entails that their speakers are more likely to react to changes in numbers of both Animal and Implement but not Substance. In contrast, Japanese speakers can only be sensitive to changes in numbers of Animal. In this task, participants were presented with five sets of stimuli, and each set contained one target picture with five alternative pictures manipulating the changes in number from different perspectives of alternates, including Animal (alternate 2), Implement (alternate 3 & 4), and Substance (alternates 5 & 6). They needed to select one alternate they thought was more like the target one based on the visual difference between the target and alternate pictures. Both

English and Japanese monolinguals were instructed in English and Japanese, respectively, and L2 speakers were instructed in English only. The results showed that English monolinguals and advanced L2 speakers reacted to changes in numbers of Implement and Substances significantly differently, whereas Japanese monolinguals and intermediate L2 speakers treated these changes more similarly. This categorization distinction of L2 learners with different proficiency levels indicates a conceptual transfer, specifically, L2 learners with lower proficiency level are more likely to be affected by their L1 when categorizing objects, that is treating these changes in number of Implement and Substances similarly.

Basseti *et al*., (2018) also found Chinese-English bilinguals (English major students recruited in China) exhibited a pattern corresponding to the linguistic characteristics in their native language when calculating calendars. Chinese and English utilize different systems in calendar calculation, specifically, Chinese is numerically transparent (rely on digits) whereas English replies on more the specific numeric information (verbal list) when calculating calendar. For instance, Chinese uses prefix *zhou* + numeral to refers to Weekday, e.g., prefix *zhou* + 1 for Monday, prefix *zhou* + 2 for Tuesday, and prefix *zhou* + 3 for Wednesday, while English uses different verbal labels to represent this numerical information; similarly, Chinese uses numeral + *yue* suffix to refer Month, e.g., 1+ *yue* suffix for January, 2 + *yue* suffix for February, and 3 + *yue* suffix for March. In this study, Chinese-English bilinguals and English monolinguals undertook two calendar calculation (Month and Weekday) tasks manipulated in two conditions sequentially: Forward (calculated month or weekday after the stimuli) and Backward (calculated month or weekday before the stimuli), and they reported their strategy afterwards. The stimuli in both conditions were also manipulated in boundary crossing as well, varied as Within-Boundary and Cross-Boundary. The participants' reaction times in these two tasks were recorded and analysed. The predictions were that Chinese-English bilinguals should reacted similarly regardless of the direction if their underlying concept was not affected by the L2 concept, whereas they should be faster when the calculation was forward than backward if they have shifted their concept from L1 to L2. In contrast, the performance in the

Cross-Boundary conditions should be the contrary. The results indicate that although the Chinese-English bilinguals were self-reported as having high proficiency in English, the patterns of their reaction times indicate an intact underlying concept which L1 dominated.

Therefore, if the influence of the L1 on the L2-based concept is observed in the current study, the L2 learners (native Chinese speakers) should exhibit reaction patterns that mirror the monolingual Chinese speakers in all three non-verbal experiments. Specifically, they should detect the *manner* and *path* manipulation in similar ways in the b-CFS experiments and verifying target stimuli with *manner* and *path* manipulation in the two verification experiments, regardless of the prime being overt linguistic labels (i.e., the SV verification experiment) or not (i.e., the VV verification experiment).

2.2.1.3 Convergence

“Another process involves convergence of L1 and L2 categories, perspectives or frames of reference, which results in bilingual participants performing differently from speakers of both the L1 and the L2 in a way often termed ‘in-between’ performance” (Pavlenko, 2011, p.247). For instance, Ameel and colleagues (2009) discovered that Dutch-French bilinguals were similar to each in typical rating tasks (e.g., 7-points rating scale, 1 = ‘very atypical’ and 7 = ‘very typical’) on common household objects, regardless of their language of instruction and they exhibited a distinct patterns compared to the Dutch and French monolingual groups. Specifically, Dutch and French are different in categorising the same objects based on names, for example, objects named *fles* in Dutch as split into two naming categories, *bouteille* or *flacon*. In their prototype rating task, the objects involved in this task were either from the “bottles” or “dishes” category. The results indicated that the Dutch-French bilinguals were more likely to rate objects on the rough information shared among themselves which suggests a convergence of identifying the prototype. That is, the rating correlations for *fles* and *bouteille* were higher (.98) than the monolingual groups (.91), suggesting that instead of relying on the strategy obligatory in Dutch or French, bilinguals have formed a unique pattern which is not correlated to patterns observed in neither of the monolingual groups.

Furthermore, Park & Ziegler (2014) investigated how Korean, English monolinguals and Korean-English bilinguals encode spatial concepts in similarity judgement and free-sorting tasks. In English, “put on” and “put in” are categorized based on the semantic meaning of containment (e.g., ‘in’) or support (e.g., ‘on’), whereas Korean related both two forms to loosing-fitting (e.g., *NEHTA*: ‘put books in bag’, or ‘put glasses in case’) and tight-fitting (e.g., *KIKA*: ‘put bookmark in book’, or ‘put gloves on’) respectively. In the similarity judgement task, participants were presented with three images per trial containing two same (e.g., ‘put bookmark in book’, and ‘put cup in bin’) and one different (e.g., ‘put glove on’) spatial concept, and they selected the one that was different from the other two within seven seconds. Participants were instructed to make judgment based on the actions displayed in the images. In the free -sorting task, participants were presented with nine pictures contained different spatial concepts (e.g., ‘put cup on table’, ‘put food in microwave’, ‘put ring on’, ‘buttoning’, ‘put hat on’, etc.), and they needed to categorize these pictures as many groups as they can. The results revealed that the Korean and English monolinguals exhibited a distinct pattern when categorizing spatial concept. Specifically, in the similarity judgement task, the Korean monolinguals rated similarity based on loosing fitting or tight-fitting, such as, ‘put books in bag’ and ‘put gloves on’ are similar to each other, whereas the English monolinguals tended to judge ‘put books in bag’ and ‘put cup in bin’ more similar. The results in the free-sorting task are consistent to those found in the similarity judgement task, that is, Korean monolinguals classify ‘put cup on table’ and ‘put ring on’ differently as the former is loose-fitting and the later is right-fitting, while English classify these two into one category as they share the same spatial concept. Korean-English bilinguals displayed a shift from L1-like to L2-like pattern with the increase of their English proficiency and frequency of language use. Importantly, two thirds of the bilinguals exhibited a novel pattern which converged from both L1 and L2 spatial concepts, that is, instead of category ‘put cup on table’ and ‘put ring on’ into one group like English monolinguals, or category ‘put cup on table’ and ‘put cup in bin’ like the Korean monolinguals, bilinguals category ‘put cup on table' in a separate group. The authors concluded this unique categorizing pattern as being due to cognitive convergence.

As a result, if cognitive convergence is discovered in the current study, L2 learners (native Chinese speakers) should exhibit reaction time patterns distinct from both the monolingual Chinese and English speakers across all three non-verbal experiments. Specifically, they should detect the *manner* and *path* manipulation differently in the b-CFS experiment, but such a temporal difference should be smaller than those discovered in the monolingual English speakers, and larger than those found in the monolingual Chinese speakers. Similar patterns should also be evident in the two verification experiments, that is, regardless of the proceeding prime is sentence or video, L2 learners should verify the *manner* and *path* manipulation differently, and this temporal difference should be smaller when compared with those evident in the English-speaking group, and bigger when compared with the Chinese-speaking group.

2.2.1.4 Restructuring of L1 concepts

"The former process, i.e. restricting of linguistic categories, perspectives and frames of reference, is of central interest in the study of bilingualism and thought. As a result of this process, bilinguals perform, verbally and non-verbally, in ways that diverge from the L1 pattern and begin to resemble, albeit not necessarily fully, that of L2 speakers" (Pavlenko, 2011, p.247). Taking Athanasopoulos (2009) as an example for illustrating this process, he discovered that Greek-English bilinguals started to shift from L1 to L2 patterns when identifying colours with cross-linguistic differences, however such shift was not driven by an underlying transformation of colour perception. Specifically, in this study, Greek speakers uses *ble* and *ghalazio* to express darker and lighter blues, respectively, whereas English has no such distinction. Two groups of Greek-English bilinguals were recruited for a prototype placement and a perceptual similarity rating experiment. In the first experiment, bilingual participants with advanced and intermediate proficiency levels completed a colour naming task, followed by selecting a coloured chip that matched the colour of *ble* or *ghalazio* with instructions in their native language (Greek). The results showed that bilinguals recognized *ble* and *ghalazio* differently compared with the Greek monolingual, and this pattern seems to resemble the English speakers in some way, for example, they the naming agreement were 82% for *ble* and 65% for *ghalazio*.

Similarly, in studies concerning language-specific effects induced by number marking, Cook and colleagues (2006) explored whether Japanese speakers of English as second language (L2) categorized objects and substances in a distinct way compared with the Japanese and English (e.g.,) monolinguals. The cross-linguistic difference between Japanese and English regarding quantifying nouns are that Japanese doesn't differentiate mass and count nouns as it uses numerical and classifiers to express quantity, such as, *koko ni issatsu no hon ga aru*, 'here is one classifier book', and *koko ni ippai no mizu ga aru*, 'here is one-classifier water'. Whereas English uses different syntactic structure to express mass (e.g., a glass of water) and count (e.g., three books) nouns. Thus, when classifying objects, Japanese speakers might tend to distribute attentions evenly for objects and substances, while English speakers should pay attention to objects and substances in different ways. In this study, Japanese L2 users of English living in English-speaking countries were recruited for a triads-matching task. Specifically, participants were presented with one target (e.g., ceramic lemon squeezer) with nonsense name and two alternates manipulated on the substance or shape (e.g., same material: pieces of broken ceramic lemon squeezer, or same shape: a wooden lemon squeezer) subsequentially. Responses for selecting the alternate corresponding to the name of the target were recorded and analysed. The results showed that Japanese L2 users resembled the pattern of Japanese speakers observed in the Imai & Gentner (1997), that is, to categorise simple objects based on their materials, however, the sub-group (those who had been staying for a longer time in English-peaking countries: from three to eight years) of the Japanese L2 learners did not exhibit a significant distinction in comparison the English native speakers. Specifically, the longer these Japanese L2 learners stayed in an English-speaking country, they were more likely to category items based on their shapes. This indicates that even though the Japanese L2 learners maintained their perception of objects (i.e., categorising items based on materials), with increasing time spent in an English-speaking country, they started to show a tendency to react more like the English native speakers (i.e., categorising items based on shapes).

Therefore, if the L2 learners (native Chinese speakers) discovered in the current study exhibit a restructuring of L1 concepts in the examination of cognitive transfer in motion event cognition, they should be distinct from the Chinese monolinguals and started to resemble the L2 English monolinguals across all three experiments. Specifically, the time taken to detect/ verify the *manner* and *path* manipulation should be different in the L2 learners, and such temporal difference should be significantly different from those observed in the Chinese monolinguals, but start to resemble (not exactly the same) the English monolinguals in both the b-CFS and two verification experiments.

2.2.1.5 Internalization/ transfer of L2 concepts

"In cases where later learned languages encode categories, perspectives or frames of reference absent in the L1, it is also legitimate to talk about internalization of new categories, perspectives, frames and/ or patters of preference that result in target-like performance" (Pavlenko, 2011, p.247). As an illustration of this, Panayiotou (2004) investigated how English-Greek bilinguals evolved after acquiring a second language which posits emotion differently compared to their native language through an emotional description task. He found that bilinguals sometimes used English words in a Greek conversation to express a specific emotion (e.g., 'frustrated') which was absent in Greek. In this study, participants were asked to describe their emotional reaction after receiving an auditory presentation of one scenario (e.g., a text describing a close friend who believes work is more important than family), and this scenario was presented in English and Greek subsequentially and the time between these two tasks was a month for avoiding repetition effects. The verbal descriptions were recorded and analysed based on three perspectives: whether participants used a direct translation strategy when presenting emotional reactions in Greek and English, whether code-switching was identified in the description, and whether they exhibited any patterns regarding the translation strategy and code-switching. The results showed that bilinguals exhibited a distinct pattern when expressing emotions. Specifically, the descriptions in two languages were not directly translated based on one to the other, and code-switching was used when participants felt the word was more appropriate to be utilized for describing specific emotions (e.g., ‘frustrated’). This indicates an intimate connection between culture and language because participants declared the reason for not using emotional expressions available in Greek when they were speaking English was the lack of corresponding feelings, rather than a different preference for the words.

Consequently, if the internalization of L2-based concepts is observed in the current study, the L2 learners (native Chinese speakers) should mirror the English monolinguals across all three experiments. That is, the time taken to detect (i.e., in the b-CFS experiment) and verify (e.g., in the two verification experiments) *manner* and *path* manipulation should be as distinct as those reflected in the English monolinguals. However, this specific cognitive process focusses more on the transformation of cognitive features that are absent in the L1, and this might not be suitable for explaining the cognitive transfer in the current study, since the main cross-linguistic difference of motion event cognition between Chinese and English are not completely different. It is the distinct degrees of *manner* and *path* saliency between Chinese and English induces language-specific cognitive differences between Chinese and English speakers.

2.2.1.6 Shifts to L2-based concepts and L1 attrition

According to Pavlenko (2011), “prolonged exposure to the L2 may also lead to another type of conceptual transfer, namely, L2 influence on L1 linguistic categories, frames of reference or patterns of preference” (p.247). A typical process of shifting to L2-based (mirror the L2 monolinguals) concept might inevitably involve an impairment of the L1 based concept as the language experience of such concept decreases given the increasing exposure to the L2 language.

To illustrate the process of shift towards L2-based concepts, the study mentioned earlier in the convergence section, Park & Ziegler (2014), found that Korean-English bilinguals of higher proficiency level showed tendency to shift from L1 to L2-based concept. Specifically, they resembled English native speakers when categorising spatial concepts. Similarly, in the second non-linguistic experiment undertook in Sachs & Coley (2006), which showed that bilinguals performed akin to the L2 (target) monolinguals regardless of the languages being tested. In English, the term “jealousy” can encompass both feelings of jealousy and envy, whereas Russian tends to differentiate between these two emotions with distinct terms, such as, '*revnuet'* for the emotion of jealousy, and '*zaviduet*' for the emotion of envy. In a triad sorting task, participants selected two situations from three options in each triad that evoked similar emotions (e.g., a sentence with an emotion-evoking feature but did not contain the exact words "jealousy" or "envy"), and subsequently provided explanations for their selections. In the free sorting task, participants were presented with all the stimuli and instructed to categorize them into as many groups as possible. They were then asked to provide explanations for their categorizations. The results showed consistent references across all Russian monolinguals, and they were able to differentiate between jealousy and envy. However, both Russian-English bilinguals and English monolinguals showed a higher preference to categorize jealousy and envy into one group. Importantly, the results observed in the bilingual group were not affected by the test languages.

Wolff & Ventura (2009) explored whether Russian-English bilinguals experience conceptual transfer in habitually describing a caused motion. Russian and English are distinct in terms of the forces driven by the entities in a caused motion event. Specifically, Russian tends to pay more attention to the internal source of the force, whereas English speakers have no such preference regardless the force is internal or external. For instance, in "John caused Mary to break the window", Russian speakers might focus on 'Mary' who generated the action "break", thus tended to choose more CAUSE verbs, whereas English speakers might interpret the action generated by external forces: initiated by 'John', thus they might use ENABLE verbs to describe the same action. Russian, English monolinguals, and Russian-English monolinguals were recruited in the study for two experiments. In the first one, participants viewed a 3D animation followed by selecting one matching sentence that best described the animation. The animation was manipulated in tendency of the force: two control animations clearly required a CAUSE (e.g., patient opposes affector) or a ENABLE (patient pushes along with affector) verb, respectively, and one critical animation indicates a vague corresponding to the types of the verbs (e.g., patient does nothing). In terms of the sentences, two options were presented with distinct choice of verbs, namely, one had CAUSE verbs (e.g., cause, drive, lead) and the other contained

ENABLE verbs (e.g., let, help, allow). Results indicated that Russian and English monolinguals exhibited a distinct preference for the animation with vague verb indication. More importantly, Russian-English speakers resemble the patterns observed in the English monolingual groups despite the test language was Russian. Specifically, instead of only using CAUSE verbs to describe the animation like the Russian monolinguals, Russian-English speakers used ENABLE verbs to describe the animation as well, and this indicates that bilinguals experienced a shift from L1-based to L2-based concepts regarding the perception of cause motion.

A second experiment was undertaken with an identical procedure but different participants and stimuli. Four groups of participants were recruited: Russian and English monolinguals, Russian-English bilinguals, and English-Russian bilinguals. The stimuli were extended to four and sentence choices contained an additional verb phrase, and one set of stimuli contained interaction between two sentient entities which was not employed in the first experiment. The number of times that CAUSE and ENABLE verbs were chosen across different language groups were recorded and analysed. The results observed in this experiment resembled the findings discovered in the first experiment. Importantly, although English-Russian bilinguals did not resemble the Russian monolinguals like the mirrored patterns between Russian-English bilinguals and English monolinguals, they did exhibit a distinct pattern compared to those found in the English monolinguals although they received instructions in English, that is, a relatively higher proportion of choosing ENABLE verbs and a lower proportion of choosing CAUSE verbs. This clearly shows a tendency of shifting from L1 concept to L2 concept when encoding caused motion.

L1 attrition is often found with an increase in the length of stay in an L2 community, and demonstrating a specific conceptual/ cognitive process that L2 learners might experience after acquiring a second language. This process differs from both L1 and L2 based concepts or cognition, focusing on a regressive situation when compared to the L1-based concepts/ cognition. For instance, in a recent paper by Ma & Vanek (2024), they explored L1 attrition by comparing Chinese teachers of English with non-English teachers in China through a lexical comprehension and a video description task.

Lexical decisions, the time it took to make those decisions, and the participants' description were analysed. The results indicated that L1 attrition was found in both tasks. Specifically, compared to the non-English teachers, English teachers took longer to make decisions for high-frequency Chinese words and had lower frequency of using sophisticated expressions. The authors concluded that L1 attrition is discovered even for L2 learners who stay in L1-speaking community, and this also might indicate an underlying conceptual transfer which is distinct from those observed in the L1 monolinguals.

Therefore, if the L2 learners (native Chinese speakers) in the current study are observed to obtain a cognitive transfer that demonstrates a shift to L2 based concept in motion event cognition, these L2 learners should resemble the monolingual English speakers across all three experiments, and this should not be limited to the cognitive feature that is absent in the L1 (cf. cognitive internalisation). That is, the time taken to detect (i.e., in the b-CFS experiment) and verify (e.g., in the two verification experiments) the *manner* and *path* manipulation should be different and such temporal difference should pattern with the English monolinguals and distinct from the Chinese monolinguals. In terms of the L1 attrition, the L2 learners in the current study should not mirror either the Chinese or the English monolinguals in all three experiments; however, this distinction is not like the in-between pattern predicted by the convergence process. For example, if the English monolinguals detect the *manner* manipulation more quickly than those with *path* manipulation, and Chinese monolinguals used a similar amount of time to detect these two, the time taken to detect the *manner* manipulation in the L2 learners should be slower than those in the *path* manipulation, and these should be observed in the two verification experiments as well.

2.2.2 Predictors for the cognitive transfer

Recall that the LRH predicts that the accumulation of language experience facilitates and strengthens the formation of underlying cognition, thus influences thought. In terms of the cognitive transfer experienced by L2 learners, both internal and external factors are found to influence the specific cognitive processes observed in L2 learners (Cook, 2003). Internal factors such L1 and L2 proficiency, and external factors consisting of the experience of using L2 seem to be vital in the prediction of specific processes of conceptual changes. These L2 language experiences include but are not limited to L2 proficiency, language context, and language contact (Pavlenko, 2011). These concepts will be discussed below.

2.2.2.1 L2 Proficiency

L2 proficiency is vital to predict specific process of conceptual transfer since the gradual transformation of conceptual transfer often happens with the increase of proficiency (Park *et al*. 2022). For instance, in Athanasopoulos (2006), which explored how numeric features affected the categorisation of objects, that is, the lack of number marking in Japanese induces Japanese speakers to only notice numeric changes in the stimuli with Animal type (*three cat/ three cats), whereas for English speakers, apart from being sensitive to numeric changes of objects with Animal type, but also to those with Implement (e.g., *three book/ three books) and Substance (e.g., *three waters/ three glasses of water) types owing to the number marking feature in English. The authors discovered that the Japanese L2 learners of English diverged into two directions in categorizing objects depending on distinct levels of L2 proficiency. Specifically, L2 learners with higher proficiency level mirrored the English monolinguals and tended to exhibit distinct patterns to identify the changes in numbers of stimuli with Implement and Substances types, whereas those with lower proficiency level resembled the patterns observed in the Japanese monolinguals: reacted to the numerical changes in stimuli similarly regardless of the manipulated types.

Athanasopoulos (2007) continues to explore the relationship between L2 proficiency and cognitive transfer in studies focussing on the effects of number features on object perception. In this study, the author also found L2 proficiency was significantly related to the preference patterns obtained by Japanese-English bilinguals in triads matching task for selecting alternates manipulated in shape and material. Specifically, the results found that bilinguals with a higher proficiency level in L2 English, tended to choose more alternates with same Substance than those with same Shape, mirroring the pattern observed in English monolinguals. However, such a correlation between L2 proficiency and non-linguistic performance was not found in

Athanasopoulos (2009). Specifically, intermediate, and advanced Greek-English bilinguals performed similarly in categorizing blue (e.g., *ble* and *ghalazio*), which paralleled the performance of English native speakers.

One potential reason for the above mixed results regarding L2 proficiency and L2 performance might be the different measurements used to identify the L2 proficiency levels. Both Athanasopoulos (2006) and Athanasopoulos (2007) used OPT (the quick Oxford Placement Test, OPT, 2001) whereas Athanasopoulos (2009) used Nation vocabulary test (Nation, 1990). It is unknown whether these two tests have equal validity and efficiency, and it is clear that they are measuring different aspects of proficiency (grammatical versus lexical knowledge). Furthermore, Park & Ziegler (2014) also failed to find an influence of proficiency in their non-verbal similarity judgement task with Korean-English bilinguals, which raises the question of whether the use of self-report provides an accurate picture of proficiency level.

2.2.2.2 Linguistic context

Current evidence also finds that conceptual transfer may be mediated by the linguistic context, or bilingual language mode (Grosjean, 1998; Bylund & Athanasopoulos, 2014), which refers to the immediate language use before the experiment. The rationale behind the effects of language mode on the non-verbal results is often explained through the label-feedback hypothesis (Lupyan, 2012), which claims that linguistic labels can activate the corresponding underlying concepts automatically. For studies concerning cognitive transfer experienced by L2 learners/ bilingual speakers, using different languages before non-verbal experiment raises the activation levels of the corresponding language. As a result, this might bring potentially affect the validity of the examination of the cognitive transfer in L2 learners/ bilingual speakers if the non-verbal performance is the key assessment for their cognitive transfer. For example, Barner *et al*. (2009) explored how mass-count syntax affects object perception in Chinese-English bilinguals through a similarity judgement task following a word extension task. The main cross-linguistic difference regarding mass-count syntax is that English speakers tended to focus more on the shape of a novel object due to the distinct number systems used to quantify count (e.g., three books) and mass nouns (e.g., three glasses of water), while nouns are quantified by using the structure of numerical and classifiers regardless they are count or mass nouns (e.g., *san* (numerical three) *ben* (classifier) *shu* (book), "three books"). The results indicated that Chinese-English bilinguals preferred the alternate with same shape rather than materials when they were tested in English, and flexibly shifted their focus on the alternate with same material when the test language was changed into Chinese. Similarly, in Sachs & Cloy (2006), when Russian-English bilinguals were examined about rating emotions based on provided stories, tended to rate jealousy and envy to the corresponding stories respectively when the experimental instruction was Russian, which is a resulted from the distinct expression for jealousy and envy emotions in Russian language. In contrast, they rated similarly about using jealousy and envy for stories expressed the emotion of envy when tested in English, since the emotion of envy can be expressed by linguistic representations representing jealousy as well in English.

However, such linguistic context mediated effects disappear when bilinguals use one language intensively more than the other one. For example, as mentioned earlier in Wolff & Ventura (2009), they failed to observe such function driven by test language in their moderately proficient English-Russian bilinguals who performed distinctly to English monolinguals although they received instructions in English. The authors explained the potential reasons might be that these bilinguals lived with Russian families who only speak Russian, which increased their language use in Russian, thus enhanced their encoding patterns in Russian way.

In the current study, to carefully control for the influence of linguistic context, the L2 learners examined in the three experiments are assigned to monolingual (i.e., Chinese) and bilingual (i.e., English) contexts evenly. If linguistic context plays a role in mediating the non-verbal performance of the L2 learners (native Chinese speakers), the time taken to detect and verify *manner* and *path* manipulated stimuli should mirror the Chinese monolinguals (i.e., process these two manipulation similarly) when the linguistic context is Chinese and switch to pattern the English monolinguals (i.e., process these two manipulation differently) when the linguistic context is changed to

English. More specific designs and predictions in each experiment will be illustrated in Chapter three.

2.2.2.3 Language contact related predictors

Previous literature also discovered that the frequency of using L1 or L2 might affect the cognitive transfer (Pavlenko, 2011; Bylund & Athanasopoulos, 2014a). According to the attentional learning account (Smith & Samuelson, 2006), one of the associative learning theories, claims that the influence of environment on language acquisition is pervasive. Thus, the underlying formation of the corresponding concepts is context bound and lies in the different language experience individual learner has gained in the past. Consequently, language contact related predictors, such as: *the amount of time of using a language,* and *cultural immersion* are essential to the specific outcomes of conceptual transfer. To better understand the underlying mechanism of the results of L2 learners, specifically non-verbal performance in the current three experiments, these two factors are recorded in a language background information questionnaire (Appendix 1) provided to L2 learners. Detailed results and analysis will be provided in the Chapter 4.

2.2.2.3.1 The amount of time of using a language

For instance, Athanasopoulos *et al*. (2011) discovered that the encoding patterns of colour perception exhibited by Japanese-English bilinguals were related to how frequent they use the specific language. In this study, Japanese speakers were found to be more sensitive to colour boundaries within the *Blue* and *Cyan Blue* ranges since Japanese has different lexical labels for light (i.e., *mizuiro*) and dark (i.e., *ao*) blues, while English speakers has no such distinction due to the absence of specific linguistic labels for the same colours. The results showed that Japanese-English bilinguals who used more Japanese tended to be more sensitive to colour boundaries, and this mirrored the patterns observed in the Japanese monolingual group. In contrast, such an advantage disappeared when bilinguals used English more often, which was consistent to the observation in the English monolingual group.

Similarly, Bylund & Athanasopoulos (2014a) found that the frequency of using an aspect language affected the categorisation of motion events. IsiXhosa speakers with multilingual knowledge were recruited in the study, and all of them learned English as an additional language. IsiXhosa speakers tend to focus more on the endpoint of a motion scene due to a lack of grammatical aspect in isiXhosa, whereas English speakers typically pay less attention to the endpoint of an event owing to rich aspectual system in English. A triads-matching task with alternates vary at the degrees of orientation. For example, if the target video illustrated ‘a man is walking towards a building’ and an alternative with high orientation labelled [+ *endpoint*] would be ‘a man is walking into a building’ which clearly illustrated that the subject reached the endpoint, while an alternative with low orientation labelled [- *endpoint*] would be ‘a man is walking towards a building’ but stopped in the middle of the trajectory when the video stopped. The results revealed that, in addition to exposure to English in primary education, the use of aspect languages was also significantly correlated with the patterns of encoding motion events. Specifically, IsiXhosa isiXhosa L1 speakers who used more aspect languages tended to select fewer [+ *endpoint*] alternates, reflecting the patterns exhibited by the English speakers.

2.2.2.3.2 cultural immersion

Another factor that is found to affect the L2 cognitive behaviour is cultural immersion, which can also refer to the length of time one stays in a country that speaks the target language. For instance, in Cook *et al*. (2006), the longer the Japanese-English bilinguals stayed in an English-speaking country the more likely they exhibited a pattern of encoding objects that resembled English native speakers. The linguistic contrast in Japanese and English requires them to prefer material and shape, respectively, when viewing the same object.

This effect of cultural immersion was also found in Athanasopoulos *et al*. (2009) who examined whether Greek-English bilinguals experienced conceptual transfer of colour perception. The results revealed the length of stay in an English-speaking country was significantly correlated with the identification of colour boundaries. Specifically, Greek speakers were more sensitive to colour boundaries within the range of different shades of *ble* or *ghalazio* since they use different linguistic labels for lighter blue (i.e., *ble*) and darker blue (i.e., *ghalazio*), but English does have such difference. In the similarity judgement task undertaken in this study, the Greek-English bilinguals rated colours within the same category as more similar compared to those across categories when they had spent less time in an English-speaking country, and this distinction disappeared as a function of length of stay.

In sum, the empirical studies discussed above indicate that conceptual transfer is observed across different semantic domains, however, the results are mixed given the evidence obtained in verbal and non-verbal experiments. As stated by Athanasopoulos & Bylund (2014b), “learning a new language goes beyond the mastery of the formal linguistic properties and sociopragmatic elements of a new linguistic system, as it may also entail learning new ways to categorize reality and the observed world, with potentially far-reaching consequences for the entire cognitive outlook of the bilingual or multilingual speaker” (p.954). It is unclear whether one can conclude that L2 learners undergo full conceptual transfer from L1 to L2, and this will remain so unless one can eliminate the bias involved when using language during non-verbal tasks. Consequently, the majority of evidence discovered in L2 learners or bilingual speakers can only support the conceptual transfer due to the linguistic involvement during non-verbal experiments, rather than the cognitive transfer, which is not biased by the potential linguistic involvement, as predicted by the linguistic relativity.

As discussed in Chapter 2, the feasibility of applying purely non-verbal tasks varies across different semantic domains. For instance, instead of examining the syntax or lexical representations of number marking, Cook *et al*. (2006) utilized a semantic correlation between number marking and objects. Their study focussed on the habitual categorization of objects and substance observed in Japanese speakers of L2 English. However, it remains unclear whether these L2 learners employed languages sub-vocally when performing the non-verbal sorting task. In the domain of colour, it was more challenging to block rapid lexical activation during the recognition of simple colour stimuli (Athanasopoulos, 2009), casting doubt on the conclusions regarding conceptual transfer in the L2 learners. Consequently, to mitigate the circularity of examining the language-specific effects on thought, there is a growing demand in SLA for non-verbal experiments that eliminate as far as possible both covert and overt language use.

2.2.3 Cognitive transfer in L2 motion events

Cognitive change in L2 motion events is somewhat evident in previous research (Kamenetski *et al*., 2022; Wang & Li, 2021b; Ji, 2017; Montero-Melis *et al*., 2016; Athanasopoulos *et al*., 2015; Flecken, Carroll *et al*., 2015; Filipovic, 2011; Kersten *et al*., 2010; Lai *et al*., 2014; Bylund *et al*., 2013), and this is not surprising, given the lack of rich cross-linguistic findings supporting the LRH in motion events (Kersten *et al*., 2010; Flecken, Athanasopoulos *et al*., 2015). That is, investigating specific cognitive processes obtained from participants (including monolingual speakers and L2 learners) in contexts that conscious linguistic recruitment is not allowed. Even though pure non-verbal context is not achieved in all these studies, specific illustrations of cognitive/conceptual transfer and potential factors influencing these processes are provided, including five specific cognitive processes (e.g., co-existence, L2 shift, restructuring, transfer of L1, and convergence) and three predictors (e.g., language mode, age of onset of bilingualism (AOB), and language use). In the following sections, these key findings are introduced with experimental details and potential predictors found to mediate these specific cognitive processes are illustrated and discussed.

In line with the possible approaches for examining language-specific effects in non-verbal context discussed in Chapter 2, neurophysiological approach (e.g., EEG) and online measurements (e.g., reaction times, eye movements) seem to be the appropriate methods to explore the potential cognitive shift experienced by advanced L2 learners or bilingual speakers (Theirry, 2016). To illustrate, Kamenetski *et al*. (2022) employed an EEG oddball paradigm to investigate whether Turkish-Dutch early bilinguals (age of onset of bilingualism (AOB), i.e., AOB < 4) and Dutch monolinguals differ in encoding *manner* and Endpoint in motion events. Dutch is a satellite-framed language which typically uses the main verb to describe *manner*, whereas Turkish is a verb-framed language which uses the main verb to express *path*. Consequently, Dutch speakers might pay more visual attention to the regions related to the *manner* component (e.g., the actions) of the event than in comparison to Turkish speakers, who might be more attentive to the regions that indicate *path* (e.g., endpoint).

During the experiment, participants viewed a motion animation clip with a schematic human figure moving towards a destination, followed by a matching judgement task with picture stimuli extracted from the earlier animation clip (e.g., a schematic human figure skate to a tunnel). There were four conditions between the animation-picture matching pairs: full match (10%, e.g., a schematic human figure skate to a tunnel), full mismatch (70%, e.g., a schematic human figure walk to a barrier), *manner* match (10%, a schematic human skate to a barrier) and endpoint match (10%, e.g., a schematic figure human walk to a tunnel). A button press was required when the picture was the same as the preceding animation clip. Behavioural results and the EEG data recorded during the making of the matching decisions were analysed. Results showed that Turkish-Dutch bilinguals and Dutch monolinguals exhibited similar patterns regarding the matching decisions, i.e., they both had more responses in the full match condition than each of the other three conditions, and this was further confirmed by the results obtained in the EEG. Specifically, a greater P300 was found in the full match condition across two language groups, a component related to task effects. More importantly, cross-linguistic results observed in the late positivity (i.e., LP: the time window between 700 and 1000 ms) demonstrated that bilinguals obtained a more positive amplitudes in the *endpoint* match condition compared to the *manner* match condition which might be due to an influence of the relatively lower *manner* saliency in Turkish, but Dutch monolinguals did have such distinction. The authors concluded that early bilinguals obtained dual attention patterns regarding the *manner* and *path* components. Specifically, they resembled the L2 native speakers at an early stage (i.e., no between-group difference was found in the P300) but returned to an L1-like pattern (i.e., between-group difference was found in LP) at a later stage.

However, a cognitive shift was not observed in Flecken, Carroll *et al*. (2015), who examined the eye-tracking patterns of late French-German bilinguals (AO > 10) before and during a verbal description task to investigate the encoding of the *path* component in motion events. The spatial contrast between French and German is lie in the distinct linguistic structure used to express *path*, namely, French mainly expresses *path* in the main verb (e.g., *se diriger vers, avancer* ‘to direct oneself toward,’ ‘to advance'), whereas German expresses *path* in adjuncts (e.g., *entlanglaufen/ fahren* 'to walk/ drive along') and particles (e.g., 'to walk/drive to/toward'). Thus, when viewing the same motion event scene (e.g., man walking toward a car), French speakers should pay more attention to the moving entity (e.g., man) and the spatial orientation (e.g., towards a car) than those of the German speakers. In this study, French-German bilinguals performed a motion description task after viewing short real-life motion films with two levels of manipulations: a short trajectory with a close endpoint but not reached (e.g., man walking towards a car), or a long trajectory with a potential endpoint showing at a distance (e.g., car driving around the bend). Their eye movements of the critical areas (i.e., entity and the endpoint) of the two levelled stimuli before and during the verbalization were recorded and analysed. Results showed that French-German bilinguals resembled the German monolinguals in verbalization, but they sticked to their L1 French habitual patterns when observing the motion events before utterance. Specifically, these bilinguals were capable to not express *path* in main verbs (i.e., mirror the L2 German monolinguals), but they focussed more on the entity and endpoint than the German monolinguals (i.e., mirror the L1 French speakers).

Using another online measure, Wang & Li (2021b) discovered a cognitive shift in early Cantonese-English bilinguals based on the on-line reaction times taken to finish a triad matching task. In this study, bilinguals verbally described the motion animation either in English (satellite-framed) or Cantonese (equipollently-framed) priori the non-verbal similarity judgement task, then they viewed the target animation first and selected a similar one from two alternative animations (e.g., both with alterations compared to the target stimuli: one with *manner* alteration and the other one with *path* alteration) displaying simultaneously. The main difference of expressing motion events between Cantonese (Chinese) and English speakers is that the saliency of *manner* is relatively higher than *path* in English because it can be expressed by the main verbs (e.g., a man walks into a room), while *manner* and *path* share equal salience since they both can be expressed by verbs in Cantonese (e.g., *a man walks enters a room). The decision they made, and the reaction times were recorded and analysed. Results indicated that both bilingual and their English/Cantonese control groups selected more alternates with same *path* than those with the same *manner*. However, the reactions times exhibited a language-specific distinction between the two control groups, and bilinguals resembled the L2 English native speakers, regardless of the languages used in the proceeding encoding phase. To examine whether such results were due to using language sub-vocally, a second experiment with a similar procedure was conducted. Additionally, participants were required to perform additional verbal interference tasks (i.e., repeat numbers out loud in their native language) during the non-verbal similarity judgement task. The results were consistent with those observed without the verbal interference task, and the author concluded that they are an indication of a cognitive shift that was in evidence even when the use of language was prohibited.

However, such conclusion needs to be interpreted with extra caution due to the complicate situation when using verbal interference tasks to investigate the motion lexicalisation (for a detailed discussion, see Chapter 2). In general, verbal interference tasks are utilised to prevent participants from using language sub-vocally by cutting off the phonological loop created by on-line linguistic involvement (Baddeley, 1992), this is contradicted to the condition of observing the language-specific effects in motion lexicalisation, that is, to activate motion details by linguistic prompt or other linguistic involvement. The participants in Wang & Li (2021b) finished a verbal description task before the non-verbal similarity judgement task, which might be the major reason behind the language-specific finding. Moreover, the function of blocking the online recruitment of linguistic labels induced by verbal interference tasks can be abolished if this recruitment was available before the verbal interference tasks intervened (Trueswell & Papafragou, 2010). Describing motion events before the similarity judgment task might provide opportunities for linguistic recruitment to dominate the online processing, thus diminish the function of the verbal interference task. Additionally, this competition between the linguistic recruitment and verbal interference continues to be evident in L2 studies, specifically, bilingual speakers exhibit language-specific patterns contradicted to the language used in the verbal interference tasks (Athanasopoulos *et al*., 2015).

Specifically, Athanasopoulos *et al*. (2015) discovered that balanced German (non-aspect) – English (non-aspect) bilinguals were able to shift their categorical behaviour depending on the languages used in a verbal interference they received. As an aspect language, English has a richer temporal system compared to German (a non-aspect language), thus endpoint in motion events is more attentive to German speakers. In this study, bilinguals were tested in German or English in a similarity judgement task examined the encoding of endpoint in motion events. That is, bilinguals viewed two alternates and one target motion videos subsequentially. The variation between the target and the two alternates is the distinct degrees of orientation: target video displayed a motion scene with an intermediate degree of orientation (e.g., a woman is walking towards a car), and two alternates showed motion videos either with a high degree of orientation (e.g., a woman is walking into a building) that shows the subject reaches the endpoint, or a low degree of orientation (e.g., a woman is walking towards a building) which shows the subject stops in the middle of the trajectory to the endpoint, respectively. Participants made decisions after viewing all three videos.

A simultaneous verbal interference task was introduced during the non-verbal similarity task to examine the influence of verbal interference tasks on the online linguistic recruitment in bilingual speakers, and the participants were instructed to switch the language in use in the middle of the experiment. Specifically, bilinguals tested in German started from repeating numbers in German and switched to English halfway. In contrast, bilinguals tested in English started from repeating numbers in English and switched to German halfway. Results indicated that bilingual speakers were able to flexibly switch to a language-specific pattern that was on the contrary to the language used in the verbal interference task. Specifically, German-English bilinguals selected more alternates with motion completion when they repeated numbers in English than when repeating numbers in German. Consequently, even though Wang & Li (2021b) claimed a cognitive shift in their L2 results even under a verbal interference task, asking participants to perform a verbal interference task in Chinese might encourage them to use the L2 to perform the non-verbal experiments, that is, to mirror the English monolinguals and spent less time to select the *manner* than the *path* manipulation in the similarity judgement task.

Furthermore, this provides an alternative interpretation for the impairment of the verbal interference task employed in the Phillips & Boroditsky (2003) (discussed in Chapter 2: section 2.1.1.3). Recall that the participants in this study were Spanish-English and German-English bilinguals, and the verbal interference task required participants to repeat digits in English. Thus, the language-specific difference observed in the rating task might be a result of being facilitated rather than inhibited by the verbal interference task. This raises doubts with regards to the authors view that their findings supporting the LRH.

In sum, in line with non-verbal evidence observed in monolinguals with distinct language backgrounds, that is, methodology is the vital solution to create pure non-verbal contexts during experiments. Thus, using approaches that prevent sub-vocalisation in L2 learners is also essential for investigating the cognitive transfer that are predicted by the LRH. Within the limited findings from research that has attempted to achieve this goal, cognitive transfer is evident consistently across various studies, even though some of the evidence needs to be re-evaluated in future studies due to the employment of verbal interference task (Wang & Li, 2021b; Philips & Boroditsky, 2003). Consequently, in an attempt to unveil authentic cognitive processes in L2 learners, the current study uses b-CFS experiment to examine the cognitive transfer in Chinese learners of L2 English reflected by their low-level visual detection patterns. This experiment is undertaken together with two further verification experiments, exploring the cognitive process in L2 learners through high-level semantic analysis mediated by overt and covert linguistic labels.

2.2.4 Predictors influencing cognitive transfer in L2 motion events

To better understand the cognitive transfer in L2 motion event cognition examined in the three non-verbal experiments in the current study, predictors found in previous studies that investigated the cognitive transfer in L2 learners/ bilingual speakers in the domain of motion events, in non-verbal context, are reviewed and discussed below (e.g., Lai *et al*., 2014; Montero-Melis *et al*., 2016; Filipovic, 2011; Wang & Li, 2021b; Kersten *et al*., 2010; Bylund *et al*., 2013). These factors include language context, age of onset of bilingualism (AOB), and language use (long-term).

2.2.4.1 Language mode (context)

In line with the evidence discovered in other domains (Bylund & Athanasopoulos, 2014; Sachs & Cloy, 2006; Wolff & Ventura, 2009), the effects of language context (mode) are consistently discovered in the domain of motion as well. For instance, in a separate experiment conducted in Athanasopoulos *et al*. (2015), which examined the categorical patterns of encoding endpoints in motion events by late German-English bilinguals. In this experiment, participants undertook a non-verbal triads-matching experiment manipulating the degrees of orientations in motion events in different language mode (e.g., English or German). Specifically, they were required to view a target motion stimulus and two alternative motion stimuli in sequence, then decide which one of the alternative stimuli was more similar to the target one. The target stimuli were motion events with intermediate degree of orientation (e.g., woman walking toward a car), and reflecting an unreached endpoint (i.e., car). The two alternative stimuli were motion scenes with a high degree of orientation (e.g., woman walking toward a building) that demonstrated a reached endpoint, and a low degree of orientation (e.g., woman walking toward a building) that displayed a visible endpoint (i.e., building) but subject stopped in the middle of the trajectory. The results showed that bilinguals tested in German selected more alternates with high degree of orientation, than those tested in English.

Similarly, Lai *et al*. (2014) discovered a significant correlation between language mode and habitual categorical patterns exhibited by Spanish-English bilinguals (native-like in both languages). English is a satellite-framed language which commonly expresses *manner* in the main verb (e.g., *Mr. Red rolled toward the rock*), whereas Spanish conveys *path* in the main verb which belongs to the verb-framed languages (e.g., *El señor Rojo se fue rodando hacia la Piedra*, “The Mister Red went rolling toward the rock”). In this study, Spanish-English bilinguals, English monolinguals, and Spanish monolinguals were examined in two subsequential tasks: a verbal repetition task and a non-verbal similarity judgement task. Participants started by viewing a target motion animation along with a corresponding oral description given by the experimenter, then they repeated the verbal description after the target animation (e.g., Mr Red rolling rightward) finished. Following this, they moved on to the similarity judgement task and were shown two alternates (e.g., same *path*: Mr Red jumping rightward; same *manner*: Mr Red rolling leftward) in turn on the screen, and participants needed to indicate which one of the two alternates was more like the target video they viewed earlier. To provide the participants with specific language modes, the bilinguals engaged in conversation with the experimenter and received instruction in English or Spanish, respectively. The results showed that the bilinguals tested in Spanish mirrored the Spanish monolinguals, and selected more *path* alternates, where those tested in English resembled the English monolinguals, preferred more *manner* alternates in the similarity judgement task.

Similarly, Montero-Melis *et al*. (2016) also explored the influence of language mode on a subsequent similarity arrangement task undertaken by Swedish (satellite-framed) learners of L2 Spanish (verb-framed). In this study, the language mode was created by a semantic prime task that used Spanish sentences highlighted *manner* or *path* component in motion events. There were two phases in each trial in the study: encoding and testing. In the encoding phase, participants read out loud a displayed Spanish sentence, which uses a *path* verb (e.g., *El senor <u>sube</u> unos escalones con una television*, 'The man <u>ascends</u> the stairs with a television'.) or a *manner* verb (e.g., *El senor <u>empuja</u> una television por unos escalones*, 'The man <u>pushes</u> a television along the stairs.') to highlight the *path* and *manner* components, respectively. After repeating the sentence, they watched a target clip (e.g., the man pushes a gift up the hill) and then described it. This is followed by the test phase, in which participants watched motion scenes and then placed them on the computer screen. They were required to arrange the motion scenes close to each other based on actions after each trial, that is, motion scenes with similar *manner* or *path*. The results indicated that the nonverbal categorical behaviour of the Swedish L2 Spanish learners was influenced by the immediate semantic priming. Specifically, L2 learners who were primed by *manner* highlighted sentences tended to put motion scenes with same *manner* together, where those who were primed by *path* highlighted sentence arranged motion scenes with same *path* together.

However, the effects of language mode were not observed in Filipovic (2011). In this study, fluent Spanish (verb-framed)-English (satellite-framed) bilinguals were randomly assigned to two encoding conditions (i.e., describing the motion events while viewing or viewed silently) before a recognition task. The main cross-linguistic distinction between Spanish (verb-framed) and English (satellite-framed) in this study was the variation in expressing *manner*. As predicted by the *manner salience* account (Slobin, 2006), languages with distinct saliency of *manner* component encode motion events in different ways. Specifically, English is a high-*manner*-salient language that uses the main verb to describe *manner* (e.g., she skipped out of the house), while Spanish is a low-*manner*-salient language that uses the main verb to describe *path* (e.g., *Salió de la casa brincando*, 'she exited the house skipping'). Thus, the manipulation in the recognition task focussed on the *manner* component only (e.g., target model: jumping over the wall, speed-walking along the path and skipping across the road; target variant: jumping over the wall, skipping along the path and skipping across the road). Specifically, the critical stimuli in this study were motion video clips displayed in pairs, and each pair consisted of one model displayed in the encoding phase (e.g., jumping, speed-walking, skipping) and one variant (e.g., jumping, skipping, skipping) with a different *manner* illustrated in the recognition stage. The language-specific predictions were that English speakers would perform better at recognising the variant than the Spanish speakers due to the higher salience of the *manner* component in English. The results confirmed the predictions and discovered that English speakers recognised stimuli with different manners better (lower rate of recognition errors) than the Spanish speakers in both encoding conditions, and bilinguals mirrored the encoding pattern exhibited by the Spanish monolinguals and were different from the English monolinguals, regardless of the languages used in the verbal encoding phase. Specifically, Spanish-English bilinguals made more errors in recognising different *manner*s, compared to their English monolingual peers, even after describing the model video in English.

2.2.4.2 Age of onset of bilingualism

One potential reason for the mixed results obtained in studies discussed above might be the age of onset of bilingualism (AOB). Previous studies concerning the motion event cognition found that the age of onset of bilingualism (AOB) played an important role to mediate the specific processes of cognitive restructuring (Bylund, 2011), and stable language-specific effects (i.e., not affected by immediate language use, such as, language context) tended to be discovered in bilinguals who started to acquire the target languages at early ages, despite the ongoing debates of the specific age for defining early and late bilinguals in various studies (Barner *et al*., 2009; Bylund, 2009b; Lai *et al*., 2014; Wang & Li, 2021b).

For example, in a follow-up analysis in Lai *et al*. (2014), they split their Spanish -English bilingual participants into early and late bilinguals based on the age of six (i.e., cut-off of the AOB), and analysed the effects of language context in each language group. The results showed that a distinct preference mediated by the test language only emerged for bilinguals who exposed to Spanish and English at/after the age of six, whereas those who acquired those two languages before six years old sticked to *path* alternates regardless of the test languages. Similarly, Kersten *et al*. (2010) also discovered that their bilingual participants seem to react differently to the language context in their non-verbal experiments after being split into two sub-groups based on the AOB. Specifically, only bilinguals whose AOB was before 5 years old were discovered to obtain a cognitive transfer which was not influenced by the language contexts. On the contrary, late bilinguals (AO > 5) displayed a flexible shift between the two languages according to the test languages. That is, late bilinguals tested in English resembled English monolinguals and performed better than bilinguals tested in Spanish (consistent with Spanish monolinguals) when categorizing novel motion events based on *manner*.

Furthermore, this also provide alternative explanation for the results in Wang & Li (2021b), that it might be the AOB (averagely at three years old) that affect those Cantonese – English bilinguals mirrored the English monolinguals in similarity judgement task despite the language context was Cantonese. Recall that in the earlier section (**Section 2.2.3**), Wang & Li (2021b) was discussed in terms of the effectiveness of using verbal interference tasks for preventing participants from using language sub-vocally in non-verbal experiments. If bilinguals acquired the target language at early ages indeed affect the influence of language context, the previous inference of ineffectiveness of the verbal interference tasks might be wrong. Specifically, letting these bilinguals to perform verbal interference tasks in their native languages might enhanced their non-verbal behaviour to mirror those discovered in the L2 English monolinguals. However, since these bilinguals acquired the L2 at an earlier stage (3 years old), the influence of immediate language use (such as language context and verbal interference task) might be unlikely to play an important role on the non-verbal results. Thus, the results might indeed support the LRH rather than the thinking-for-speaking concluded by the authors.

2.2.4.3 Language use (long-term)

Another extra-linguistic factor observed to influence cognitive outcomes in L2 motion events is long-term language use. To illustrate, Bylund *et al*. (2013) investigated how Afrikaans speakers (self-reported as upper intermediate level of L2 English learners), English, and Swedish monolinguals encode motion events in non-verbal similarity judgement tasks. Afrikaans and Swedish are non-aspect languages which direct their speakers to pay more attention to the endpoint in a motion scene. In contrast, English has grammatical aspect which guides English speakers to focus more on the ongoing process in a motion scene, thus they are predicted to encode endpoint relatively less than Afrikaans and Swedish speakers. For example, in a motion scene: ‘a man is walking to a building’, English speakers might pay more attention to the *walking* process, while Afrikaans and Swedish speakers might focus more on the *building*. In the similarity judgment task, participants viewed two alternates and one target video with distinct degrees of goal orientation. For example, the target videos displayed an intermediate level of goal-orientation which contain a visible endpoint without being reached by the moving subject (e.g., a man is walking toward a building), alternates (i.e., [+*endpoint*]) with high level of goal-orientation illustrated a visible endpoint and it’s reached by the moving subject (e.g., a man is walking into a building), and alternates (i.e., [-*endpoint*]) with low level of goal-orientation did not show a visible endpoint (e.g., a man is walking along a road, and the end of the road is a building).

Participants viewed two alternates and one target video one by one in sequence, and they were required to judge whether the target video they viewed at the end of each trial was more like the first alternate or the second one. The results indicated that Afrikaans speakers resembled the Swedish speakers in both verbal and non-verbal tasks. Specifically, they not only mentioned more endpoints in their verbal descriptions, but also selected more [+*endpoint*] alternates in the similarity judgement task. Moreover, the results obtained from the analysis of predictors for such outcomes showed that the Afrikaans speakers maintained the linguistic patterns specific in their native language and this was not influenced by any of the predicted factors: frequency of use of English, self-reported English proficiency, and age of acquisition of English. However, the results for the non-verbal similarity judgement task found a significant correlation between the frequency of endpoint and the use of English. Specifically, the more Afrikaans speakers use English the more selection they made for [-*endpoint*] alternates.

To sum up, current evidence obtained in non-verbal experiments examining bilinguals suggest several processes related to cognitive or conceptual shift. For example, the flexibility of switching on one cognitive/ conceptual pattern under certain conditions indicates a co-existence of L1- and L2-based concepts (e.g., Athanasopoulos, *et al*. 2015; Lai *et al*. 2014); stable L2-like cognitive patterns exhibited by early bilinguals suggests a complete cognitive shift to L2 cognition (e.g., Wang & Li, 2021b; Kersten *et al*., 2010); a tendency towards the L2 cognition implies a conceptual restructuring (e.g., Bylund *et al*., 2013); sticking to L1-like cognitive patterns even with overt L2 linguistic activation demonstrates a transfer of L1-based concepts (e.g., Filipovic, 2011); and categorical patterns with features shared in both L1 and L2 reveals a conceptual convergence (Lai *et al*., 2014). Moreover, these specific processes of cognitive transfer appear to be influenced by three main extra-linguistic factors: language mode, AOB, and (long-term) language use. Note that the effects of linguistic proficiency in L1 and L2 are not mentioned above since the participants examined in these studies were mainly advanced L2 learners or bilinguals with equivalent high proficiency levels in both L1 and L2. However, these three extra-linguistic factors all arguably contribute to linguistic proficiency. Specifically, AOB is typically related to the length of staying in a L2 speaking country and the frequency of using a second language, these are the core elements influencing a general linguistic proficiency, which directly mediate the degrees of cognitive or conceptual transfer experienced by L2 learners or bilinguals (Bylund *et al*., 2012; Athanasopoulos, 2011).

Therefore, to further understanding the effects of language on thought in the domain of motion event cognition in L2 speakers, the current study also focusses on the examination of *manner* and *path* encoding in Chinese learners of L2 English in three novel non-verbal experiments, including a b-CFS experiment focussing on the low-level detection of motion signals which maximally reduces the possibility of using language sub-vocally during non-verbal experiments, a sentence-video verification experiment aiming to investigate motion event cognition when overt linguistic labels are available during non-verbal experiments, and a video-video verification experiment exploring motion event cognition when covert linguistic representations are available during non-verbal experiments. Through comparing the motion encoding between L2 English speakers and monolingual (i.e., Chinese and English) speakers, the current study illustrate the potential cognitive/ conceptual transfer predicted by the LRH.

2.3 The current study

2.3.1 Encoding manner and path in Mandarin Chinese and English

As briefly mentioned in Chapter 2, Chinese is an equipollently-framed language (Slobin, 2004) which typically uses verbs with equivalent weight to express *manner* and *path*, respectively. Such semantic features of expressing motion events are different from the satellite-framed English. This is due to a unique syntactic structure in Chinese: a serial verb construction (Chen & Gao, 2009). In a typical voluntary motion event, as illustrated in [1a], *zou3* (i.e., walking) and *jin4* (i.e., enter) are both main verbs and indicate actions of the agent: *nan2ren2* (i.e., a man). [1b] and [1c] are two typical expressions of caused motion events in Chinese, which illustrate a *BA* construction and a *ZHE* construction, respectively (Ji *et al*., 2011b).

1a. 一个男人走进厕所。

Yi2ge4 nan2ren2 zou3 jin4 ce4suo3.

One man walk enter a toilet

‘A man is walking into a toilet.’

1b. 一个男人把箱子搬进屋。

Yi2ge4 nan2ren2 ba3 yi2ge4 xiang1zi0 ban1 jin4 wu1.

One man BA one box move enter house

‘A man is moving a box into a room.’

1c. 一个男人搬着箱子进屋。

Yi2ge4 nan2ren2 ban1zhe0 xiang1zi0 jin4 wu1.

One man move ZHE box enter house

‘A man enters a room while moving a box.’

The *BA* construction often appears with ‘a resultative verb compound (Tusun, 2023, p.5)’, for example, in 1b, *move* encode Cause + *manner*, and *enter* encodes *path*. In contrast, the *ZHE* construction is typically regarded as a representation belonging to Verb-framed language (Tamly, 2000, Ji & Hohenstein, 2014a), for instance, in [1c], *enter* encodes the main verb, and Cause/*manner* is encoded by *move* (Ji *et al*., 2011b). To maximally reduce the semantic bias of overt linguistic labels, the current study only used *BA* construction in caused motion events as experimental stimuli in the sentence-video verification experiment (the detailed design is explained in the following section: sentence-video verification experiment).

Unlike the equal weight of encoding *manner* and *path* in Chinese, English is a satellite-framed language (e.g., man walking into room), which usually conveys *manner* (e.g., walking) via the main verb, and the satellite element within the same clause (e.g., into) represents *path*. As predicted by Cognitive Grammar (Langacker, 2008), Chinese and English should have distinct cognitive processes to process *manner* and *path* due to the different distribution of salience between *manner* and *path* in English and Chinese (Tamly, 2000). Specifically, when processing the same motion events, Chinese speakers should be attentive to both *manner* and *path* to a similar extent, whereas English speakers should be more attentive to *manner* than *path*.

Within the limited amount of research that has examined the cognition of caused motion events between Chinese and English, Ji (2017) is a pioneer. She was the first to explore how Mandarin Chinese and English speakers, and Chinese learners of L2 speakers (three subgroups with different proficiency levels in L2: low, intermediate, advanced) encode *manner* and *path* in a non-verbal experiment. In this study, participants engaged in a similarity judgement task while undertaking a verbal interference task. Specifically, they viewed one target motion animation (e.g., man pushing box up-hill) first, then selected a similar alternate based on two alternative motion animations manipulated in *manner* (e.g., man pushing box down-hill) or *path* (e.g., man pulling box up-hill) component in the motion events, and they were displayed simultaneously side-by-side on the screen. Meanwhile, they repeated random numbers as the interference task in order to eliminate the possibility of using language sub-vocally during the main non-verbal experiment. Their responses to the similarity judgment task and the time taken to make those judgments were recorded as analysed. The results indicated that alternates with the same *path* (e.g., '*man pushing box up-hill*' vs. '*man pulling box up-hill*') were selected more than those with the same *manner* (e.g., '*man pushing box up-hill*' vs. '*man pushing box down-hill*') across all three language groups. However, the reaction times showed distinct patterns between monolingual Chinese and English speakers. Specifically, the English group spent less time to make decisions when the target and alternate shared the same *manner*, compared to those with shared *path*. In contrast, as predicted, the Chinese speakers spent an equal amount of time making both judgments on both types of stimuli. In terms of the Chinese learners of L2 English, the more highly proficient group mirrored the English monolinguals, in contrast to the less proficient learners. In other words, low- and intermediate-level learners patterned more akin to the Chinese monolinguals and spent a similar amount of time selecting alternates with same *manner* and *path*, whereas advanced L2 learners patterned the English monolinguals and selected alternates with same *manner* more quickly than those with the same *path*.

Following Ji (2017), Wang and Li continued to investigate the encoding of *manner* and *path* between Cantonese (Chinese) and English speakers through the non-verbal similarity judgement tasks in their two separate studies: Wang & Li (2021a) and Wang & Li (2021b). These two studies shared similar procedures with a few differences. The main difference between these two studies is that Wang & Li (2021a) also included monolingual Japanese (a verb-framed language) speakers, and Japanese-English-Cantonese multilingual speakers, no interference tasks were involved during the similarity judgement task, and the motion stimuli was caused motion events. Whereas Wang & Li (2021b) focussed on the monolingual Cantonese, English, and Cantonese-English bilinguals, and the verbal interference task was examined carefully in two separate experiments, and the stimuli were voluntary motion events only.

Specifically, in Wang & Li (2021a), five groups of participants were recruited, including monolingual Japanese, English, and Cantonese groups, Cantonese-English bilingual speakers, and multilingual Japanese-English-Cantonese multilingual speakers. The inclusion of Japanese speakers sheds light on encoding motion events with a more contrasting feature, that is, Japanese is a verb-framed language, which uses the main verb to describe *path* (e.g., *Kara wa ni-o oshite michi-o watalimashita*, 'she crosses the road pushing the goods'), thus might direct Japanese speakers to be more attentive to the *path* manipulation when viewing the same motion event, compared to the English speakers (i.e., satellite-framed language, and uses main verb to describe *manner*, e.g., she pushes the goods across the road). In this study, participants performed a verbal descriptive task and a subsequent similarity judgement task. The detailed procedure of the similarity judgment task followed Ji (2017), but without an interference task. The results in the verbal description task were in line with the cross-linguistic predictions made by the authors, that is, the cross-linguistic differences of encoding motion events should be evident in the *manner* component, that is, monolingual English speakers should be more attentive to *manner* since English uses main verbs to describe *manner*, and Cantonese speakers should encode *manner* and *path* similarly, while the Japanese speakers should be more attentive to *path* since Spanish uses the main verb to describe *path*. The results confirmed the prediction and showed that the encoding of *path* was similar across all five groups. In contrast, there was a main effect of language group which correlated with the encoding of *manner* of cause. Specifically, bilinguals mirrored English monolinguals and encoded *manner* significantly more than Cantonese monolinguals. Multilinguals patterned with the Japanese monolinguals encoded in a *manner* less than the Cantonese monolinguals. The results obtained from the semantic decisions in the non-verbal experiment were contradicted to the cross-linguistic predictions, and all five language groups preferred alternates with the same *path* rather than *manner*. However, the time took to make such decisions indicated a language-related difference. Specifically, the English and bilingual speakers were similar in that they spent less time to select alternates with the same *manner* (e.g., *man pushing box down-hill*) than those with same *path* (e.g., '*man pulling box up-hill*'). In contrast, the Japanese and multilingual speakers patterned together, and were quicker to select alternates with same *path* compared to those with same *manner*. The Cantonese monolinguals used similar amount of time to select *manner*/ *path* matched alternates, as predicted by the authors.

However, semantic decisions of selecting alternates with match *manner* and *path* are found to be affected by the language-specific effects in Wang & Li (2021b), and such effects disappeared with introducing a verbal interference task. In this study, following Wang & Li (2021a), an identical procedure was applied in the first experiment. Specifically, monolingual Cantonese, English, and Cantonese-English bilinguals speakers performed a verbal description task before the non-verbal similarity judgment task, and the bilingual speakers were examined in monolingual and bilingual contexts evenly. In the second experiment, participants performed the same verbal description and similarity judgement tasks with identical stimuli, but they had to undertake a verbal interference task throughout the non-verbal experiment. Participants were instructed to perform the verbal interference task in their native language. The results showed that participants vary at selecting alternates with different manipulations in the first experiment. That is, the monolingual English speakers and Cantonese-English bilinguals (in both language contexts) selected more alternates with same *manner* (e.g., '*man pushing box down-hill*') than those with same *path* (e.g., '*man pulling box down-hill*') after viewing the target motion video (e.g., '*man pushing box up-hill*'). Whereas Cantonese did not exhibit preference among the two alternates.

However, such language-specific preference disappeared in the second experiment when verbal interference was introduced. That is, all participants selected more alternates with matching path component. Surprisingly, in line with the Ji (2017), language-specific differences were discovered in the reaction times taken to make the similarity judgement in both experiments. The authors concluded that their results were consistent with the *thinking-for-speaking* hypothesis (Slobin, 2003), which claims that the language-specific effects can only be discovered when language is in operation.

To sum up, it seems that the RT observed in these three studies share the same language-specific pattern, which is consistent regardless the stimuli is voluntary or caused motion events, and these effects were persistent under a verbal interference task. In contrast, the semantic decision discovered in these studies are distinct. Specifically, monolingual English and Cantonese speakers seems to exhibit language-specific difference when the motion stimuli is voluntary motion events, and this language-specific effects were mediated by the online linguistic recruitment (Wang & Li, 2021b). Whereas when the stimuli are caused motion events, which is supposed to exhibit more language-specific difference because of the distinct motion typologies in Chinese and English (Slobin, 2004), however, both Ji (2017) and Wang & Li (2021a) failed to support their language-specific predictions. The reasons are unclear. Even though Ji (2017) concluded that the absent cross-linguistic differences in the semantic decisions in the non-verbal similarity judgment task was due to the verbal interference task, but this needs to be interpreted with caution. Since this study did not have a control experiment that used identical procedure without employing a verbal interference task. Previous literature indicates that the cross-linguistic difference discovered in the encoding of motion lexicalisation often emerges with the assistance of linguistic prompts (Papafragou & Selimis, 2010; Montero-Melis & Bylund, 2017) or a proceeding verbal description task (Gennari *et al*., 2002; Wang & Li, 2021b). Ji (2017) did not include this condition; thus, it might be reasonable to have the missing language-specific effects on the sematic analysis of *manner* and *path*. However, this inference failed to explain the missing language-specific difference in the semantic selections in Wang & Li (2021a), which did employ a verbal description task to boost the potential language-specific effects expected in the subsequent similarity judgment task.

Therefore, the current study focusses on the caused motion event, and further explores the language-specific effects in non-verbal experiments which maximally reduces the possibility of using language sub-vocally, and examine the influence of linguistic labels on the non-verbal semantic encoding of manner and path among monolingual Chinese and English speaker, and Chinese L2 English learners, the current study uses three novel non-verbal experiments which aims to examine the low-level processing (i.e., via the b-CFS experiment) of *manner* and *path* when online linguistic involvement is forbidden, and also their high-level processing when overt (i.e., via the sentence-video verification experiment) and covert (i.e., via the video-video verification experiment) online linguistic recruitment are available, respectively. More details on these methodological tools are set out below.

2.3.2 Breaking Continuous Flash Suppression (b-CFS) paradigm

In this section, details of the b-CFS experiment are introduced, including the rationale of this paradigm and the attempts to employ it in the current exploration of motion event. The rationale for b-CFS is that 'more conspicuous stimuli are harder to suppress by CFS and, as such, break through suppression faster than less conspicuous stimuli (Gayet *et al*., 2014, p.3)'. Specifically, the manipulated saliency of *manner*/*path* in the target stimuli will affect the processing time, which is needed to break the continuous flash suppression, i.e. the time in which the stimuli emerge into awareness.

Continuous Flash Suppression (CFS) (Tsuchiya and Koch, 2005) is a psychophysical technique used to detect the unconscious processing of visual stimuli, which is developed based on binocular rivalry and flash suppression (Stein 2019). In CFS experiments, participants see different images per eye, normally one eye is presented with a low contrast target stimulus, and the other eye is presented with a high contrast dynamic mask flashing continuously around 10Hz aimed at distracting the participants and preventing them from consciously visualising the target stimulus, and this can last for a relatively long time (e.g., up to 3 minutes).

According to Yang *et al*. (2014), there are three main CFS paradigms (e.g., visual adaptation, visual priming, and breaking suppression) used since the first CFS experiment was introduced. Both visual adaptation and visual priming paradigms manipulate the unconsciousness of the proceeding procedure to examine its influences on a later categorical decision, detection sensitivity, and time used to make conscious judgment were recorded and analysed. To illustrate, Baumann & Valuch (2022) examined whether CFS affects conscious processing of different scenes (e.g., indoor: a picture of a kitchen; outdoor: a picture of a car park). By implementing CFS phase proceeding a categorical task, participants viewed a critical stimulus presented to their non dominant eye and a flickering mask presented to their dominant eye with full contrast throughout the CFS phase. Subsequently, they moved to a decision-making phase by indicating the category of the target stimulus presented (e.g., indoor or outdoor). Responses were recorded and analysed. The results showed that participants were faster in making decisions when the stimulus presented in the CFS phase was categorically congruent with the target stimulus. However, even though CFS can be used to examine the influence of unconscious priming on of processing the target stimulus, the categorical judgement procedure required conscious awareness.

In contrast, breaking suppression does not involve conscious semantic decisions being made throughout the experiment, which maximally reduces the possibility of using language sub-vocally in non-verbal experiments investigating the effects of language on thought. In a typical b-CFS paradigm, the low contrast target stimulus is presented to the non-dominant eye, while the high-contrast flickering Mondrian-like mask is presented to the dominant eye. The contrast of the target stimulus increases from 0% to 100% in 10 seconds, while a high contrast flickering mask is displayed on the other eye. The target stimuli are located on the right or left side of the vision area, and the task for the participants indicates the location of the target stimuli, rather than simply recognising whether it is present or not. Moreover, the time that is used to break through the suppression of the flickering mask and reach awareness (i.e., indicating the location) is analysed, effectively reflecting a maximal time frame of unconscious processing. Thus, b-CFS allows researchers to capture the automatic process of a stimulus entering consciousness from unconsciousness with enough time (Stein, 2019), given that the boundaries between these two often were challenging to detect due a rapid transformation (Moors & De Hower, 2006).

The b-CFS experiment can be employed to investigate the influence of familiarity of the visual stimuli (e.g., Jiang *et al*., 2007) or the violation between upcoming visual stimuli and concurrent information (e.g., Costello *et al*., 2009). For instance, Jiang *et al*. (2007) used b-CFS to examine whether familiarity facilitates the unconscious processing of faces and characters by comparing the reaction times between an upright face, an inverted face, a Chinese character, and a Hebrew word. The results show that the participants spent less time processing familiar and recognisable stimuli unconsciously. Specifically, both Chinese and Hebrew speakers processed upright faces more quickly than those with inverted faces; and more interestingly, Chinese speakers spent less time breaking through the suppressing mask with stimuli that displayed Chinese characters rather than Hebrew words, and vice versa for Hebrew speakers, which indicated that the familiarity exerted by language can also be detected by the b-CFS. Moreover, Costello et al. (2009) found that visual stimuli paired with congruent proceeding primes allowed participants to break through the suppression of the flickering mask more quickly than those paired with incongruent primes in a b-CFS experiment. Specifically, participants were required to identify the presence of a target word obscured by a flickering mask after viewing a prime word. The time taken to identify the target word was recorded and analysed. The target and prime words either shared congruent or incongruent semantic meanings. The results confirmed that participants were able to identify target word faster when the prime word conveyed a congruent semantic meaning. To summarise, b-CFS records and analyses the time required to identify visual stimuli suppressed by a flickering mask, with reaction times varying based on the familiarity of the visual stimuli, or the congruency between a prime and target stimulus.

2.3.3 The b-CFS in language and cognition

The current study is the first attempt to employ b-CFS to examine the potential language-specific effects on motion event processing in monolingual English and, Chinese speakers, and Chinese learners of English speakers. Although there is no direct evidence regarding the cross-linguistic differences obtained by b-CFS in motion, there are a few studies that performed b-CFS or CFS experiments to examine the effects of auditory linguistic labels on visual recognition or discrimination (Paffen *et al*., 2021; Lupyan & Ward, 2013; Ostarek & Huettug, 2017), suggesting that effects of language on perception can be observed both in 'higher-level processes such as recognition and in lower-level processes such as discrimination and detection' (Lupyan *et al*., 2020, p.930).

Very recently, Paffen *et al*. (2021) utilised b-CFS to examine whether congruent and incongruent verbal linguistic labels affect the automatic processing of simple object objects and colours. In this study, the participants performed a b-CFS with an auditory verbal linguistic cue (e.g., 'red') mapped with three alternative targets with different colours: congruent (e.g., a red colour), incongruent (e.g., a green colour), or neutral (e.g., a neutral noncolor word). The participants started from a fixation point, and heard the cue at one of the three stages (i.e., congruent, incongruent, and neutral) illustrated earlier, then pressed one of two preassigned buttons to indicate which side of the screen displayed the coloured targets as soon as they could identify them. The results indicated that participants were faster to break through the mask and identify the location of the target stimuli when carried out by a congruent cue, compared to those in the neutral and incongruent conditions. However, they did not react more slowly than in the neutral condition when the cue and target were incongruent.

Similarly, Lupyan & Ward (2013) found that verbal linguistic labels affect the unconscious processing of the presence of simple objects via a CFS paradigm. In this study, participants listened to auditory cues (i.e., a matching label or a noise) before moving on to the CFS phase, which presented object (e.g., a pumpkin) to the dominant eye and a distract mask to the nondominant eye for 1.5 seconds. The participants were required to make decisions about whether they saw an object as soon as possible and continued to answer a question regarding the object displayed in the CFS (e.g., 'was the object you saw a pumpkin?'). The results showed that participants were more sensitive to stimuli with a valid label compared to those without, and they also spent less time identify a hidden object after listening to a congruent linguistic label. This indicates that verbal cues facilitate the detection of stimuli with congruent semantic meanings.

In sum, current evidence obtained by use of b-CFS or CFS indicates that language affects low-level cognitive activities, i.e., visual detection. Specifically, the automatic process of a visual target entering consciousness is mediated by verbal linguistic labels. Congruent linguistic cues facilitate the suppression of the target to break through the CFS (i.e., the flickering mask) and enter into awareness faster than those with incongruent linguistic cues, whereas incongruent linguistic cues function in the opposite way, i.e., they inhibit processing. To implement b-CFS in the examination of linguistic effects on motion event, participants should exhibit different patterns when processing motion events processing as in the current study, participants should exhibit different patterns when processing motion events with distinct proceeding cues; specifically, they should be faster to recognise the hidden target with congruent proceeding linguistic cues than those with incongruent linguistic cues. For instance, they should be faster to break through the suppression of a motion event (e.g., man walking into room) when the proceeding cue was congruent (e.g., man walking into room) rather than incongruent (e.g., woman jumping out of room).

In terms of examining language-specific effects on motion event cognition among monolingual Chinese and English speakers, and Chinese learners of English L2 speakers, b-CFS is sensitive to detect suppressed stimuli with distinct degrees of familiarity (Jiang *et al*., 2007). Consequently, it has the potential to detect distinctions in language-specific processing among speakers in these three language groups, i.e., they should take different amounts of time to unconsciously process the same motion events. More importantly, Chinese monolinguals, low, low, and intermediate L2 learners, should react similarly to break through the CFS mask suppressing stimuli with both highlighted *manner*-salient and *path*-salient targets due to the equal saliency of *manner* and *path* in the equipollently-framed Chinese (Slobin, 2006; Ji, 2017). On the contrary, English (satellite-framed) monolinguals, advanced L2 learners, should use different amounts of time to enter consciousness with *manner* and *path*-salient stimuli, respectively (Tamly, 2000; Ji, 2017; Wang & Li, 2021a, 2021b). To summarise, , this chapter has three main sections: one discussing the effects of language on thought cross-linguistically, and the other focusing the effects of language on thought in second language acquisition. In the first section, cross-linguistic evidence supporting the LRT was introduced first, followed by discussions of potential reasons of not observing language-specific effects, such as the characteristic of the linguistic domains, and approaches utilized in these studies. Moving forward to the focus of the present study: investigating the predictions of the LRH in motion events, key research in motion event processing was carefully reviewed. A research gap was identified, that is, the conflicting results between offline responses and online processing patterns (Ji, 2017; Wang & Li, 2021b; Soroli, 2024) observed in research focussing on motion lexicalisation might illustrate a more intriguing picture for understanding the specific processes induced by distinct pronounced motion expressions. Therefore, the current study focusses on *manner* and *path* encodings in Chinese and English. Through three carefully designed non-verbal experiments, the present study attempts to support the LRH by providing evidence observed at both low and high levels of cognitive processes. The second section started with a detailed introduction of studies examining the cognitive transfer in L2 learners. However, due to various reasons, mainly methodological, most L2 research concerns conceptual transfer in L2 learners when using the second language. Since the present study focusses on the LRH, only those attempts to investigate conceptual/ cognitive transfer in non-verbal experiments were reviewed, but with no restriction for studies which may have involved covert linguistic recruitment. Predictors associated with these specific processes of cognitive/ conceptual transfer were also introduced and discussed. Following the same rationale, specific processes of cognitive/ conceptual transfer, specifically in the domain of motion, were presented, as well as their associated predictors. Investigations of cognitive transfer in L2 learners at lower L2 proficiency levels is greatly lacking in the relevant research, and this research gap will be addressed in the current study. That is, the aim is to explore potential factors that mediate the cognitive transfer in L2 learners of both intermediate and advanced L2 proficiency in a low-level detection paradigm (b-CFS). Additionally, L2 learners were also examined in two verification experiments, with and without overt linguistic labels, in the hope that this will provide more evidence regarding any conceptual transfer that may occur when the language recruitment is not a concern (Berthele & Stocker, 2017; Bylund *et al*., 2013; Aveledo & Athanasopoulos, 2023). The third section presents a demonstration and brief discussion of studies examining motion event cognition specifically between Chinese and English monolinguals, as well as conceptual/ cognitive transfer in Chinese learners of L2 English. This is followed by an explanation of the experimental rationale and a discussion of employing the b-CFS in the current study. Details on experimental design are set out in the following Chapter 3.

Chapter 3: Methodology

Through an exploration of low- and high-level encoding patterns of *manner* and *path* among monolingual English and Chinese speakers, and Chinese learners of L2 English, the present study aims to investigate the effects of language on motion event cognition. Specifically, caused motion events are examined in three non-verbal experiments, including a low-level visual detection experiment: b-CFS, and two high-level visual processing experiments: video-video (without overt linguistic involvement) and sentence-video (with overt linguistic labels) verification experiments. This Chapter sets out the methodology of the current study, including a demonstration and brief discussion of studies investigating motion event cognition specifically between Chinese and English monolinguals, and cognitive transfer in Chinese learners of L2 English. Following this, general research questions of the current study are presented, followed by detailed methodology (e.g., rationale, detailed research questions, predictions, materials, participants, procedure) in each experiment conducted in the current study introduced in sequence: the b-CFS experiment, the sentence-video verification experiments, and the video-video verification experiment.

3.1 Research questions and hypotheses

There are three main research questions in the present study:

1. To what extent do monolingual Chinese and English speakers vary in the processing of *manner* and *path* components in motion events without overt or covert language involvement? Specifically, will monolingual English and Chinese speakers break through the suppression and detect the *manner* and *path-manipulated* information in different ways?
2. **2.** To what extent do monolingual Chinese and English speakers vary in the processing of *manner* and *path* components in motion events with (overt) covert language involvement? Specifically, will monolingual English and Chinese speakers spend different amounts of time identifying (sentence) video-video (in)consistency in *manner* and *path-manipulated* information?
3. **3.** To what extent do Chinese learners of L2 English experience cognitive transfer when processing motion events? Specifically, will they change from

L1- to L2-like cognitive patterns? Will this process be affected by *L2 proficiency*, *language context*/ or other extralinguistic factors, such as *cultural immersion*, *onset of L2 acquisition*, *language use in L2* and *length of stay in an L2-speaking country*?

There are two main hypotheses in the present study:

1. If attentional biases to specific motion components (*manner*/*path*) are predictable according to the speakers' native language, as predicted by the LRH (Wholf, 1956), monolingual Chinese and English speakers will vary in the processing of *manner* and *path* components, regardless of whether this process is conscious or unconscious. Such cross-linguistic differences will emerge consistently from an early unconscious stage (i.e., reflected in the b-CFS experiment) to a later conscious stage (i.e., reflected in the video-video and sentence-video verification experiments).

   Specifically, monolingual English speakers are expected to exhibit distinct patterns in breaking through the suppression for *manner* and *path* components, while Chinese monolinguals should demonstrate similar patterns in their responses to those two components. Consistent cross-linguistic patterns should also be observed in the two verification experiments, but with a stronger effect in the sentence-video verification experiment, attributable to the enhanced influence of linguistic labels (Lupyan, 2008).

   Monolingual English speakers should spend different amounts of time processing picture/video-video (mis)matches with highlighted *manner* information compared to those with highlighted *path* information. Detailed variation of processing the English *manner* between low-level detection (b-CFS) and high-level visual discrimination (sentence-video verification and sentence-video verification) would be expected to be different due to distinct levels of perceptual processing examined by the current study (Meteyard *et al*., 2007). In contrast, monolingual Chinese speakers should spend similar amounts of time verifying (mis)matches regardless of the highlight *manner/path* information.

2. If attentional biases to specific motion components (*manner*/*path*) are predictable according to the speakers' native language, Chinese learners of L2 English should experience consistent cognitive transfer across three experiments. Moreover, these specific processes are expected to be influenced by *L2 proficiency*, *language context*, and other extra linguistic predictors (Pavlenko, 2011). Specifically, L2 learners with high L2 proficiency should mirror the cognitive patterns exhibited by monolingual English speakers and process the *manner* and *path* components in distinct ways. Whereas L2 learners with low L2 proficiency should stick to L1-like patterns and process *manner* and *path* components in a similar way.

3.2 The b-CFS experiments

3.2.1 Pilot study

To test the feasibility of using b-CFS to examine the *manner* and *path-manipulated* information in monolingual Mandarin Chinese and English speakers, a pilot study was designed and conducted.

3.2.1.1 Predictions

According to the LRH, the Monolingual Chinese and English speakers are expected to process motion events differently in the b-CFS pilot study. Specifically, English speakers should use a different amount of time to process *manner*-manipulated information compared to *path* because the *manner* is more salient in English, whereas Chinese speakers would not resemble this pattern due to the equal salience of those two components in Chinese. Specifically, the difference between breaking through the suppression to identify the corner of the appearing stimuli with *manner*-manipulated information and those with *path*-manipulated information should be bigger in English speakers, when compared to Chinese speakers.

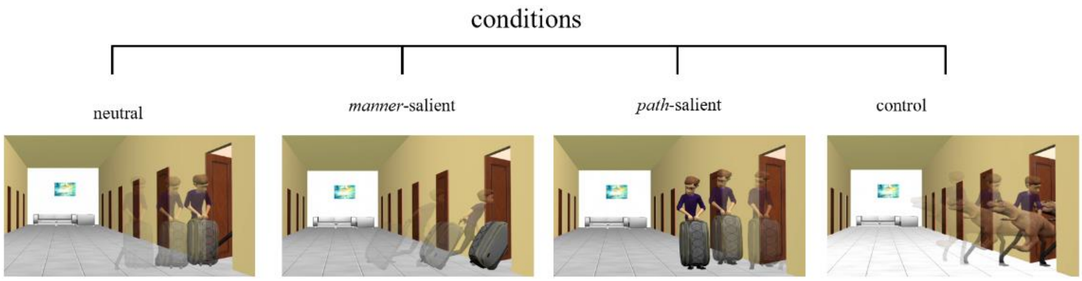
3.2.1.2 Participants

Forty-eight participants were recruited at the University of York, UK. 24 were native English speakers aged 18 to 22 (M = 19.25, SD = 0.99, female = 12), and the other 24 participants were native Mandarin Chinese speakers aged 18 to 33 (*M* = 21.96, *SD* = 2.99, female = 16). All participants were over 18 years of age and had normal or corrected-to-normal vision. The Mandarin Chinese speakers were Chinese students at the University of York who had arrived in the UK shortly before the experiment. They had knowledge of English as their L2 (IELTS scored 6 or below) and were self-assessed as Mandarin-dominant speakers. The English speakers were students at the University of York and self-assessed as English-dominant speakers.

3.2.1.3 Materials

A total of forty-eight critical stimuli (Appendix 2) illustrating caused motion events were used in the pilot study, which consisted of 12 different motion scenarios displayed by 3D animations with the same agent (e.g., a man) doing different actions. In Figure 3.1, each motion scenario had four conditions with identical information (e.g., background, object) except the manipulated component: *neutral* (e.g., '*A man is carrying a snowball up the hill*'), *manner-salient* (e.g., '*A man is rolling a snowball up the hill*'), *path-salient* (e.g., '*A man is carrying a snowball down the hill*') and *control* (e.g., '*A man is carrying a dinosaur up the hill*'), and these four conditions shared the same information in each specific motion scenario except the manipulated information. The *manner* and *path* saliency were manipulated based on five *manner* alternates: *pushing*, *pulling*, *rolling*, *carrying*, and *dragging*; and five pairs of *path* alternates: *towards*/ *away from*, *up*/ *down*, *into*/ *out of*, *along*/ *across*, *around*/ *across*. These alternates were designed based on Ji (2007) and Wang & Li (2021a). To avoid the repetition effect, each participant only viewed one condition per scenario. Therefore, each participant viewed a total of 36 trials in the pilot study, including 12 critical stimuli and 24 fillers that were not related to the critical ones (e.g., a woman is doing yoga). Four of the 24 fillers were viewed as practice trials. To be consistent with previous studies which also examined the linguistic-specific effects of euqipollently-framed Chinese and satellite-framed English on motion event cognition in Chinese and English speakers, the current study only used caused motion events as stimuli.

Figure 3.1 *Illustration of experimental conditions in the pilot study*


conditions
neutral
manner-salient
path-salient
control

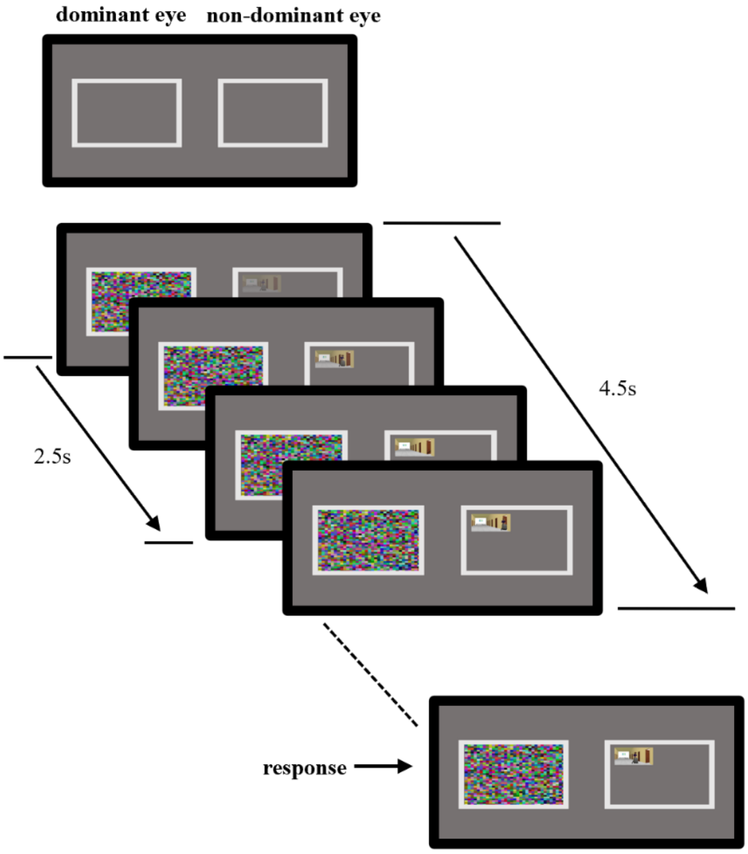
3.2.1.4 Procedure

The main equipment for the b-CFS pilot experiment was a set of mirror stereoscope goggles with four adjusted mirrors inside. The angle of the mirrors was adjusted according to the distance between the participants and the displayed stimuli (e.g., on a computer screen); thus, they viewed the mask and the stimuli with their dominant and non-dominant eye, respectively. The experiment was operationalised via Matlab (MathWorks, Natick, MA) using Psychotoolbox (Brainard, 1997).

Before participating in the experiment in a linguistic lab, participants were instructed to perform a self-check for their dominant eye via a video (AllAboutVisionVideo, 2018) on YouTube demonstrating how to find their dominant eyes. This aimed to familiarise the participants with the idea of dominant eyes, because the mask should be presented to the dominant eyes of the participants, thus maintaining the high efficiency of the suppression. Self-checking prior to the face-to-face experiment also saved time as it removed the need to explain the detailed rationale for identifying dominant eyes in the lab. To further check if the participants fully understood the procedure of checking their dominant eyes and found their dominant eye correctly, they were asked to demonstrate the process one more time before the experiment session. Following the experimental session, the participant completed a short language background questionnaire.

In the experimental session, participants sat in front of a mirror stereoscope and place their heads on a chin rest. The mirror stereoscope was held by a fixed tripod, and a Micro-soft laptop was placed behind it. The distance between the laptop screen and the mirror stereoscope was approximately 40 cm, and the distance between the participant's eyes and the mirror stereoscope was approximately 2 cm. These two distances were slightly different between participants, as their eyes and heights varied. The dominant eye (e.g., the right eye) was presented with a dynamic Mondrian-like mask with colourful noise (right side of the screen), which flickered around at 10Hz. The non-dominant eye (e.g., the left eye) was presented with an animation clip (4.5 seconds) placed in one of the four corners of the left side of the screen randomly. The target stimuli appeared to go from transparent to opaque with a contrast increase from 0% to 100% in 2.5 s (Jiang *et al*., 2007).

Figure 3.2 *Illustration of the b-CFS procedure in the pilot study*



The participants viewed 36 trials (4 practice trials and 32 real trials) in total, and it took around 10-20 min depending on individual differences. As shown in Figure 3.2, every trial started with two white rectangle frames showing on the screen, which aimed to help the participant slightly adjust the mirror stereoscope goggle if needed. To obtain this, they should see only one white rectangle frame on the screen through the goggles once they have been properly set up. Following a button press to proceed, the b-CFS mode began. In the b-CFS mode, a flickering mask was presented to the dominant eye (i.e., one side of the screen), and an animation appeared gradually appearing (reach full contrast in 2.5 seconds) to the other eye (i.e., at one of the four corners of the other half of the screen) simultaneously. The participant's task was to identify the location of the animation by pressing one of the four preassigned buttons on the laptop's keyboard indicating the four positions of the videos, respectively: top left ('J'), top right ('k'), bottom left (‘n’) and bottom right (‘m’). The stimuli were presented counterbalanced with a randomised order, and the reaction times used to indicate the location of the stimuli were analysed.

3.2.2 A control experiment

In a typical b-CFS experiment, the mask plays a vital role in suppressing the stimuli effectively, and a specific mask should be selected based on careful considerations of the target stimuli (Pournaghdali & Schwartz, 2020). Current studies of employing the b-CFS mainly used static stimuli (Paffen *et al*., 2021; Lupyan & Ward, 2013; Ostarek & Huettug, 2017), and in the limited studies using dynamic stimuli (Hong, 2015), the stimuli contain simple spatial information, such as, directions. Whereas the current study uses motion event animation as stimuli in the b-CFS, which contains more information, such as, spatial, action, etc., so it is vital to utilise this control experiment for validating the function to suppress in the current study (Moors *et al*., 2014).

Therefore, a small-scale control experiment using the same sets of stimuli and procedure was conducted separately. The aim of the control study is to examine the sufficiency of mask to suppress the presented stimuli, so there are two conditions in the control experiment: with mask and without mask. The prediction is that participants (regardless of the language-backgrounds) would use significantly longer time to detect the corner of the presented stimuli in the condition with a mask, compared to those without a mask.

Ten participants (Chinese monolinguals, age, M = 19.1, SD = 0.74), including 6 females and 4 males, were recruited in China for the control experiment. The stimulus and procedure were as identical as the b-CFS experiment in the main study. Half of the stimuli were presented to the participants with a suppression driven by the mask (i.e., regular design described in procedure in pilot study section) and the other half without a suppression driven by the mask (i.e., an additional mask was presented to the nondominant eyes of participants as well, thus abolished the suppression of the mask). The stimuli with and without mask suppression were distributed counterbalanced and played in a randomised order.

3.2.3 Main experiment

The feasibility of the b-CFS was confirmed in the pilot study. However, examining the *manner* and *path* encoding through b-CFS phase alone seems to create more difficulty in the interpretation of the results. For example, in the pilot study, the language-specific effects were observed by comparing the difference between the RT (*manner*-salient) – mean (RT neutral), and the RT (*path*-salient) – mean (RT neutral). Moreover, using b-CFS phase alone requires participants to retrieve motion information from long-term memory, which seems not highly reliable, given no language-specific effects were discovered in previous studies require to retrieve motion memory (Gennari *et al*., 2002; Papafragou *et al*., 2002). Thus, to maximally boost the language-specific effects for readily availability in motion, and reflecting the results of reaction times, a prime procedure with sequence images abstracted from each stimulus was added proceeding the b-CFS phase. The names of each condition were adjusted due to the added prime procedure as well, to demonstrate the relationship between the prime and target stimuli. For instance, *neutral* is changed into *full match* since the target video and the prime photo are identical, *manner*-salient is changed into *manner mismatch* since the target videos in this condition have different *manner* components compared to those in the *full match* condition, *path*-salient is changed into *path* mismatch based on the comparison between stimuli in this condition and those in the *full match* condition, which are different *path* components, and *control* is changed into *full march* as the stimuli in this condition have a different object, specifically: dinosaur, to probe an immediate evident comparison. More details with examples are explained in the later section: 3.2.2.3. Materials. In both pilot and the main experiment, the saliency of motion components, such as *manner* and *path*, refers to the ability to recognise difference manners or paths. For instance, *manner* is more salient than *path* for English speakers, as the two components carry different weights in linguistic expressions in satellite-framed English. In contrast, *manner* and *path* should be equally salient to Chinese speakers due to the use of both verbs to describe these components in Chinese.

Full match and *full mismatch* conditions are two control conditions, and they would also provide evidence regarding low-level visual detection if different results emerged in the other two verification experiments, which examined the higher-level processing of caused motion events with the same stimuli. As reflected in Ji (2017), both English and Chinese speakers tended to choose *path* alternates in the similarity judgment task, but the time taken to make these judgments differed between the two language groups. Specifically, language-specific differences might emerge at an early stage and then be obscured by the universal saliency of *path* in motion events unless they are boosted by overt linguistic labels.

3.2.3.1. Predictions

3.2.3.1.1 Task-specific predictions:

The English and Chinese speakers are expected to demonstrate largely comparable processing patterns between the two control conditions (*full match:* e.g., ‘a man is carrying a suitcase into a room’ and *full mismatch:* e.g., ‘a man is carrying a dinosaur into a room’) in the b-CFS experiment. This is because the language-irrelevant effect is expected to emerge at a later stage, subsequent to the manifestation of language-specific effects (Trueswell & Papafragou, 2010). However, b-CFS is sensitive to processing at an early stage, and thus should mainly capture any potential language-specific differences. Specifically, both English and Chinese speakers should not exhibit differences in RT between stimuli in the *full mismatch* and *full match* conditions.

3.2.3.1.2 Cross-linguistic predictions:

If language-specific low-level effects are indeed observed between monolingual Chinese and English speakers, monolingual Chinese and English speakers should process motion events differently in the b-CFS study. According to the predictive processing account (Lupyan & Clark, 2015), the construction of mental representations results from the dynamic interplay between predictions from higher-level cognitive processes and incoming low-level information; thus, b-CFS monitors how differences in the representation of motion across languages (e.g., English vs. Chinese) can influence the predictions used to evaluate visual stimuli. Specifically, English speakers should spend less time breaking through the suppression and detecting the corner of test stimuli with *manner-manipulated* information in relation to the prime image compared to those with *path-manipulated* information, as the *manner* is more salient in English. In contrast, Chinese speakers are expected not resemble this pattern due to the equal salience of *manner* and *path* in Chinese.

If Chinese learners of English L2 speakers experience a cognitive shift, they should mirror monolingual English speakers with the increase of their L2 proficiency (Athanasopoulos, 2007; Athanasopoulos & Kasai, 2008). Specially, lower-level L2 learners should stick to the processing pattern obtained in monolingual Chinese speakers, i.e., they should break through the suppression similarly and spend a similar amount of time to spot the locations of the stimuli in both *manner* and *path* manipulated conditions. Whereas higher-level L2 learners are more likely to mirror the monolingual English speakers, i.e., they should break through the suppression differently and spend less time to spot the locations of stimuli with *manner-manipulated* information than those with *path-manipulated* information.

In line with the findings on extralinguistic factors which were found to potentially affect the cognitive transfer in bilinguals (Pavlenko, 2011), the present study also included the *language context* in the design and controlled the *cultural immersion* for the L2 learners (Athanasopoulos, 2007). Other factors, such as the *onset of learning English as a second language (AO), the daily speaking/writing of L2 English* and *time spent in the UK* were also recorded in a language background information questionnaire. L2 proficiency is the primary condition for having flexible cognitive switch between L1 and L2-like patterns. Cultural immersion, language use, and time spent in the L2-speaking country are factors related to L2 proficiency, which might influence our results in the b-CFS experiment. However, AO and language context are unlikely to be reflected in the b-CFS experiment unless our L2 learners have balanced levels of L1 and L2 proficiency (Bylund & Athanasopoulos, 2014).

Specifically, if *language context* affects the low-level visual detection of *manner* and *path-manipulated* stimuli in the L2 learners, they should switch between L1 and L2 patterns based on the tested languages, that is, the language used to introduce the study and the instructions before starting the b-CFS experiment (Sachs & Cloy, 2006). Specifically, L2 learners tested in English should tend to mirror monolingual English speakers, whereas those who tested in Chinese should follow the pattern of monolingual

Chinese speakers in the b-CFS experiment. However, this is highly unlikely, since b-CFS examines the low-level processing which should demonstrate a stable processing pattern (Jiang *et al*., 2007), which is not mediated by immediate language use.

If the *age of onset* (AO) of L2 acquisition affects the low-level visual detection of *manner*- and *path-manipulated* stimuli in the L2 learners, they should exhibit distinct processing patterns depending on the AO. Specifically, late (e.g., AO > 5, Wang & Li, 2011b) L2 learners have been found to be more likely to be influenced by immediate language use, whereas early L2 learners should stick to the L2-like pattern regardless of the tested languages once their proficiency reaches advanced or near-native level, otherwise they are very likely to pattern with L1 native speakers regardless of the tested languages due to low L2 proficiency (Pavlenko, 2011). Specifically, advanced L2 learners should spend less time breaking through the suppression and detecting the corner of test stimuli with *manner-manipulated* information in relation to the prime image compared to those with *path-manipulated* information, as they mirror the English learners, thus, acquiring a relatively more salient *manner*. In contrast, lower level L2 learners are expected not resemble this pattern observed in monolingual Chinese group, which exhibit similar amount of time to process stimuli with *manner* and *path* manipulation due to the equal salience of *manner* and *path* in Chinese.

If *cultural immersion* affects low-level visual detection of *manner* and *path-manipulated* stimuli in L2 learners, distinct patterns should be observed between L2 learners (with equivalent L2 proficiency) resident in York and China (Cook *et al*., 2006). Specifically, L2 learners who lived in England should tend to mirror monolingual English speakers, whereas those who lived in China should follow the pattern of monolingual Chinese speakers in the b-CFS experiment.

If *language use* (long-term language use, cf. *language context*) affects the low-level visual detection of *manner* and *path-manipulated* stimuli in L2 learners, they should shift from sticking to the L1-like pattern to mirroring the L2-like pattern with increasing daily use (speaking and writing) of L2 English (Athanasopoulos *et al*., 2011).

3.2.3.1.3 Lateralisation

The effects of lateralisation were not designed to be examined intentionally, but this was achieved since the suppression mask must be presented to the dominant eye, and the other (nondominant) eye sees the stimuli in the b-CFS experiment. This is consistent with a typical design in lateralisation. As discussed in Chapter 2 the left hemifield (LH) of the brain is primarily responsible for language processing, thus cross-linguistic differences observed in the right visual field (RVF) rather than the left visual field (LVF) are regarded as influences driven by language (Gilbert *et al*., 2006). Thus, if the lateralisation affects the low-level processing of motion events between Chinese and English, the language-specific patterns observed in the b-CFS should emerge when the mask was placed on the left eye.

3.2.3.2 Participants

One hundred and ninety-four participants were recruited in the main experiment. All participants were over 18 years of age and had normal or correct-to-normal vision. One hundred and seven participants were monolingual English and Chinese speakers. Among monolingual speakers, 55 (age, *M* = 19, *SD* = 1) were native Chinese speakers (female: *N* = 21) recruited from three universities in the city of Zhengzhou. Although English is compulsory for students in China, these participants were regarded as monolinguals, as they had relatively low score on the national English tests. There were 52 (*M* = 20 years old, *SD* = 2.02) English native speakers (female: *N* = 28) recruited in York, who claimed to be monolingual in the self-reported questionnaire. Fifty-one (*M* = 23.67 years old, *SD* = 2.56) Chinese learners of English L2 (female: *N* = 40) with IELTS scores of 6 or higher were recruited at the University of York. Another 36 (female: *N* = 31) Chinese learners of L2 English (*M* = 19.03 years old, *SD* = 1.10) were recruited in Zhengzhou, China. These participants were Mandarin Chinese speakers who had never been to an English-speaking country before. However, as English major students, their English proficiency level was relatively higher in comparison to average Chinese students in other majors.

3.2.3.3 Materials

The materials used in the b-CFS stage were identical to those used in the pilot study, which were designed based on the caused motion events examined in Ji (2017), Wang & Li (2021a). However, to boost the language-specific differences, prime images (abstracted from the 12 critical motion scenarios and 24 fillers) were introduced proceeding the b-CFS stage. As shown in Figure 3.3, each prime image highlighted the actions of the agent by integrating six abstracted frames from the motion animation, thus illustrating a complete process of the motion scenario. The temporal information along the trajectory was reflecting by the contrast of the man in the image, with lowest contrast indicates the start of the animation and highest contrast refers to the end of the animation.

Figure 3.3 *An example of the prime image*

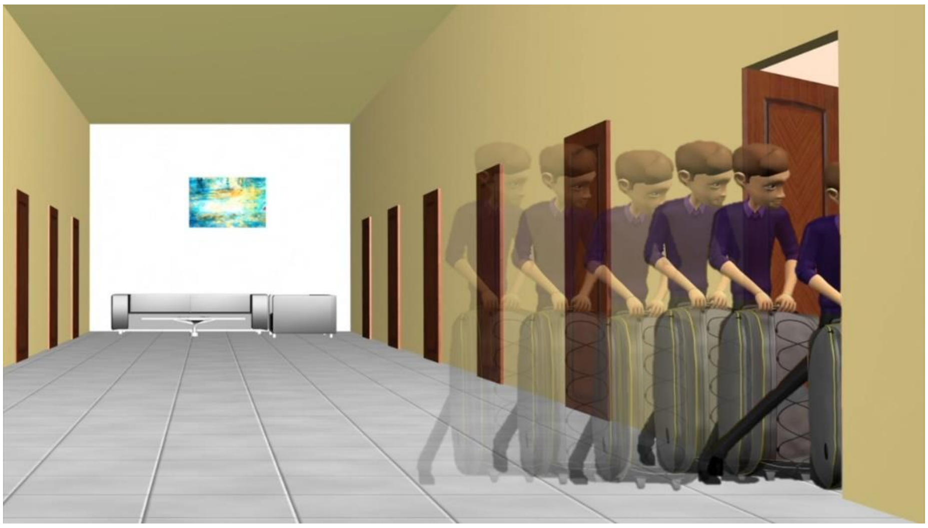


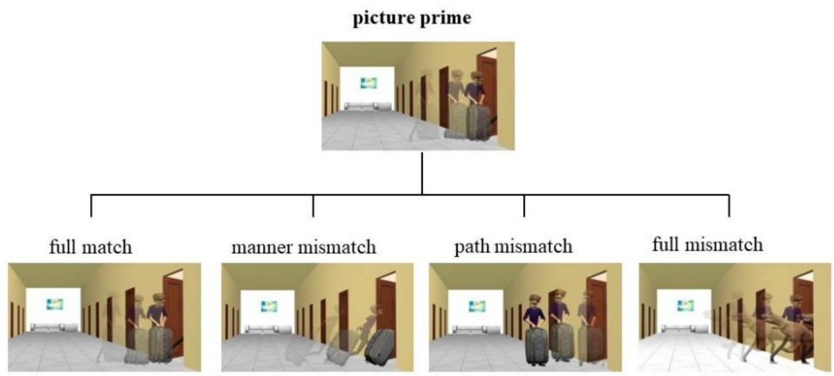
Table 3.1 *Critical items with manner and path alternates*

Items	
1	A man is pushing (rolling) a tyre along (across) the street.
2	A man is rolling (carrying) a log toward (away from) the cabin.
3	A man is rolling (pushing) a haystack around (across) a puddle.
4	A man is carrying (rolling) a snowball up (down) the hill.
5	A man is pushing (carrying) a giftbox into (out of) a truck.
6	A man is carrying (carrying) a wheelchair towards (away from) a house.
7	A man is carrying (pulling) a suitcase into (out of) a room.
8	A man is pulling (carrying) a skateboard away from (toward) the skate park.
9	A man is dragging (carrying) a sack out of (into) a tunnel.

10	A man is dragging (pulling) a cart across (around) a garden.
11	A man is carrying (dragging) a toy bear across (around) a playground.
12	A man is pulling (dragging) a toy car along (across) a river.

A total of 48 critical event animations (4.5 seconds long per video) and 12 critical prime images were included in the experiment. There were four conditions in the target animations in relation to the prime images. As shown in Figure 3.4, an identical match between prime and target: *full match* condition (e.g., A man is **carrying** a suitcase **into** a room), a critical mismatch between prime and target: *manner mismatch* condition (e.g., A man is *pulling* a suitcase into a room), the other critical mismatch between prime and target: *path mismatch* condition (e.g., A man is carrying a suitcase *out of* a room) and a full mismatch between prime and target: *full mismatch* condition (e.g., A man is carrying a *dinosaur* into a room). The *manner* words were *pushing*, *pulling*, *rolling*, *carrying*, and *dragging*, and the *path* words were *along/across*, *towards/away from*, *around/across*, *up/down*, *into/out of*. The list of critical items is given in Table 3.1. The full match and the full mismatch were two control conditions that illustrate the baselines of the match and mismatch between the prime and target, which are not driven by language-specific differences.

Figure 3.4 *Conditions in the b-CFS experiment*



Note: The picture primes were sequential photos generated on the videos in the full match condition. The four conditions after the picture prime were 4.5s animated videos. The match and mismatch were a comparison between the picture primes and the videos.

The participants viewed 36 trials in total. Each trial contained one prime-target pair. These 36 trials contained 12 critical trials and 24 filler trials. Each prime-target pair had four alternates for the target animations and, to avoid the repetition effect, each participant viewed one alternate target animation for each prime. Filler trials presented simple motion scenes which were not relevant to the critical stimuli, e.g., *Two men are hiking*. The materials and locations of the target stimuli were counterbalanced, and the order was randomised.

Participants completed a language background questionnaire (modified from Zhang & Vanek, 2021) after the experiment. The questionnaires for monolingual Chinese and English speakers and Chinese learners of L2 English speakers were in Appendix 1. Additionally, Chinese learners of L2 English completed an Oxford placement test, Grammar 1. Details of the test are given in Appendix 3.

3.2.3.4 Equipment

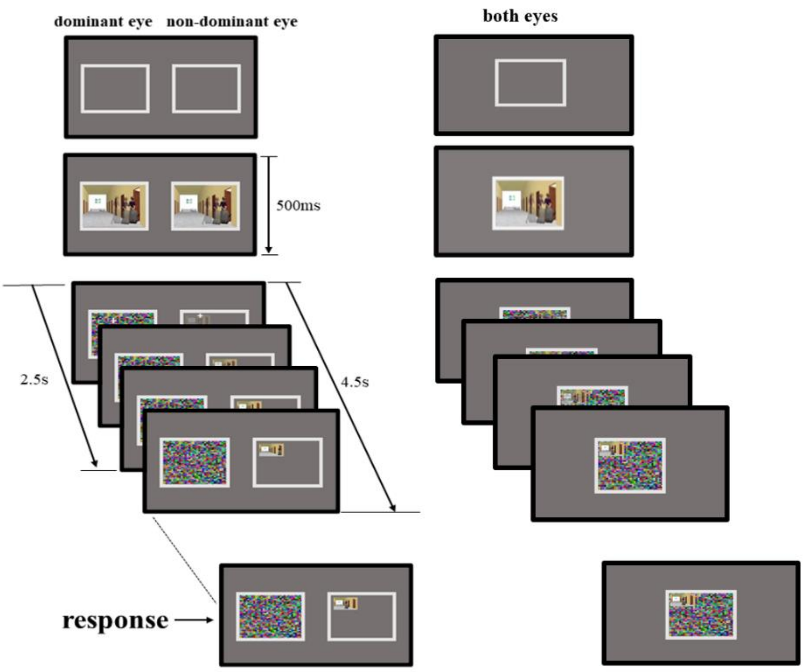
The equipment and set-up were identical to those used in the pilot study.

3.2.3.5 Procedures

The whole procedure was similar to that of the pilot study, but with an additional prime stage between the fixation stage (reflected by two white frames) and the b-CFS stage: i.e., a 500-ms prime image was presented to the participants. As shown in Figure 3.5, the schematic procedure illustrated on the left side shows what was shown on the laptop screen, and the other on the right side displays the images the participants saw through the mirror stereoscope goggles. To start by viewing two white rectangles on the screen, participants checked the suitability of the set-up by confirming that they saw one white frame through the goggles. To proceed to the next step, participants were instructed to press a button (any button) on the keyboard. A prime (including critical and filler) image was presented to the participants' both eyes for 500 milliseconds, and the b-CFS procedure was immediately followed, which was identical to that used in the pilot study.

Another difference between the main study and the pilot study was that to further confirm the reliability of the mask, participants were also required to provide verbal feedback after the experiment. For example, they were asked ‘was the area inside the rectangle fully covered by the colourful mask?’ The answer should be ‘yes’. The rationale for this addition to the procedure is that when the mirror stereoscope worked, the mask, rectangle, and stimulus should overlap and merge into one image as shown on the right side in Figure 3.5. Specifically, the mask should cover the entire area inside the rectangle, and the stimulus should be on top (i.e., break through) of the mask when the participant was able to see the stimulus. All of this should stay inside the rectangle frame. Finally, the participants completed a questionnaire to check their language background.

Figure 3.5 *Illustration of the b-CFS procedure in the main study*



Finally, two additional procedures were included for the Chinese learners of L2 English speakers: a test to confirm their L2 English proficiency (i.e., Oxford placement test, Grammar 1); and, in order to examine the effects of language context, they were randomly assigned to two subgroups with different languages tested languages (e.g., half of the L2 learners received instructions in Chinese only, and the other half L2 learners received instructions in English) before and during the experiment, that is, monolingual mode: the experimenter contacted these L2 learners in Chinese and stuck to Chinese during introducing the experiment and assisting the adjustment of the set up; bilingual mode: the experimenter contacted these L2 learners in Chinese before they arrived in the lab, and spoke English to them during introducing the experiment and assisting the adjustment of the set-up. The b-CFS experiment did not require any overt language use, unless the participants had any questions.

3.3 Verification experiments

The LRH predicts that the effects of language on thought should be stable across different cognitive experiments (Casasanto, 2006); thus, two additional verification experiments (video-video verification and sentence-video verification) will be introduced to measure the higher-level processing of motion events in monolingual Chinese and English, and Chinese learners of English speakers without and with overt linguistic labels, respectively. Both verification experiments are modified based on Zwaan *et al*. (2002), which uses sentence/video as a memory retrieval cue for a later recognition of the target motion animation.

The rationale of the current verification experiments lies in the situation model theory (Zwaan & Radvansky, 1998), which claims that there is a dynamic interaction between long-term working memory (LTWM) and short-term working memory (STWM): "Readers keep the integrated situation model in LTWM while the current model is constructed in STWM. During the construction process, there is transient activation in STWM to retrieve cues for parts of the integrated model. Update occurs by forming links between the current model and the retrieved elements of the integrated model. At this point, the current model has been integrated, and the integrated model has been updated, so that a new current model can be constructed in STWM. This process continues until the complete model is stored in long-term memory" (p.166). Consequently, (mis)match between a sentence (video) and a subsequent video exerted distinct immediate attention to the participants for *updating* their integrated model. For example, viewing a sentence/video prime like: *man carrying suitcase into room* would create a current model of a motion scenario, and viewing a subsequent target video contained different information (e.g., man carrying *dinosaur* into room) would 'force' participants to update the current model by replacing the different information.

In terms of the difference between the stimuli in the *manner mismatch* condition and the stimuli in the *path mismatch* condition, monolingual Chinese and English speakers should react differently in the verification experiments due to the distinct motion typologies (Tamly, 2000; Slobin, 2004) in Chinese and English. Specifically, Chinese speakers should use similar amount of time to verify target video with a different *manner* (e.g., man pulling suitcase into room) or *path* (e.g., man carrying suitcase out of room) since these two components are equally salient in Chinese, whereas English speakers should spend different amounts of time to verify the same stimuli with alterations in these two components due to the asymmetric saliency between them in English. That is, when the sentence/ video prime activated the language-specific *manner* and *path* stored in the LTWM, to update the current model, verifying a target video with different a *manner* or *path* results in different patterns between Chinese and English speakers. More detailed language-specific predictions are presented in the following.

3.3.1 Self-paced Video-Video (VV) Verification Experiment

To further examine the LRH, the current verification experiment aims to investigate the effects of language on event cognition among monolingual Chinese and English speakers, and Chinese learners of L2 English from a higher-level (i.e. categorical) perspective, a video-video verification (VV) experiment was conducted. In the VV verification experiment, identical stimuli (videos) were utilised except that sequence photos were replaced by videos as memory retrieval cues (Zwaan & Radvansky, 1998) for the later verification stage.

3.3.1.1 Predictions

3.3.1.1.1 Task-specific predictions

In contrast to the task-specific predictions proposed in the b-CFS experiment, if high-level processing is observed in the current VV verification experiment, participants should exhibit similar processing patterns in the two control conditions (*full match* and *full mismatch*) across all three language groups (Zwaan *et al*., 2002). Specifically, participants should be more sensitive in verifying video-video pairs in the *full mismatch* than those in the *full match* conditions. This concept is grounded in previous research on the impact of categories on visual processing, which discovers that the shortest reaction times are expected when the stimulus belongs to a distinct category (Lupyan *et al*., 2010). Additionally, such differences should be more prominent than those in the critical conditions (*manner* and *path mismatch* conditions) due to the absence of overt linguistic labels (Trueswell & Papafragou, 2010).

3.3.1.1.2 Cross-linguistic predictions

If high-level language-specific effects are indeed observed between monolingual Chinese and English speakers, they would process motion events differently in the video-video verification experiment. Following the predictive processing account proposed by Lupyan and Clark (2015), the formation of mental representations is shaped by the continuous interaction between anticipations that arise from higher-level cognitive functions and visual input. Therefore, the VV verification experiment aims to demonstrate how distinct aspects of motion events, highlighted or understated in language, influence the development of expectations when promptly identifying (mis)matches between videos and videos. Specifically, according to Ji (2017), English speakers should spend more time verifying video-video pairs with *manner* than *path* alterations, since if higher-level linguistic categories affect visual perception (Lupyan *et al*., 2010), the shortest reaction times are expected when the stimulus belongs to a distinct category. In contrast, the Chinese speakers should not resemble this pattern due to the equal salience of *manner* and *path* in Chinese.

If Chinese learners of L2 English experience a conceptual shift, they should mirror monolingual English speakers with increasing in line with L2 proficiency (Athanasopoulos, 2007; Athanasopoulos & Kasai, 2008). Specifically, the L2 learners with lower L2 proficiency should stick to the processing pattern in monolingual Chinese speakers, i.e., they should spend a similar amount of time to verify video-video pairs with stimuli in both *manner* and *path* manipulated conditions. Whereas those with higher L2 proficiency are expected to mirror the monolingual English speakers, i.e., they should spend more time to verify stimuli with *manner* manipulated information than those with *path* manipulated information.

Consistent with the evidence regarding the influence of extralinguistic factors which have been found to potentially affect the conceptual transfer in bilinguals (Pavlenko, 2011), the present VV experiment also included *language context* in the design (Athanasopoulos, 2007). Other factors, such as the *age of learning English as a second language (AO), the daily speaking/writing of L2 English* and *time spent in the UK* were also recorded in the language background information questionnaire. L2 proficiency is the primary condition for showing a flexible conceptual switch between L1 and L2-like patterns. Language context, language use, and time spent in the L2-speaking country are factors related to L2 proficiency, which might influence our results in the VV verification experiment. However, AO is unlikely to be reflected in the VV verification experiment unless our L2 learners have balanced levels of L1 and L2 proficiency (Bylund & Athanasopoulos, 2014).

Specifically, if *language context* affects the low-level visual detection of *manner* and *path-manipulated* stimuli in L2 learners, they should switch between L1 and L2 patterns based on the tested languages, that is, the language used to introduce the study and the instructions before starting the VV verification experiment (Sachs & Cloy, 2006). That is, the L2 learners tested in English should tend to mirror monolingual English speakers, whereas those who are tested in Chinese should follow the pattern of monolingual Chinese speakers.

If the *age of onset* (AO) of L2 acquisition affects the low-level visual detection of *manner* and *path-manipulated* stimuli in L2 learners, they should exhibit distinct processing patterns depending on the AO. Specifically, late (e.g., AO > 5, Wang & Li, 2011b) L2 learners were more likely to be influenced by the immediate language use, whereas early L2 learners should stick to the L2-like pattern regardless of the tested languages once their proficiency reaches advanced or near-native level, otherwise they are very likely to pattern with L1 native speakers regardless of the tested languages due to low L2 proficiency (Pavlenko, 2011).

If *language use* affects high-level visual processing of *manner* and *path* in L2 learners, the L2 learners should shift from sticking to the L1-like pattern to mirroring the patterns exhibited by the L2 native speakers with the increase of frequency of the use (speaking and writing) of L2 English in daily life.

3.3.1.2 Participants

A different group of 150 participants were recruited in the VV verification experiment. There were 51 (age: *mean* = 19.49, *SD* = 1.38; female = 37) Chinese monolingual speakers were recruited in Zhengzhou, China, and 50 (age: *mean* = 21.80, *SD* = 2.50; female = 23) English monolingual speakers recruited in York, UK. The rest of the 51 participants (age: *mean* = 24.31, *SD* = 2.75; female = 41) were Chinese learners of English L2, who were also recruited in York.

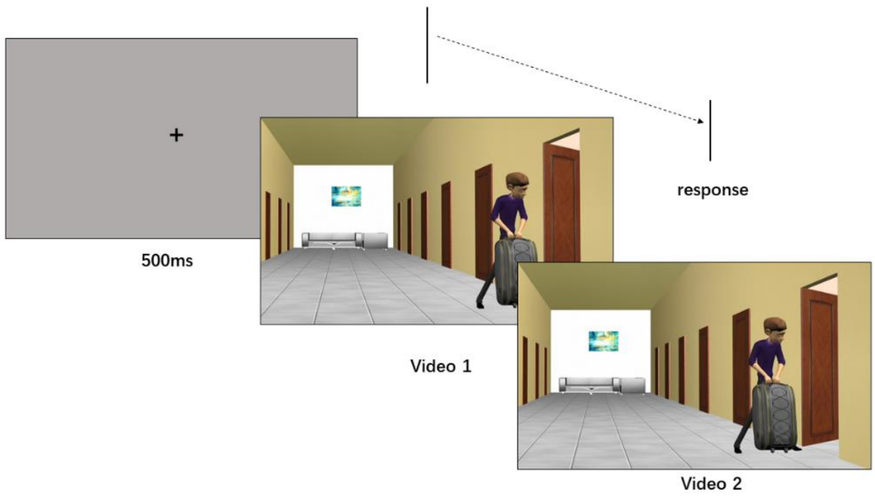
3.3.1.3 Materials

The materials used in this VV verification experiment were all motion event videos, which were identical (both critical stimuli and fillers) to those used in the b-CFS experiments. There were 12 motion scenes (cues) paired with four conditions reflecting in specific alterations (target videos, e.g., *A man is carrying a suitcase into a room.*), namely, *full match* (identical to the proceeding video cues, e.g., *A man is carrying a suitcase into a room.*), *manner mismatch* (*manner* alterations, e.g., *A man is pulling a suitcase into a room.*), *path mismatch* (*path* alterations, e.g., *A man is carrying a suitcase out of a room.*), and *full mismatch* (object alterations, e.g., *A man is carrying a dinosaur into a room.*). A total of 36 trials were presented to each participant, including 12 critical items and 24 fillers. All the fillers were also cartoon videos, for example, 'two *men are hiking*'. The materials were presented counterbalanced in random orders. The language background questionnaires (Appendix 1) and the OPT (the quick Oxford Placement Test, 2001) test used in this experiment were identical to those in the b-CFS experiments. As in the b-CFS, the VV verification experiment was also operationalised via Matlab (MathWorks, Natick, MA) using Psychotoolbox (Brainard, 1997).

3.3.1.4 Procedure

In Figure 3.6, the VV verification experiment consisted of two primary stages. Initially, participants received centrally displayed video cues (video 1) on a laptop screen following a brief fixation period (500 ms) period. These video cues depicted scenarios involving motion events and persisted on the computer screen until participants signalled their comprehension by pressing any key to proceed. Subsequently, the participants entered the second stage where they viewed a target video (video 2) which either aligned (i.e., *full match condition*) or contradicted (i.e., *full mismatch*, *manner mismatch*, and *path mismatch* conditions) the preceding video. The aim of the task was to determine whether the target video matched the proceeding video by pressing the left arrow (‘YES’) or right arrow (“NO”) key on the keyboard. The experimenter also clarified the mismatches between the two videos would be related to actions. Both videos looped continuously until participants made their decisions. The presentation of stimuli followed a counterbalanced randomised order. After the experiment, the participants completed a language background questionnaire, while Chinese English learners additionally undertook the same OPT test. Furthermore, in the VV verification experiment, the monolingual Chinese and English speakers received instructions in Chinese and English, respectively, while L2 learners were divided into two subgroups based on the tested languages: a monolingual group and a bilingual group. The researcher used Chinese to guide the participants in the monolingual group and English to guide the participants in the bilingual group.

Figure 3.6 *Illustration of the VV verification experiment*



3.3.2 Self-paced Sentence-Video (SV) verification experiment

To continue to explore the language-specific differences in *manner* and *path* encoding in non-verbal experiments, but with cover linguistic involvement, the current verification experiment measures the verification patterns of same/different motion components between prime and target motion stimuli among monolingual English and Chinese, and L2 (native Chinese) English speakers. In the current verification experiment, identical procedures, and stimuli (videos) were used, except that the videos were replaced by written sentences as memory retrieval cues (Zwaan & Radvansky, 1998) for the later verification task. The general predictions are in line with those in the earlier video-video verification experiment; however, specific differences between control and critical conditions would be anticipated due to the presence of overt linguistic labels. Specifically, language-specific effects should be more prominent compared to those observed in the sentence-video verification experiment (Lupyan, 2008). Detailed predictions, research designs are set out below.

3.3.2.1 Predictions

3.3.2.1.1 Task-specific predictions

If high-level processing was observed in the current SV verification experiment, participants should exhibit similar processing patterns in the two control conditions (*full match* and *full mismatch*) across all three language groups (Zwaan *et al*., 2002). Specifically, participants should be more sensitive to verify sentence-video pairs in the *full mismatch* than those in the *full match* conditions. Additionally, such differences should be as comparable as those in the critical conditions (*manner* and *path mismatch* conditions) due to the presence of overt linguistic labels (Lupyan, 2008).

3.3.2.1.2 Cross-linguistic effects

If high-level language-specific effects are indeed observed between monolingual Chinese and English speakers, they would process motion events differently in the sentence-video verification experiment. According to the predictive processing proposed by Lupyan and Clark (2015), mental representations are formed and shaped by the continuous interaction between anticipations arising from higher-level cognitive functions and visual input. Hence, the SV experiment aims to show how various elements of motion events, emphasised or de-emphasised in language, impact the formation of expectations when swiftly discerning (in)consistencies between sentences and videos. Specifically, as noted by Ji (2017), English speakers tend to spend more time verifying sentence-video pairs featuring alterations in *manner* rather than *path*, consistent with the influence of categorical effects on visual perception (Lupyan *et al*., 2010), wherein the shortest response times are anticipated for stimuli belonging to distinct categories. In contrast, Chinese speakers are not expected to in this way pattern because of the equal prominence of *manner* and *path* components in the Chinese language.

If Chinese learners of L2 English have experienced a conceptual shift, they should resemble those of monolingual English speakers as their L2 proficiency increases (Athanasopoulos, 2007; Athanasopoulos & Kasai, 2008). Specifically, lower-level L2 learners are expected to exhibit processing patterns akin to monolingual Chinese speakers, spending similar amounts of time verifying sentence-video pairs across both *manner* and *path* manipulated conditions. In contrast, higher-level L2 learners should mirror the behaviour of monolingual English speakers, investing more time in verifying stimuli with *manner* manipulated information compared to those with *path* manipulated information.

According to the findings on extralinguistic factors which were discovered to potentially affect conceptual transfer in bilinguals (Pavlenko, 2011), the present SV experiment also included the *language context* in the design (Athanasopoulos, 2007). Other factors, such as the *onset of learning English as a second language, daily speaking/writing of English, and the length of time spent in the UK* were also recorded in a language background information questionnaire.

Specifically, if *language context* affects the high-level visual processing of *manner* and *path* in L2 learners, this effect should only be observed in late (e.g., AO > 5, Wang & Li, 2011b) higher level L2 learners. These learners should switch between L1 and L2 patterns based on the tested languages, that is, the sentences in the sentence-video pairs. Lower-level L2 learners should stick to an L1-like pattern regardless of the language contexts. Higher level L2 learners should pattern with the L2 native speakers regardless of the tested languages. If *language use* affects high-level visual processing of *manner* and *path* in L2 learners, L2 learners should shift from sticking to the L1-like pattern to mirror the L2-like pattern with the increase of frequency of using (speaking and writing) L2 English in daily life.

3.3.2.2 Participants

A different set of 154 participants were recruited in the self-paced sentence-video verification experiment (SV). All participants were 18 or over 18 years old and had normal or corrected-to-normal vision. Fifty-one monolingual Mandarin Chinese speakers (age, M = 21.30, SD = 2.90) were recruited in Zhengzhou, China and 28 of them were female. Another 51 monolingual English speakers (age, M = 19.49, SD = 1.23) were recruited in York, UK, and 34 of them were female. Fifty-two participants were Chinese learners of L2 English (age, M = 24.31, SD = 2.75) recruited in York, England, and 38 of them were female.

3.3.2.3 Materials

The materials (including both critical items and fillers) used in the SV experiment were identical to those employed in the VV verification and b-CFS experiment. However, instead of using written sentences, the current experiment uses videos in the memory-retrieval stage. The aim of the SV experiment was to heighten the involvement of linguistic labels in a high-level non-verbal setting and assess its impact on a verification task. In this experiment, motion videos were substituted with written sentences (e.g., man carrying suitcase into room) displayed on the laptop screen. The prime sentences served the purpose of aiding participants in constructing mental situation models (Zwaan & Radvansky, 1998), which were utilised for comparison with the motion event depicted in the subsequent video. Thus, the determination of *match* (i.e., *full match* condition, e.g., man carrying suitcase into room) and *mismatch* (i.e., *manner mismatch* condition, e.g., man pulling suitcase into room; *path mismatch* condition, e.g., man carrying suitcase out of room; *full mismatch* condition, e.g., man carrying dinosaur into room) conditions was based on the correspondence between the sentences and the videos. Both English and Chinese sentences utilised the present continuous tense, with Chinese sentences exclusively featuring the *BA* structure (further rationale provided in the first part of the current Chapter), such as, *yi2ge4 nan2ren2 ba3 xiang1zi0 ban1jin4 wu1zi0*, ‘a man is carrying a suitcase into a room’. A detailed list of sentences can be found in Appendix 4.

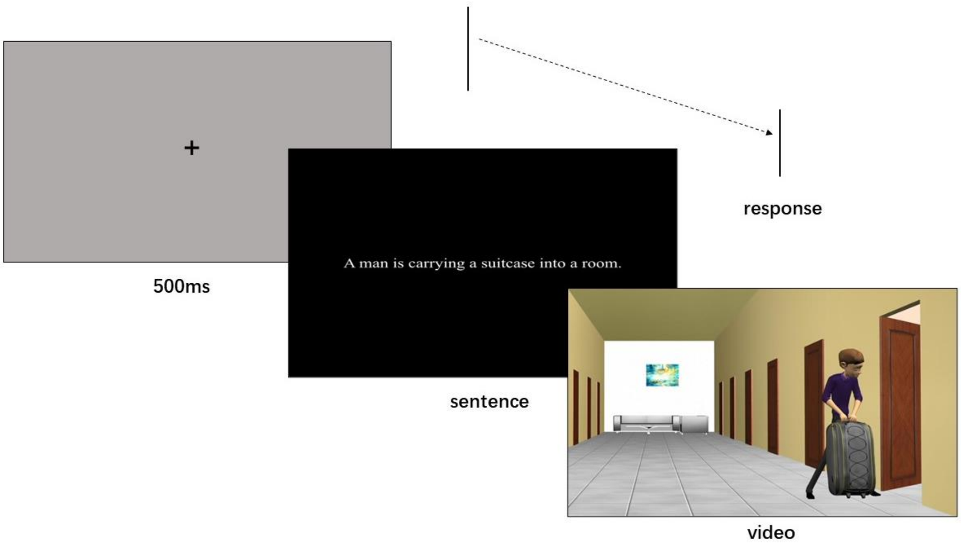
To establish distinct language contexts, L2 learners of English were divided into two subgroups for the SV experiment: the bilingual group and the monolingual group. In the bilingual group, the participants were presented with an equal distribution of Chinese and English sentences, with all 12 critical items presented in English. Additionally, six more filler sentences were in English, while the remaining 18 filler sentences were in Chinese. In contrast, in the monolingual group, all sentences were displayed in Chinese.

In line with the design of the VV verification and b-CFS experiment, each participant viewed 36 trials in total, comprising 12 critical pairs and 24 filler pairs. Each sentence-video pair consisted of one sentence and one subsequent video (i.e., animation clip), and they were exposed to only one condition within each motion scene to avoid the repetition effect. The materials were presented counterbalanced with a randomized order. Furthermore, the language background questionnaires presented to the participants after the experiment were identical to those used in the b-CFS and VV verification experiments.

The SV experiment was conducted in Matlab (MathWorks, Natick, MA) using the Psychophysics toolbox (Brainard, 1997) to collect face-to-face data from the monolingual Chinese speakers. The Gorilla Experiment Builder (www.gorilla.sc) (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2018) was used to create and host the SV experiment to collect online data from or the English speakers and Chinese learners of L2 English speakers.

3.3.2.4 Procedure

Figure 3.7 *Illustration of the sentence-video verification procedure*



The procedure in the SV verification experiment was identical to that in the VV verification experiment. Specifically, as shown in Figure 3.7, there were two main stages in the SV verification experiment. Participants viewed one sentence presented on the centre of the laptop screen shortly after a brief fixation (500 ms). This sentence depicted a scenario involving a motion event and remained on the computer screen until participants pressed any key to proceed, signifying their complete understanding of the sentence. Subsequently, the participants viewed a dynamic video (animation) in the second stage, which either corresponded to or conflicted with the preceding sentence.

The task was to determine whether the video matched the proceeding sentence by pressing the left arrow (‘YES’) or right arrow (“NO”) key on the keyboard. The video was continuously looped until a decision was reached. The stimuli were presented counterbalanced with a randomized order. Following the experiment, participants completed a language background questionnaire, and a Chinese learners of English completed an additional same OPT test.

To sum up, the current Chapter set out the rationale and specific methodology of each experiment undertaken in the present study. The research questions, hypotheses, predictions, participants, and procedures were also laid out. In the subsequent Chapter, the results of each experiment are presented in detailed and briefly discussed.

Chapter 4: Results

The current Chapter sets out the results of the experiments undertaken in the present study. This includes results from the pilot study of the b-CFS experiment, control experiment, main study of the b-CFS experiment, video-video verification experiment, and sentence-video verification experiment. Brief discussions are also presented after illustrating each specific finding.

4.1 The b-CFS pilot experiment

To minimise any potential effects of stimulus features not related to *manner* or *path* on visual processing differences, reaction times (RTs) were compared solely within each motion scenario. In this context, *manner* saliency refers to the RT difference calculated by subtracting the average RT for the motion-neutral animation from the raw RT for the corresponding *manner-salient* animation. Similarly, *path* saliency denotes the RT difference obtained by subtracting the average RT for the motion-neutral clip from the raw RT for the corresponding *path-salient* clip. To examine the effects of language groups, stimuli conditions and their interactions, linear mixed-effect regression models (LMMs) with main fixed effects (Language Group and Condition) and random effects (Participant and Item) were built and converged by using the lme4 packages (Baayen *et al*. 2008) in the R software.

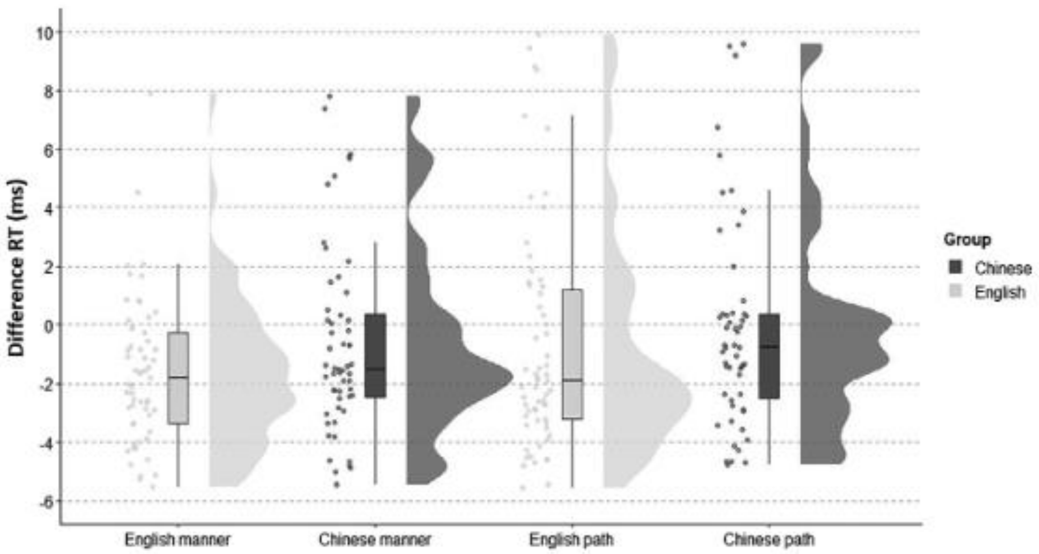
4.1.1 Results

Generally, the results were in line with the predictions. Both monolingual English (*M* = 95.59%, *SD* = 4.83) and Chinese (*M* = 94.06%, *SD* = 6.99) obtained a high accuracy rate in identifying the correct corner where the stimuli appeared. English speakers were faster in the *manner-salient* condition (*M* = 5.03s, *SD* = 3.33) compared to the *path-salient* condition (*M* = 5.63s, *SD* = 4.35). In contrast, Chinese speakers were slower under the *manner-salient* condition (*M* = 7.41s, *SD* = 5.85), compared to the *path-salient* condition (*M* = 6.67s, *SD* = 4.11).

The statistical analysis was based on the residual RT rather than the raw RT, as the detection advantage was obtained by comparing the differences between the *manner* and *path-salient* with the neutral condition, respectively. As shown in Figure 4.1, the

English group exhibited the lowest negative average RT difference for *manner-salient* stimuli (*M* = -1.75, *SD* = 3.60), suggesting that when *manner* was prominent, English speakers, on average, detected it more quickly than neutral stimuli. In contrast, the detection advantage was less pronounced for *path-salient* stimuli compared to neutral stimuli (RT difference *M* = -0.39, *SD* = 4.71). However, Mandarin speakers showed slower detection times for both *manner* and *path-salient* stimuli compared to *neutral* stimuli (RT difference *M* = 1.04, *SD* = 6.17; *M* = 1.05, *SD* = 5.29, respectively).

Figure 4.1 *RT differences in manner and path salient stimuli in Chinese and English speakers*



Note: The detection times of motion events are expressed as difference RTs (RT Manner minus RT neutral; RT Path minus RT neutral) for each condition and group. The box plots illustrate the medians and 50% of the transformed motion event detection times within the boxes. Raincloud and violin plots were added to help visualise data distribution patterns between conditions and groups.

To further examine the statistical significance of the effect of language groups, a maximal model was built and converged (Appendix 5: LMM model 1). A significant effect of the *language group* was found (β = -2.75, *SE* = 1.02, *t* value = - 2.68, *p* < .05). To further confirm whether the effect was driven by detection in the differences in detecting *manner-salient* motion animation between language groups, another model was built. The results showed that English speakers were significantly faster to detect motion animation compared to Chinese speakers (residuals, β = -2.83, *SE* = 1.02, *t* value = -2.77, $p < .05$). However, no significant results were found in the *path-salient* condition between language groups (residuals, β = -1.45, *SE* = 0.91, *t* value = -1.59, $p > .05$). This indicated that the detection differences between Chinese and English speakers were of a *manner*.

4.1.2 Discussion

In sum, the cross-linguistic difference observed between monolingual Chinese and English speakers was indeed driven by the way manipulated information, consistent with the language-specific features in English (satellite-framed) and Chinese (equipollently-framed) (Ji, 2017; Wang & Li, 2011b). Consequently, this indicates that b-CFS is sensitive to detect differences in low-level visual features between language groups and between stimuli conditions within each group. However, such differences were not readily observed or interpreted.

For example, in the current pilot study, the language-specific effects were observed by comparing the difference between the RT (*manner*-salient) – mean (RT neutral), and the RT (*path*-salient) – mean (RT neutral). This difference was based on another RT difference obtained from the difference value between the raw RT of detecting the manner manipulation and the mean value of raw RT of detecting the stimuli with no manipulation. Such indirect analysis process brings concerns for interpreting the results, such as, the results was not exactly predicted by the motion typologies (Tamly, 2000; Slobin, 2004) in each language (i.e., equipollently-framed Chinese vs. satellite-framed English). Moreover, using the b-CFS phase alone requires participants to retrieve motion information from long-term memory, which seems not highly reliable, given no language-specific effects were discovered in previous studies require to retrieve motion memory (Gennari *et al*., 2002; Papafragou *et al*., 2002). Thus, to maximally boost the language-specific effects for readily availability in motion, and reflecting the results of reaction times, a prime procedure with sequence images abstracted from each stimulus was added proceeding the b-CFS phase.

4.2 The b-CFS control experiment

The average accuracy rate in the control experiment was 0.88 (*SD* = 0.10). A participant was removed due to the low accuracy rate (< 0.75). The average suppression time of the group with a suppressed mask was 8113 ms (*SD* = 4390), and the average suppression time of the group without a suppressed mask was 5730 ms (*SD* = 1351). A significant main effect of the *mask* was found by comparing the full model (Appendix 5: LMM model 2) with the reduced model (Appendix 5: LMM model 3) ($\chi 2$ (1) = 9.31, p = .02), indicating that the *mask* designed in the current b-CFS experiment effectively suppressed the stimuli in a relatively longer time, thus expanding the time to detect automatic processing (Jiang *et al*., 2007). Consequently, providing evidence to the validity of the employment of b-CFS in exploring the motion cognition in the present study.

4.3 The b-CFS main experiment

4.3.1 Monolingual English and Chinese speakers

4.3.1.1 Results

The Chinese (mean: 92%, *SD* = 0.09) and English (mean: 93%, *SD* = 0.06) speakers reached a very high accuracy rate in the b-CFS experiment. The cut-off point of the accuracy rate was 75% (Franken *et al*., 2011; Slivac *et al*., 2021). Three Chinese and one English native speaker were removed due to the low accuracy rate (<=75%). Four native English speakers were removed because they were fluent in another language, as assessed via their responses on the language background questionnaire. Only correct trials were included in the data analysis. Any data (including fillers) point that was greater than 2.5 standard deviations after removing the mean RT of each participant were excluded as outliers. Moreover, any data point larger than 35 seconds, a cut-off based on the scatter plot, was also removed due to individual differences in b-CFS. Therefore, a total of 119 (9%) data were identified and deleted.

Reaction times (RT) in the present study refer to the duration from stimulus onset until button press. The observed results align with the predictions. Specifically, as shown in Table 4.1, the disparity in average response time to break through suppression between stimuli presented in the *manner mismatch* and those in *path mismatch* conditions was smaller in the Chinese speakers (589 ms) compared to the English speakers (1655 ms). Moreover, both Chinese (26 ms) and English speakers (65 ms) exhibited marginally longer average response times in *full mismatch* conditions compared to *full match* conditions.

Table 4. 1 *Mean and SD of the reaction time (ms) observed in the monolingual Chinese and English speakers under four conditions from the b-CFS experiment*

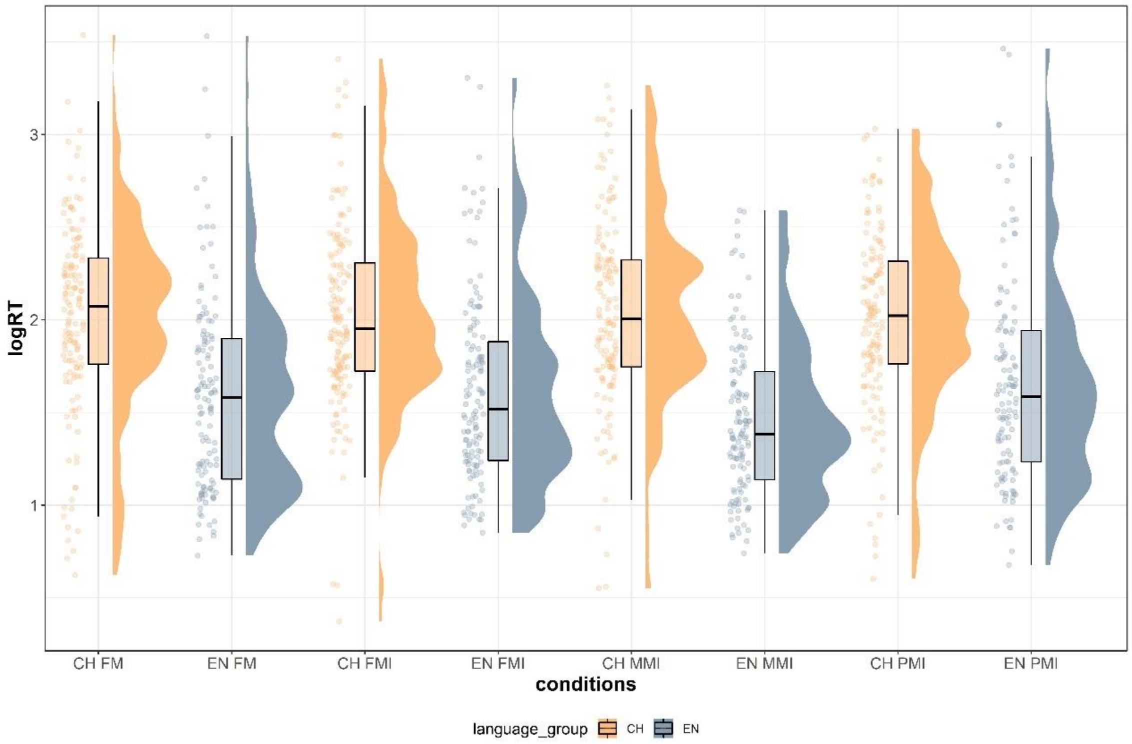
	Mean	*SD*
Chinese		
Full match	8501	4469
Manner mismatch	8809	4791
Path mismatch	8220	4741
Full mismatch	8536	4791
English		
Full match	5815	4389
Manner mismatch	4707	2325
Path mismatch	6362	5039
Full mismatch	5880	4046

Table 4.1 illustrates that the Chinese speakers exhibited similar RTs for detecting stimuli under *manner* and *path mismatch* conditions, while English speakers showed shorter RTs in the *manner mismatch* compared to the *path mismatch* condition. To statistically examine the effects of *language group, condition,* and their *interaction,* linear mixed effects regression models were built using the lme4 package (Bates *et al*., 2015) in R (R Core Team, 2020). The regression models included *language group* (English and Chinese) and *condition* (full match, manner mismatch, path mismatch, and full mismatch) as dummy-coded fixed effects, and the reference level was *full match* condition. *Participant* and the *Item* were included as random effects. As a standard statistical procedure used in related studies measuring RTs to track low-level motion

perception (e.g., Slivac *et al*., 2021), raw RT was log-transformed to reduce skewness in a nonnormal distribution, which served as dependent variables.

A significant main effect of the *language group* was observed when comparing a fully converged full model (Appendix 5: LMM model 4) with two reduced model (Appendix 5: LMM model 5, 6) ($\chi2$ (4) = 46.67, p < .001), indicating that Chinese exhibited overall slower RT compared to English speakers in all four conditions. Furthermore, a significant interaction effect was found between the *language group* (*full match* vs. *manner mismatch*) and the *condition* (*Chinese* vs. *English*) (β = -0.18, *SE* = 0.06, t = -2.854, p = .004; raw RT β = -835 ms), suggesting that the difference in RT between Chinese and English speakers varied depending on the conditions considered.

Figure 4.2 *Logged RTs from the b-CFS experiment in the Chinese and English monolingual groups*



Note: CH = Chinese speakers, EN = English speakers, FM = full match, FMI = full mismatch, MMI = manner mismatch, PMI = path mismatch. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

As shown in Figure 4.2, English speakers appear to vary specifically between *manner* and *path mismatch* conditions, whereas this pattern was not observed in the

Chinese speakers. To assess the specific differences between conditions within each language group, a post hoc test was conducted using the *emmeans* package in R (Lenth, 2021). This analysis provided comprehensive comparisons for both language groups. In the English group, the reaction times in the *manner mismatch* condition was significantly shorter compared to those used under the *path mismatch* condition (β = -0.19, *SE* = 0.07, *t* = -2.82, *p* = .006, raw RT β = -827 ms). Furthermore, the RTs taken in the *manner mismatch* condition were significantly shorter compared to the *full match* (FM) (β = -0.16, *SE* = 0.07, *t* = 2.38, *p* = .020, raw RT β = -852 ms) and the *full mismatch* (FMI) conditions (β = -0.16, *SE* = 0.07, *t* = 2.30, *p* = .024, raw RT β = -852 ms). However, no significant differences were observed between the *manner mismatch* and the *path mismatch* conditions in the Chinese monolingual group (β = 0.03, *SE* = 0.07, *t* = 0.48, *p* = .63, raw RT β = 970 ms). Furthermore, no significant differences were observed in any other comparisons, including *full match* and *mismatch* conditions (β= -0.04, *SE* = 0.07, *t* = -0.06, *p* = .97, raw RT β = -991 ms), *manner mismatch* and *full match* (β = 0.02, *SE* = 0.07, *t* = -0.33, *p* = .75, raw RT β = 980 ms), and also *manner mismatch* and *full mismatch* conditions (β = 0.02, *SE* = 0.07, *t* = -0.27, *p* = .79, raw RT β = 980 ms).

4.3.1.2 Effects of lateralization

The effects of the *mask* were examined by adding it to the full model (Appendix 5: LMM model 7) as a fixed effect, since the rationale for examining the lateralization is to confirm whether the language-specific effects discovered above were only available in the right visual field (that is, when the mask was presented to the left eye in the present study). We relevelled the *manner mismatch* condition as the reference level in the model based on its demonstrated language-specific salience in previous analyses (see **Section** 4.3.1.1). It is important to note that the mask positions were not counterbalanced, since it was dependent on the dominant eyes of the individual participants (left mask: *N* = 29; right mask: *N* = 70), and b-CFS functions well only as long as the mask was presented to the dominant eye; therefore, this leads to unbalanced data to examine the effects of lateralisation: 312 critical data under the left mask condition and 746 critical data under the right mask condition.

No significant main effect of *mask* ($\chi 2$ (8) = 8.60, p = .38) was found when comparing this maximally converged full model with the reduced model (Appendix 5: LMM model 8). However, there were two significant interactions between the *language group* (*manner mismatch* vs. *full match*), *condition* (Chinese vs. English) and *mask* (right vs. left) (β = 0.30, *SE* = 0.14, t = 2.17, p = .03; raw RT β = 741 ms), and between *Language group* (*manner mismatch vs. path mismatch*), *condition* (Chinese vs. English) and *Mask* (right vs. left) (β = 0.28, *SE* = 0.14, t = 1.98, p = .048; raw RT β = 756 ms).

To further examine the statistical difference between conditions within each language group under separate mask condition, a post-hoc test was conducted using the *emmeans* package in R (Lenth, 2021), which displayed all possible comparisons among the four conditions for both language groups within each mask condition. In line with the early results obtained in the monolingual groups, significant differences between conditions were only discovered in English speakers, and such effects emerged only when the mask was placed in the right eyes, suggesting a perceptual advantage in the left vision field, which turns counter to the predictions of laterization. Specifically, in the English group, the RTs in the *mismatch* condition were significantly shorter than in the *path mismatch* condition (β = -0.25, *SE* = 0.08, t = -3.29, p = .001, raw RT β = -779 ms). Moreover, RTs in the *manner mismatch* condition were also significantly shorter compared to the *full match* (β = -0.22, *SE* = 0.08, t = 2.83, p = .005, raw RT β = -803 ms) and the *full mismatch* conditions (β = -0.20, *SE* = 0.08, t = 2.62, p = .01, raw RT β = -819 ms). In contrast, no significant differences were found between the *manner mismatch* and the *path mismatch* conditions when the mask was presented to the left eye (β = -0.07, *SE* = 0.10, t = -0.71, p = .48, raw RT β = -932 ms). No significant differences were found among any comparisons for Chinese speakers, regardless of whether the mask was presented to the left or right eye.

4.3.1.3 Discussion

The RT data shows cross-linguistic differences between Chinese and English consistent with those observed in the pilot study, thus supporting the *manner* rather than *path* salience. That is, the English speakers exhibited a variation in the time taken to emerge into consciousness with stimuli involving manipulation between *manner* and *path*, whereas Chinese did not pattern in this way, arguably due to the equal salience of *manner* and *path* in Chinese. Moreover, through the use of b-CFS whiling employing a prime image as a non-verbal cue, the detection advantage of English *manner* observed in English speakers exhibited a pervasive influence of language on motion cognition at a lower level. Thus, extending the influence of linguistic labels on visual perception in motion event cognition (Luypan *et al*., 2020), even when the labels were reflected in a non-verbal mode (i.e., prime image). More strikingly, the speed advantage in detecting variation in *manner* exhibited by English speakers in the b-CFS suggests a distinct processing pattern in the early stage of visual perception, compared to those observed at a later stage (Ji, 2017). This effect may be attributed to the automatic activation of *manner* labels, enhancing the salience of *manner* information in visual input and facilitating its detection (Perry & Lupyan, 2013), which is consistent with evidence in Papafragou *et al*. (2008), who also discovered that English speakers tended to inspect the *manner* rather than the *path* component at first sight.

Specifically, the time taken to detect changes in motion type in the b-CFS serves as a psychophysiological correlation of higher-level representation. Rapidly assessing changing sensory signals with gradually emerging motion events tends to rely more on *manner*-based processing since *manner* encoding is more prominent in English. Conversely, in Chinese, where information receives less attention in verbal encoding, this linguistic modulation might influence Mandarin speakers' detection of motion events emerging from visual noise. Furthermore, and more importantly, the RT results in the b-CFS were aligned with the assumption proposed by the LRH without being trapped by the circularity, that is, language might indeed affect thought, even without covert language use (Athanasopoulos & Bylund, 2020). However, to further examine the LRH, consistent supporting evidence should be provided in processing motion events at higher level between Chinese and English through utilising the same material. Consequently, two verification experiments with a manipulation of linguistic involvement were conducted, and details will be illustrated in the Chapter 3.

In line with Trueswell and Papafragou (2010), the task effects did not emerge at the early stage of motion encoding, reflecting a low-level advantage that b-CFS leverages. Specifically, both Chinese and English speakers used slightly shorter time in the *full mismatch* (e.g., man carrying dinosaur into room) condition compared to the *full match* (e.g., man carrying suitcase into room) condition, and this difference was minor and did not reach statistical significance, suggesting automatic processing which did not involve higher-level discrimination.

Interestingly, the overall RTs across all four conditions were slower for the Chinese group compared to the English group. This does not appear be related specific cross-linguistic differences related to motion event cognition, as those differences pertain to the processing of *manner* and *path* components. Further explanations will be provided in the discussion chapter.

In terms of the effect of lateralisation, the low-level detections observed in b-CFS contradicted the notion that language-specific effects are predominantly manifest in the right vision field (with the left mask). However, these findings should be interpreted with caution. First, the uneven distribution of the data in the present study warrants consideration, which could affect the robustness of the conclusion. Second, b-CFS involves the presentation of stimuli in an illusionary rather than a clear vision field, differing form early studies examining lateralisation where stimuli were presented on distinct sides of the screen (Mo *et al*., 2011), and participants perceived them as such. Despite presenting stimuli and masks to separate the eyes, they merge into a unified visual field in b-CFS. Additionally, the stimuli in the present study were moving actions, not static stimuli, such as colour patches (Gilbert *et al*., 2008), and according to the feedback collected after the test session, a subset of participants (8%) reported that upon the stimuli breaking through suppression, the stimuli seem to move from one side to another in the vision field within the limited time of spotting the location.

4.3.2 Chinese learners L2 English

RTs obtained from Chinese learners of L2 English were analysed following the data cleaning procedure as set out above. The primary objective of the L2 analysis was to determine whether there was a cognitive transfer in L2 learners within the b-CFS experiment and to what extent this occurred. Initially, L2 learners were compared with the two groups of monolingual speakers in four conditions. Subsequent exploration expanded to potential the factors, as illustrated in Table 4.2, which could potentially influence the RT patterns observed in L2 learners. In addition, the effects of *cultural immersion* and the *language context* were examined. The former was investigated by incorporating data from L2 learners obtained in China. Detailed procedures for the analyses will be provided in the subsequent sections. Given the negative evidence and potential biases associated with investigating lateralisation in b-CFS, its influence on L2 learners is not analysed.

4.3.2.1 Overall Results

The inclusion criteria were identical to those used in the analysis in monolingual English and Chinese speakers. The average accuracy rate in the L2 learners recruited in York was 93.46% (SD = 0.05). Three participants were eliminated due to the low accuracy rate (< 75%) in detecting the correct corners in the b-CFS experiment. Regarding the L2 learners recruited in China, their average accuracy rate was 92.85% (SD = 0.09) and two participants were removed due to a low accuracy rate (< 75%). One extreme (>35 s) was removed based on the scatter plot. Thus, a total of 34 (3%) data was removed.

Table 4.2 *Mean and SD of the reaction time (ms) observed in the L2 learners under four conditions from the b-CFS experiment*

	Mean	*SD*
Full match	5878	3047

Manner mismatch	6741	4622
Path mismatch	7234	5122
Full mismatch	6901	5126

Table 4.2 illustrates the average RTs observed in L2 learners, demonstrating a pattern similar to that observed in monolingual English speakers. Specifically, L2 learners exhibited quicker (493 ms) detection of appearing stimuli in the *manner mismatch* compared to the *path mismatch* condition.

In line with the statistical analysis in the two monolingual groups, the analysis of the Chinese learners of L2 English speaker data were also carried out using linear mixed-effects regression via the package lme4 package (Bates *et al*., 2015) in R (R Core Team, 2020). In the maximally converged regression model (Appendix 5: LMM model 9), *condition* (full match, full mismatch, manner mismatch, and path mismatch) and *language group* (monolingual English, Chinese speakers, and Chinese learners of L2 English) were dummy coded as fixed effects. The *Manner mismatch* condition and *L2 learners* are the reference levels. *Participant* and *item* were random effects.

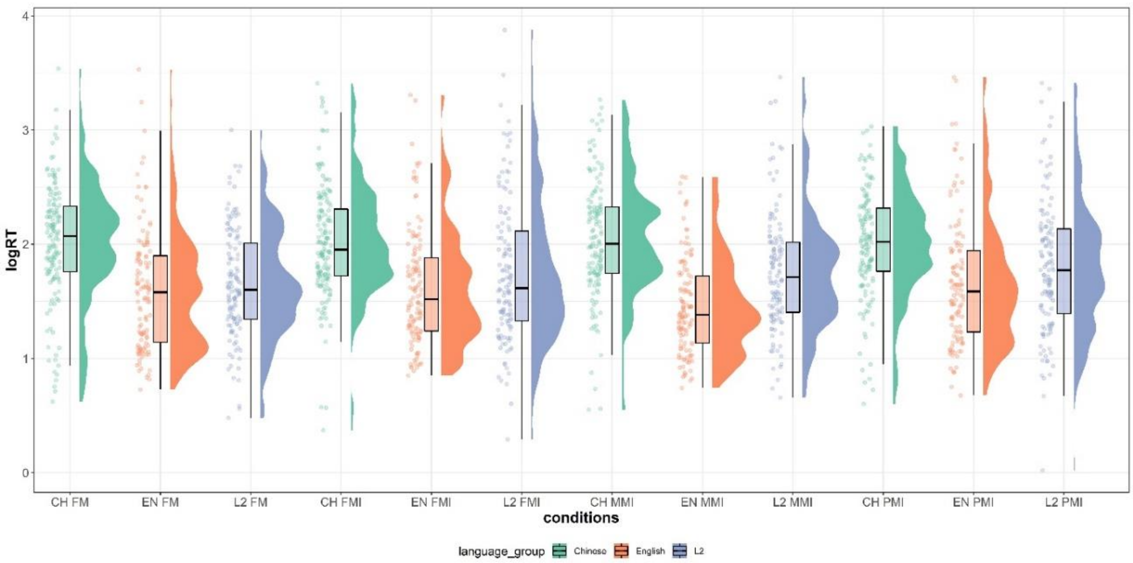
Several significant effects were observed when comparing the full model with three reduced models (Appendix 5: LMM model 10, 11, 12), respectively, including a significant main effect of *language group* ($\chi2$ (8) = 59.35, p < .001), indicating that the L2 learners were different from the other two language groups overall. Specifically, L2 learners exhibited generally significantly slower RT compared to English speakers (β = -0.29, SE = 0.08, t = -3.71, p < .001; raw RT β = -748 ms) but faster than Chinese speakers (β = 0.28, SE = 0.08, t = 3.59, p < .001; raw RT β = 756 ms). Additionally, a significant main effect of the *condition* was also obtained ($\chi2$ (9) = 20.28, p = .02), indicating that the time taken in the *mismatch* condition was observed to be different from the other three conditions in general. Furthermore, two significant and one marginal significant effects of the *interaction* were identified. These include a significant interaction between the *condition* (*manner* mismatch *vs.* full match) and the *language group* (L2 vs. English) (β = 0.26, SE = 0.07, t = 3.52, p = .0004; raw RT β = 771 ms), a significant interaction between *condition* (manner mismatch *vs.* full mismatch) and the *language group* (L2 vs. English) (β = 0.18, *SE* = 0.07, t = 2.47, p = .014; raw RT β = 835 ms), and a marginally significant interaction between the *condition* (manner mismatch *vs.* path mismatch) and the *language group* (L2 vs. English) (β = 0.14, *SE* = 0.07, t = 1.91, p = .06; raw RT β = 869 ms), demonstrating the *manner mismatch* interacted with all the other three conditions between L2 learners and English speakers.

To further explore the statistical difference within the L2 group, a post-hoc test was conducted using the *emmeans* package in R (Lenth, 2021). RTs in the *manner mismatch* were not significantly different from those in the *path mismatch* condition (β = -0.05, *SE* = 0.08, t = -0.66, p = .51, raw RT β = -951 ms). Furthermore, the RTs in the *full match* condition were quicker in the *path mismatch* condition, but this difference did not reach significance (β = -0.14, *SE* = 0.08, t = -1.87, p = .06, raw RT β = -869 ms) than. Consistent with the findings in monolingual speakers, the L2 learners also exhibited slightly faster stimuli processing in the *full match* condition compared to the *full mismatch* condition; however, this difference was not statistically significant (β = -0.07, *SE* = 0.08, t = -0.85, p = .40, raw RT β = -932 ms).

In Figure 4.3, L2 learners seem to pattern like the English speakers in the *full match*, *full mismatch*, and *path mismatch* conditions, and they display a pattern in the *manner mismatch* condition that appears in between that seen for the English and the Chinese monolinguals. An additional post-hoc test was conducted to further investigate the statistical differences between L2 learners and two monolingual speakers within each condition. In the *manner mismatch* condition, L2 learners were significantly faster than Chinese speakers (β = -0.28, *SE* = 0.08, t = -3.57, p < .001, raw RT β = -756 ms) and significantly slower than English speakers (β = 0.30, *SE* = 0.08, t = 3.68, p < .001, raw RT β = 741 ms). Whereas in the other three conditions, L2 learners were consistently significant from Chinese speakers but mirrored the English speakers. Specifically, in the *path mismatch* condition, L2 learners were significantly faster than Chinese speakers (β = -0.20, *SE* = 0.08, t = -2.55, p = .01, raw RT β = -819 ms), but only reached marginal significance compared to English speakers (β = 0.15, *SE* = 0.08, t

= 1.92, p = .06, raw RT β = 861 ms). In the *full match* condition, the L2 learners were significantly faster than the Chinese speakers (β = -0.35, SE = 0.08, t = -4.52, p < .001, raw RT β = -705 ms), but comparable to the English speakers (β = 0.04, SE = 0.08, t = 0.47, p = .64, raw RT β = 961 ms). In the *full mismatch* condition, the L2 learners were also significantly faster than Chinese speakers (β = -0.29, SE = 0.08, t = -3.77, p < .001, raw RT β = -748 ms), but patterned with English speakers (β = -0.14, SE = 0.08, t = 1.45, p = .15, raw RT β = -869 ms).

Figure 4.3 *Logged RTs observed in monolingual speakers and L2 learners from the b-CFS experiment*



Note: CH = Chinese speakers, EN = English speakers, FM = full match, FMI = full mismatch, MMI = manner mismatch, PMI = path mismatch. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

4.3.2.2 Overall Discussion

RTs observed in L2 learners obtained from the b-CFS experiment suggest that the cognitive transfer in L2 learners were constrained by their native language. Specifically, L2 learners were found to break through mask suppression and detect the positions of the stimuli significantly faster than Chinese speakers and slower than English speakers in general. This general distinction between monolingual speakers and L2 learners demonstrates a cognitive transfer of the encoding *manner* and *path* in motion events, diverging from the cognitive process based on L1 and L2. However, language-specific differences between *manner* and *path mismatch* conditions within the L2 group were not observed; specifically, although L2 learners spent less time detecting the locations of stimuli in a *manner mismatch* than *path mismatch* condition, this distinction did not reach statistical significance. This suggests that L2 learners in the present study were transitioning from L1-based L2-based cognitive processes, but that this process was strongly guided by their native language (Pavlenko, 2011).

Furthermore, upon closer examination of the significant differences between language groups within each stimuli condition, it was observed that the significant divergence from two monolingual groups was only found in the *manner mismatch* condition. In contrast, a distinct pattern with consistent results was obtained in the other three conditions, where the L2 learners mirrored the English speakers, but not the Chinese speakers. This provides supporting evidence for the claim that cognitive transfer is influenced by similarity and differences in salient features between the source and target languages (Wang & Li, 2019), demonstrating a gradual process of transferring from L1 to L2-like cognitive patterns. Specifically, Chinese and English share a similar weight of salience of path, while English has a higher salience in manner (Feist, 2016). Therefore, L2 learners should acquire a target-like cognitive process more readily in the *path mismatch* compared to the *manner mismatch* condition. B-CFS detects the early stage of visually processing *manner* and *path* components in motion; consequently, the in-between consequence in the *manner mismatch* condition and L2-like outcome in other conditions demonstrates a gradual transformation of the underlying concepts. Specifically, shared identical or similar concepts between source and target languages are acquired more quickly than those with distinct features (Ringbom, 2007).

More importantly, the results obtained in b-CFS in L2 learners align with the LRH, demonstrating that without conscious linguistic involvement, language (experience) shapes thoughts. Furthermore, language indeed influences visual perception, and b-CFS captured delicate changes in the L2 learner when processing low-level visual features driven by higher-level categorical cognitive patterns specific to source and target languages, respectively. However, such results are inconsistent with earlier literature, which found a cognitive shift to the L2-based pattern in their advanced L2 learners (Ji, 2017) and Chinese-English bilinguals (Wang & Li, 2011b). Consequently, in the following sections, potential factors that may affect the L2 results in the current b-CFS experiment are analysed and discussed.

4.3.2.3 Predictor for the cognitive outcomes in L2 learners

4.3.2.3.1 Language context results

Table 4.3 *Mean and SD of the reaction time (ms) observed in the L2 learners in language contexts under four conditions from the b-CFS experiment*

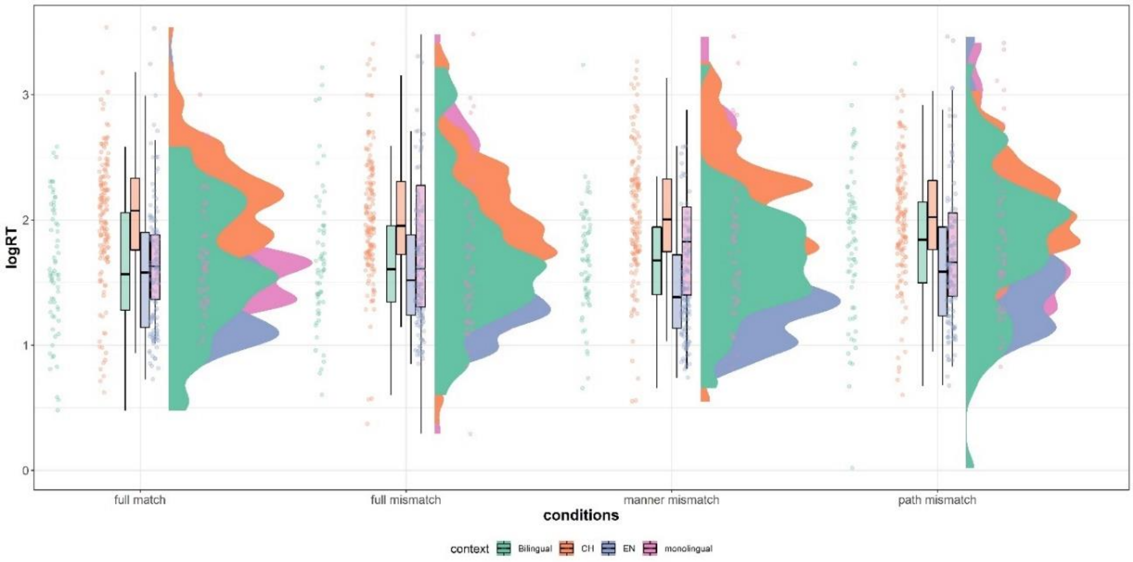
Language group	*Mean*	*SD*
Bilingual context		
Full match	5672	2831
Manner mismatch	5882	3254
Path mismatch	7218	4368
Full mismatch	6499	4753
Monolingual context		
Full match	6091	3265
Manner mismatch	7586	5554
Path mismatch	7251	5848
full mismatch	7271	5459

Table 4.3 illustrates that the L2 learner tested in a bilingual context spent less (1336 ms) time on average in the *manner mismatch* than those in the *path mismatch* condition, while those who were tested in monolingual context spent more time (335 ms) processing stimuli in the *manner mismatch* condition compared to those in the *path mismatch* condition. To directly examine the effects of language-specific differences, the *manner mismatch* condition and L2 learners tested in *monolingual context* were set as reference levels in a maximally converged LMM (Appendix 5: LMM model 13). Three significant effects emerged after comparing the full model with three reduced models (Appendix 5: LMM model 14, 15, 16), including a main effect of *language*

group, suggesting that the L2 learners tested in *monolingual context* were overall different from the rest of the three language groups ($\chi 2$ (12) = 63.42, p < .001); a main effect of *condition*, suggesting that the time taken in the *manner mismatch* condition was overall different from those in the rest of the conditions ($\chi 2$ (12) = 23.48, p = .02); and a significant interaction between *language group* and *condition*, demonstrating that the difference between conditions can vary across different language groups ($\chi 2$ (9) = 22.21, p = .01). Furthermore, L2 learners tested in the *monolingual contex*t were found to interact significantly with *English speakers* in three comparisons, including *manner mismatch* and *path mismatch* (β = 0.23, *SE* = 0.09, t = 2.56, p = .01, raw RT β = 795 ms), *manner mismatch* and *full match* (β = 0.30, *SE* = 0.09, t = 3.31, p < .001, raw RT β = 741 ms), *manner mismatch* and *full mismatch* (β = 0.21, *SE* = 0.09, t = 2.33, p = .02, raw RT β = 811 ms). Also, a marginally significant effect emerged between *condition* (manner mismatch *vs.* path mismatch) and *language group* (L2 in monolingual context vs. L2 in bilingual context) (β = 0.18, *SE* = 0.10, t = 1.74, p = .08, raw RT β = 835 ms); however, no significant difference between the two L2 groups was observed regardless of the language contexts assigned to them (β = -0.16, *SE* = 0.11, t = -1.46, p = .15, raw RT β = -852 ms), indicating that these two L2 groups shared similar patterns of overall RTs used in all conditions.

To further investigate the statistical differences between conditions within each L2 group, a post-hoc test was performed using the *emmeans* package in R (Lenth, 2021), which showed all possible comparisons for each L2 group. In the *bilingual context*, RTs taken in the *path mismatch* were significantly longer than those in the *full match* condition (β = -0.19, *SE* = 0.09, t = -2.10, p = .04, raw RT β = -827 ms), but not significantly longer than those in the *manner mismatch* condition (β = -0.14, *SE* = 0.09, t = -1.51, p = .13, raw RT β = -869 ms). No other comparisons were statistically significant, neither in a the L2 learners tested in *monolingual context*.

Figure 4.4 *Logged RTs in L2 learners in monolingual and bilingual contexts and two monolingual groups from the b-CFS experiment*



Note: CH = Chinese speakers, EN = English speakers. Bilingual = bilingual context, monolingual = monolingual context. The box plots show the medians and the box demonstrates 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

As shown in Figure 4.4, contrasting patterns were observed between L2 learners in a monolingual and bilingual context. Specifically, the L2 learners tested in a bilingual context had demonstrated a quicker reaction time in the *manner mismatch* compared to *path mismatch* conditions, whereas the L2 learners tested in monolingual reacted faster in the *path mismatch* compared to the *manner mismatch* conditions. Using the same package, an additional post-hoc analysis was conducted to explore the statistical differences between language groups within each condition. In the *manner mismatch* condition, both L2 groups were significantly faster than Chinese speakers (monolingual: β = -0.20, *SE* = 0.10, *t* = -2.05, *p* = .04, raw RT β = -819 ms; bilingual: β = -0.36, *SE* = 0.10, *t* = -3.74, *p* < .001, raw RT β = -698 ms), and slower than English speakers (monolingual: β = 0.38, *SE* = 0.10, *t* = 3.84, *p* < .001, raw RT β = 684 ms; bilingual: β = 0.21, *SE* = 0.10, *t* = 2.17, *p* = .03, raw RT β = 811 ms). In the *path mismatch* condition, both groups of L2 learners were significantly faster than the Chinese speakers (monolingual: β = -0.21, *SE* = 0.10, *t* = -2.16, *p* = .03, raw RT β = -811 ms; bilingual: β = -0.19, *SE* = 0.10, *t* = -2.01, *p* = .05, raw RT β = -827 ms), but comparable to the

English speakers (monolingual: $\beta = 0.15$, $SE = 0.10$, $t = 1.48$, $p = .14$, raw RT $\beta = 861$ ms; bilingual: $\beta = 0.16$, $SE = 0.10$, $t = 1.65$, $p = .10$, raw RT $\beta = 852$ ms). In *full match* condition, both L2 groups were significantly faster than Chinese speakers (monolingual: $\beta = -0.31$, $SE = 0.10$, $t = -3.20$, $p = .002$, raw RT $\beta = -733$ ms; bilingual: $\beta = -0.39$, $SE = 0.10$, $t = -4.11$, $p < .001$, raw RT $\beta = -677$ ms), but comparable to the English speakers (monolingual: $\beta = 0.08$, $SE = 0.10$, $t = 0.81$, $p = .42$, raw RT $\beta = 923$ ms; bilingual: $\beta = -0.01$, $SE = 0.10$, $t = -0.05$, $p = .96$, raw RT $\beta = -990$ ms). In the *full mismatch* condition, the L2 learners tested in monolingual mode were significantly faster than the Chinese speakers ($\beta = -0.24$, $SE = 0.10$, $t = -2.49$, $p = .01$, raw RT $\beta = -787$ ms), and emerged marginally significantly different when compared to the English speakers ($\beta = 0.17$, $SE = 0.10$, $t = 1.75$, $p = .08$, raw RT $\beta = 844$ ms), while those test in bilingual mode were significantly faster than the Chinese speakers ($\beta = -0.35$, $SE = 0.10$, $t = -3.61$, $p < .001$, raw RT $\beta = -705$ ms), but comparable to the English speakers ($\beta = -0.06$, $SE = 0.10$, $t = 0.60$, $p = .55$, raw RT $\beta = -942$ ms).

4.3.2.3.2 Language context discussion

No main effect of language context was observed between the L2 learners tested in monolingual and bilingual contexts in the b-CFS experiment, which is consistent with our prediction. However, the L2 learners tested in bilingual mode were observed to break through mask suppression significantly more slowly in the *path mismatch* than the *full match* condition. No significant differences were observed in the L2 learners tested in the monolingual context. There are three possible reasons for the absence of effects from the language context. Firstly, the b-CFS paradigm examined the automaticity of processing *manner* and *path* manipulated components in motion in the current study, and such automaticity is independent of the influence of immediate language use (i.e., language context) (Jiang *et al*., 2007). If this is the case, we might be able to find the effects of language context in the two verification experiments later, since they examine the language-specific effects from a higher categorical level (Zwaan *et al*., 2002), these two non-verbal experiments were not able to prevent participants from using language in silence, which provides opportunities for observing the influence of immediate language use. Secondly, the primary condition of observing the flexible switch between source and target language based cognitive processes is the availability of two equally weighted cognitive processes in L2 learners (Wolff & Ventura, 2009), for instance, L2 learners with near native L2 proficiency. However, the L2 learners in the present study were of mixed intermediate and advanced proficiency, they might not be proficient enough to acquire *manner* salient concepts since it's not shared between Chinese and English (Ringbom, 2007; Ji, 2017). Consequently, only L2 learners tested in bilingual mode exhibited a significantly slower response when spotting locations with *path* manipulated stimuli than those without manipulations. Third, AO might be another potential predictor of the current outcome. Specifically, early bilinguals (e.g., AO < 12) tend to exhibit L2-like cognitive patterns regardless of the tested languages (Kamenetski *et al*., 2022; Bylund, 2009), while late bilinguals were found to be mediated by immediate language use (Athanasopoulos *et al*., 2015). This is also seen in the results of age of onset of learning an L2 in **Section: 4.3.2.4**. below. The average AO in the current study was 8.56, which classified our L2 learners as early L2 learners. Thus, the effects of language context might indeed be likely to be absent.

4.3.2.3.3 Cultural immersion results

In the present study, cultural immersion refers to the experience of living in an L2-speaking country (Cook *et al*., 2006). To examine the effects of cultural immersion, thirty-six participants in the L2 group with comparable OPT scores (<= 78, and >=50) were selected to compare with L2 learners lived in China.

In Table 4.4, both L2 groups spent on average less time in the *manner mismatch* than in the *path mismatch* condition. Furthermore, the difference in RTs between *manner mismatch* and *path mismatch* was smaller (539ms) in L2 learners lived in China, compared to those living in the UK (938 ms). To directly examine potential language-specific effects, the *manner mismatch* condition and the L2 group who *lived in China* were set as reference levels in a maximally converged LMMs (Appendix 5: LMM model 17). By comparing the full model with three reduced models (Appendix 5: LMM model 18, 19, 20), a main effect of the *language group* was found ($\chi2$ (12) =

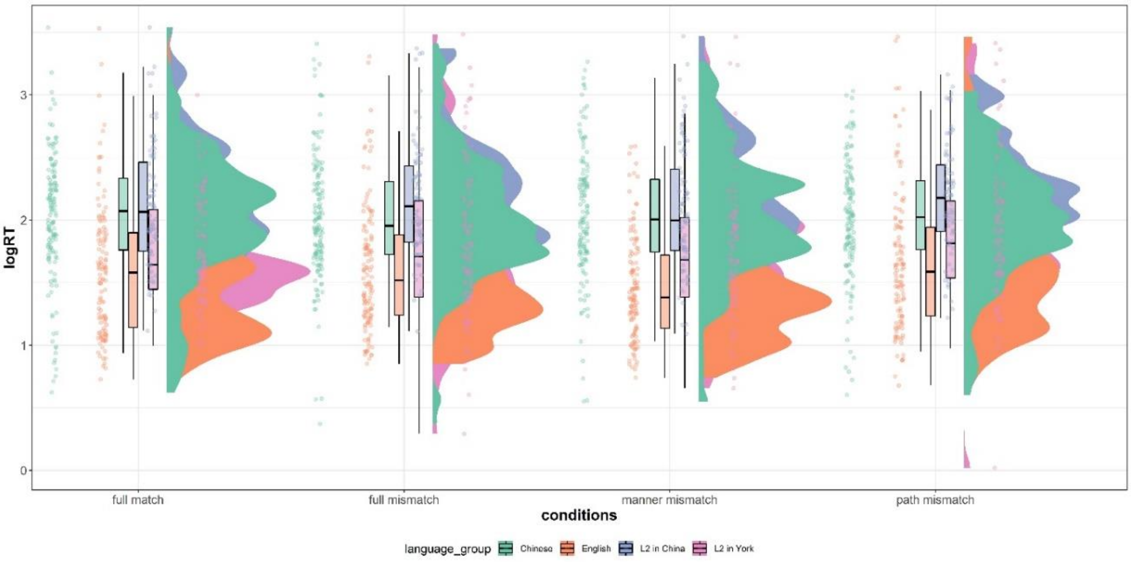
76.36, p < .001), suggesting that the L2 learners who lived in China in general performed differently from the other language groups. Specifically, the L2 learners in China were generally significantly different from L2 learners living in the UK (β = -0.39, SE = 0.09, t = -4.16, p < .001, raw RT β = -677 ms) and the English speakers (β = -0.66, SE = 0.09, t = -7.47, p < .001, raw RT β = -517 ms), but similar to the Chinese speakers (β = -0.09, SE = 0.09, t = -1.10, p = .27, raw RT β = -914 ms). No main effect of *condition* emerged ($\chi 2$ (12) = 15.99, p = .19), indicating the RTs taken in the *mismatch* condition was overall not in general significantly different from any of the other conditions. No effects of *interaction* were observed either.

Table 4.4 *Mean and SD of the reaction times (ms) observed in the L2 learners who lived in York and China under four conditions from the b-CFS experiment*

	Mean	*SD*
L2 learners in York		
Full match	6296	3025
Manner mismatch	6757	4957
Path mismatch	7740	5512
Full mismatch	7323	5387
L2 learners in China		
Full match	9363	5426
Manner mismatch	9090	5216
Path mismatch	9629	4481
full mismatch	9608	5354

To further investigate the differences between conditions within each L2 group, a post-hoc test was performed using the *emmeans* package in R (Lenth, 2021), which displayed all possible comparisons between conditions for each L2 group. No comparisons were statistically significant in L2 learners lived either in the UK or China.

Figure 4.5 *logged RTs in L2 learners in York and China, and two monolingual groups from the b-CFS experiment*



Note: L2 in China = L2 learners lived in China, L2 in York = L2 learners lived in York. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

Figure 4.5 illustrates that the L2 learners in the UK and those living in China seem to resemble English and Chinese speakers in the *full match* condition, respectively. Whereas in the *manner mismatch* condition, only the L2 learners in China mirrored the Chinese speakers, and the UK L2 learners diverged from English speakers. An additional post-hoc analysis was conducted using the *emmeans* package in R (Lenth, 2021), which displayed all possible comparisons for each condition. In the *manner mismatch* condition, L2 learners living in China were significantly slower than the English speakers (β = 0.64, *SE* = 0.09, *t* = 7.27, *p* < .001, raw RT β = 527 ms) and the UK L2 learners (β = 0.37, *SE* = 0.09, *t* = 3.93, *p* < .001, raw RT β = 691 ms), but comparable to the Chinese speakers (β = 0.07, *SE* = 0.09, *t* = 0.80, *p* = .43, raw RT β = 932 ms). In the *path mismatch* conditions, the L2 learners living in China were also significantly slower than English speakers (β = 0.53, *SE* = 0.09, *t* = 5.96, *p* < .001, raw RT β = 589 ms), the UK L2 learners (β = 0.31, *SE* = 0.09, *t* = 3.34, *p* < .001, raw RT β = 733 ms) and the Chinese speakers (β = 0.17, *SE* = 0.09, *t* = 2.00, *p* = .047, raw RT β = 844 ms). In the *full match* condition, the L2 learners living in China were significantly slower than the English speakers (β = 0.52, *SE* = 0.09, *t* = 5.96, *p* < .001, raw RT β = 595 ms), the L2 learners lived in UK (β = 0.40, *SE* = 0.09, *t* = 4.29, *p* < .001, raw RT β = 670 ms), but not the Chinese speakers (β = 0.14, *SE* = 0.09, *t* = 1.60, *p* = .11, raw RT β = 869 ms). In the *full mismatch* condition, the L2 learners who were living in China were also significantly slower than the UK L2 learners (β = 0.54, *SE* = 0.09, *t* = 6.16, *p* < .001, raw RT β = 583 ms), the English speakers (β = 0.36, *SE* = 0.09, *t* = 3.92, *p* < .001, raw RT β = 698 ms), but not the Chinese speakers (β = 0.13, *SE* = 0.09, *t* = 1.56, *p* = .12, raw RT β = 878 ms).

4.3.2.3.4 Cultural immersion discussion

Cultural immersion was found to significantly affect the RTs in the b-CFS between the L2 learners living in the UK and those who had equivalent L2 proficiency but were living in China. Specifically, the UK L2 learners were overall faster in detecting stimuli locations in the four conditions in the b-CFS experiment. Although the L2 learners living in China did not show significantly faster RT patterns in the *manner mismatch* compared to the *path mismatch* condition, as did the UK L2 learners, they responded significantly more slowly in comparison to all other three language groups, including the Chinese speakers, in the *path mismatch* condition, indicating a cognitive restructuring toward an English-like cognitive process mediated by cultural immersion (Pavlenko, 2011). A surprising RT pattern observed in the L2 learners in China is that they were slower than Chinese speakers in all four conditions, even though it did not reach statistical significance. This is consistent with Ji (2017), who also discovered that even though advanced L2 learners (recruited in China) used significantly different amounts of time to select stimuli with *manner* and *path* manipulations, respectively, the RTs recorded for both selections were slower than Chinese monolinguals. In contrast, the bilingual Cantonese-English bilinguals recruited in the UK in Wang and Li (2021b) displayed a divergent RT pattern when selecting *manner* and *path* alternates in a similarity judgment task. Specifically, they exhibited an in-between RT pattern when compared with English and Cantonese monolinguals for selecting alternates with same *path* (different *manner*), whereas they were even faster than English speakers when selecting alternates with same *manner* (different *path*). Therefore, apart from language-specific effects, a general cognitive process independent of cross-linguistic differences may have occurred in the L2 learners who lived in different language-dominant communities, that is, UK L2 learners were overall quicker than the L2 learners in China in detecting low-level stimuli signals displaying in the b-CFS experiment.

4.3.2.4 Other predictors for the cognitive outcomes in L2 learners

Table 4.5 illustrates the *mean* and *SD* for each potential predictor that may influence L2 performance, i.e., RTs obtained to break suppression and detect the correct corner of the stimuli in b-CFS. To directly explore the influence of these factors, using the same method, a maximally LMMs (Appendix 5: LMM model 21) was converged by including all these factors as fixed effects in the model, and *manner mismatch* condition as the reference level.

Table 4.5 *Mean and SD of potential predictors observed in L2 learners from the b-CFS experiment*

Measure	*Mean* (*SD*)
Time in the UK (months)	14.81 (34.90)
Daily speaking of English (%)	36.75 (22.41)
Daily writing of English (%)	53.42 (28.85)
Onset of learning English as L2 (AO)	8.56 (2.67)
Oxford Placement Test score (maximum 100)	63.92 (11.78)
IELTS score	6.46 (0.71)

Table 4.6 *Statistical results of potential predictors recorded in L2 learners from the b-CFS experiment*

Predictors	*Estimate*	*SE*	*t*	*p*
Time in the UK (months)	0.001	0.001	0.51	0.61
Daily speaking of English (%)	0.00003	0.002	0.01	0.99
Daily writing of English (%)	-0.002	0.001	-1.52	0.14
Onset of learning English as L2 (AO)	-0.02	0.02	-0.92	0.37

Oxford Placement Test score (maximum 100)	0.003	0.003	0.67	0.51
IELTS score	-0.17	0.08	-2.08	0.04

The statistical results of the full model are presented in Table 4.6. Only IELTS score was found to significantly affect the RTs in L2 learners. Furthermore, a marginally significant interaction between *condition* (manner mismatch and full match) and *daily use of L2 written English* emerged (β = -0.005, *SE* = 0.003, *t* = -1.95, *p* = .06, raw RT β = -995 ms), when adding each predictor as a fixed factor in a series of converged models (e.g., *logRT ~ 1 + condition * use of L2 written English + (1 | participant) + (1 | item)*). To further explore these two effects, separate analyses were conducted and presented in the following sections.

4.3.2.4.1 The effects of L2 English proficiency results

To examine the effects L2 English proficiency, the L2 learners were divided into two subgroups based on the criteria established by the International English Language Testing System (IELTS score). Specifically, L2 learners who scored above 6.5 on IELTS score were classified as advanced learners (*N* = 31), while the remaining L2 learners were classified as intermediate learners (*N* = 17).

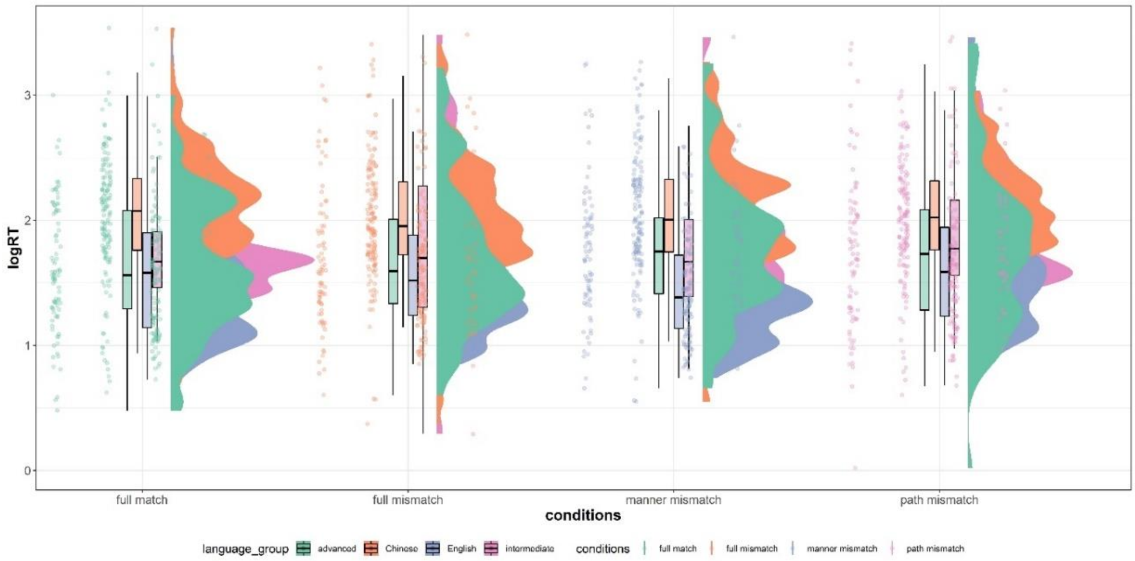
In Table 4.7, both the advanced and intermediate L2 learners spent less time in the *manner mismatch* condition than in the *path mismatch* condition. To directly examine the effects of language-specific differences, the *manner mismatch* condition and *intermediate L2 learners* were set as reference levels in a maximally converged LMMs (Appendix 5: LMM model 22). When comparing the full model with three reduced models (Appendix 5: LMM model 23, 24, 25), several significant effects emerged. Firstly, a main effect of the *language group* ($\chi 2$ (12) = 61.93, *p* < .001), indicating substantial differences in log-transformed RTs across various language groups. The divergence of intermediate L2 learners from the general pattern observed in the other three language groups. Second, a main effect of *condition* was evident ($\chi 2$ (12) = 22.06, *p* = .04), suggesting significant differences between different conditions. In particular, the time taken in the *manner mismatch* condition differed from the overall pattern observed in the other three conditions. Third, a main effect of the interaction between *language group* and *condition* emerged ($\chi 2$ (9) = 20.78, p = .01), indicating that the differences between conditions may vary among different language groups. However, no significant difference was observed between these two L2 groups (β = -0.01, SE = 0.12, t = -0.10, p = .92, raw RT β = -990 ms), indicating these two L2 groups that they shared overall similar patterns of RTs observed across all conditions.

Table 4.7 *Mean and SD of the reaction time (ms) observed in intermediate and advanced L2 learners from the b-CFS experiment*

	Mean	*SD*
Advanced L2 learners		
Full match	5720	3156
Manner mismatch	6791	4575
Path mismatch	7204	5661
Full mismatch	7105	6599
Intermediate L2 learners		
Full match	6206	2816
Manner mismatch	6658	4749
Path mismatch	7290	4007
Full mismatch	7461	5834

To further investigate the differences between conditions within each L2 group, a post-hoc test was performed using the *emmeans* package in R (Lenth, 2021), which displayed all possible comparisons for each L2 group. No significant differences between *manner* and *path mismatch* were found in either of the two L2 groups (intermediate: β = -0.11, SE = 0.10, t = -1.03, p = .30, raw RT β = -896 ms; advanced: β = -0.02, SE = 0.09, t = -0.19, p = .85, raw RT β = -980 ms), suggesting that although both L2 groups processed stimuli in *manner mismatch* faster than those in the *path mismatch* conditions, none of these reached statistical significance. However, in the advanced L2 learners, a marginally significant difference was discovered between *full match* and *full mismatch* conditions (β = -0.16, SE = 0.09, t = -1.88, p = .06, raw RT β = -852 ms), indicating that advanced learners were slower to process stimuli with path alteration than those without alteration. Furthermore, a marginal significant difference was also observed between the *manner mismatch* and the *full match* conditions (β = 0.15, *SE* = 0.09, *t* = 1.68, *p* = .096, raw RT β = 860 ms), suggesting that the advanced learners were also slightly slower to process stimuli with manner alteration than those without. No significant differences between any conditions were observed in intermediate L2 learners.

Figure 4.6 *Logged RTs in advanced and intermediate L2 learners, and two monolingual groups from the b-CFS experiment*



Note: advanced = advanced L2 learners, intermediate = intermediate L2 learners. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

In Figure 4.6, it can be seen that the advanced and intermediate L2 learners shared similar patterns in the *full match*, *full mismatch*, and *manner mismatch* condition, but diverged in two directions in the *path mismatch* condition. Using the same package, an additional post-hoc analysis was conducted to explore the differences between language groups within each condition. In the *manner mismatch* condition, both L2 groups were significantly faster than the Chinese speakers (intermediate: β = -0.27, *SE* = 0.10, *t* = -2.47, *p* = .01, raw RT β = -763 ms; advanced: β = -0.28, *SE* = 0.09, *t* = -3.14, *p* = .002, raw RT β = -756 ms) and slower than the English speakers (intermediate: β = 0.30, *SE* = 0.10, *t* = 2.77, *p* = .006, raw RT β = 741 ms; advanced: β = 0.29, *SE* = 0.10, *t*

= 3.21, p = .002, raw RT β = 748 ms). In the *path mismatch* condition, the advanced L2 learners were significantly faster than Chinese speakers (β = -0.23, *SE* = 0.09, t = -2.65, p = .01, raw RT β = -795 ms), but comparable to the English speakers (β = 0.12, *SE* = 0.10, t = 1.28, p = .20, raw RT β = 887 ms). Whereas intermediate L2 learners were comparable to the Chinese speakers (β = -0.13, *SE* = 0.11, t = -1.22, p = .22, raw RT β = -878 ms), but significantly slower than the English speakers (β = 0.22, *SE* = 0.11, t = 1.98, p = .049, raw RT β = 803 ms). In the *full match* condition, both L2 groups were significantly faster than the Chinese speakers (intermediate: β = -0.25, *SE* = 0.11, t = -2.23, p < .001, raw RT β = -779 ms; advanced: β = -0.40, *SE* = 0.09, t = -4.60, p < .001, raw RT β = -670 ms), and comparable to the English speakers (intermediate: β = 0.14, *SE* = 0.11, t = 1.25, p = .21, raw RT β = 869 ms; advanced: β = -0.02, *SE* = 0.09, t =-0.18, p = .86, raw RT β = -980 ms). In the *full mismatch* condition, both L2 groups were significantly faster than the Chinese speakers (intermediate: β = -0.24, *SE* = 0.11, t = -2.25, p = .02, raw RT β = -787 ms; advanced: β = -0.32, *SE* = 0.09, t = -3.59, p < .001, raw RT β = -726 ms), and comparable to the English speakers (intermediate: β = 0.16, *SE* = 0.11, t = 1.48, p = .14, raw RT β = 852 ms; advanced: β = 0.09, *SE* = 0.09, t =0.99, p = .32, raw RT β = 914 ms).

4.3.2.4.2 The effects of L2 English proficiency discussion

L2 proficiency (operationalised by IELTS scores) was found to significantly influence RT patterns in the b-CFS experiment within the L2 learners, however, this main effect disappeared when we separated them into intermediate (IELTS score < 6.5) and advanced (IELTS score > = 6.5) groups. This might be due to the pervasive influence of the L1. As we can see in the *path mismatch* condition, the advanced and intermediate L2 learners were indeed observed to vary and mirrored the English and Chinese speakers, respectively. This suggests that the L2 learners were more likely to have undergone some cognitive transfer in motion features shared by both source and target languages. The L1 influence constrained the cognitive shift when the motion feature is absent or less salient (at a greater degree) (Pavlenko, 2011). Specifically, the L2-based pattern observed in L2 learners when detecting *path* manipulated stimuli might be due to the relatively equivalent salience of *path* in Chinese and English. Whereas the L2 learners exhibit a distinct pattern compared with both monolingual speakers when detecting the *manner* manipulation might be due to that English *manner* is more salient than Chinese *manner* (Slobin, 2004; 2006). Conversely, this also brings into doubt the validity of claiming cognitive transfer in previous studies investigating this issue and which used verbal (Sachs & Coley, 2006; Panayiotou, 2004) or non-verbal (Cook et a., 2006) experiments but without measures to prevent participants from using language in sub-vocally.

4.3.2.4.3 Daily use of L2 written English results

To explore the effects of writing in the L2 in daily life on RTs in L2 learners, they were divided into low frequency (LF, N = 17) of use of written language (< 50%) and high frequency (HF, N = 31) of written language use (>= 50%) groups based on the median (50%) of the frequency of using L2 writing by each participant.

Table 4.8 *Mean and SD of the reaction times (ms) observed in L2 learners with high and low frequency of language use in L2 written from the b-CFS experiment*

	Mean	*SD*
HF of using written L2		
Full match	5551	2808
Manner mismatch	6643	4125
Path mismatch	7240	5219
Full mismatch	6640	4556
LF of using written L2		
Full match	6447	3380
Manner mismatch	6936	5534
Path mismatch	7223	5005
full mismatch	7352	6017

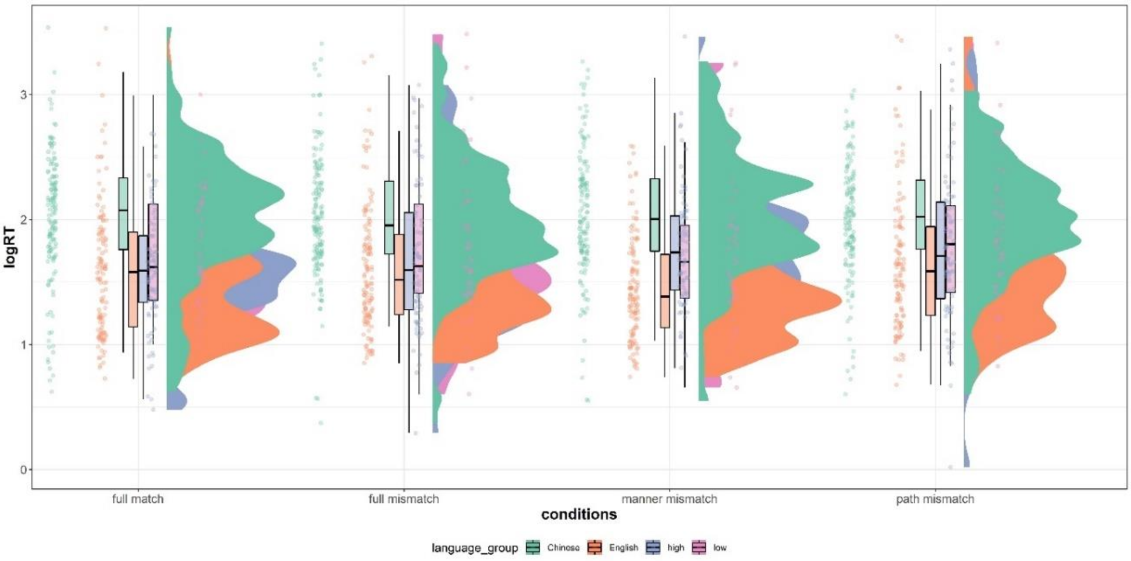
In Table 4.8, L2 learners who use more written L2 English in daily life exhibited a larger difference (597 ms) between the *manner* and *path mismatch* conditions compared to those who wrote less (287 ms). To directly examine the effects of language-specific differences, the *manner mismatch* condition and L2 learners with *LF of writing L2 English* were established as reference levels in a maximally converged LMMs (Appendix 5: LMM model 26). The results revealed that L2 learners who used less written English in daily life were significantly different from the English (β = -0.27, *SE* = 0.11, *t* = -2.42, *p* = .02, raw RT β = -763 ms) and Chinese speakers (β = 0.30, *SE* = 0.11, *t* = 2.77, *p* = .006, raw RT β = 741 ms), but not significantly different from the L2 learners who used relatively more L2 written English in daily life (β = 0.04, *SE* = 0.12, *t* = 0.30, *p* = .76, raw RT β = 961 ms).

By comparing the full model with three reduced models (Appendix 5: LMM model 27, 28, 29), a main effect of *language group* emerged ($\chi 2$ (12) = 62.16, $p < .001$), suggesting that the L2 learners who used relatively less written L2 English in daily life were significantly different from the other language groups overall. Furthermore, a significant main effect of *condition* ($\chi 2$ (12) = 22.71, *p* = .03) and a significant main effect of interaction between the *language group* and *condition* ($\chi 2$ (9) = 21.44, *p* = .01) were also observed. These two results demonstrate that the time taken in the *manner mismatch* condition differed overall from the other three conditions across four language groups, and the difference between conditions may vary across different language groups.

To further investigate the differences between conditions within each L2 group, a post-hoc test was conducted through using the *emmeans* package in R (Lenth, 2021), which displayed all possible comparisons for each L2 group. L2 learners who used L2 written English relatively more exhibited a marginally significant difference to detect between stimuli in the *full match* and those in the *path mismatch* condition (β = -0.16, *SE* = 0.09, *t* = -1.86, *p* = .07, raw RT β = -852 ms). No other significant differences were observed in any of the comparisons (detailed results are presented in Table 4.6).

In Figure 4.7, the L2 learners used more written L2 English in daily life that appear to resemble English speakers in the *full match*, *full mismatch*, and *path mismatch* condition, but not in a *manner mismatch* condition. Using the same package, an additional post-hoc analysis was conducted to explore the statistical differences between language groups within each condition.

Figure 4.7 *Logged RTs in L2 learners with low and high frequency of writing L2 English, and two monolingual groups obtained from four conditions from the b-CFS experiment*



Note: high = L2 learners who wrote more L2 English, low = L2 learners who wrote less L2 English. Box plots show the medians, and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

In the *manner mismatch* conditions, both L2 groups were significantly faster than Chinese speakers (low: β = -0.30, SE = 0.11, t = -2.74, p = .01, raw RT β = -741 ms; high: β = -0.27, SE = 0.09, t = -3.03, p = .003, raw RT β = -763 ms) and significantly slower than English speakers (low: β = 0.27, SE = 0.11, t = 2.40, p = .02, raw RT β = 763 ms; high: β = 0.31, SE = 0.09, t = 3.39, p = .001, raw RT β = 733 ms). In the *path mismatch* condition, the L2 learners who used less written L2 English were comparable to the Chinese speakers (β = -0.16, SE = 0.11, t = -1.49, p = .14, raw RT β = -852 ms) and also comparable to the English speakers (β = 0.19, SE = 0.11, t = 1.73, p = .09, raw RT β = 827 ms), while those used more L2 written English were comparable to English speakers (β = 0.13, SE = 0.09, t = 1.48, p = .14, raw RT β = 878 ms) but significantly faster than Chinese speakers (β = -0.22, SE = 0.09, t = -2.47, p = .01, raw RT β = -803 ms). In the *full match* condition, both L2 groups were significantly faster than Chinese speakers (low: β = -0.28, *SE* = 0.11, *t* = -2.63, *p* = .01, raw RT β = -756 ms; high: β = -0.39, *SE* = 0.09, *t* = -4.37, *p* < .001, raw RT β = -677 ms) and comparable to English speakers (low: β = 0.10, *SE* = 0.11, *t* = 0.93, *p* = .35, raw RT β = 905 ms; high: β = -0.001, *SE* = 0.09, *t* = -0.01, *p* = .99, raw RT β = -999 ms), regardless of the frequency of using written L2 in daily life. In the *full mismatch* condition, both groups of L2 learners were significantly faster than the Chinese speakers (low: β = -0.22, *SE* = 0.11, *t* = -1.99, *p* = .047, raw RT β = -803 ms; high: β = -0.33, *SE* = 0.09, *t* = -3.78, *p* < .001, raw RT β = -719 ms) regardless of the frequency of writing L2 in daily life, and learners using more L2 English were comparable to the English speakers (β = 0.13, *SE* = 0.09, *t* = 1.48, *p* = .14, raw RT β = 787 ms), but those who used less L2 written English in daily life were marginally slower than the English speakers (β = 0.19, *SE* = 0.11, *t* = 1.74, *p* = .08, raw RT β = 827 ms).

4.3.2.4.4 Daily use of L2 written English discussion

The main effect of *daily written L2 language use* was not observed in the RT patterns obtained from the L2 learners between different conditions in the b-CFS experiment. Specifically, the L2 learners who used more written L2 English were not found to be significantly different from those who wrote less L2 English in daily life. However, the L2 learners who wrote more L2 English were found to exhibit longer RTs in the *path mismatch* and *full match* conditions, although this did not reach significance. Furthermore, in the *path mismatch* condition, the L2 learners who wrote more L2 English in daily life resembled English speakers, whereas those who wrote less were more similar to Chinese speakers. This indicates a cognitive transfer in L2 learners that is mediated by the frequency of language use (Pavlenko, 2011). Specifically, the more L2 learners wrote English, the more likely they would experience a cognitive shift from L1 to L2.

After illustrating the results observed in the b-CFS experiment, the findings observed in the video-video verification experiment are presented in the subsequent section.

4.4 The VV verification experiment

4.4.1 Monolingual Chinese and English speakers

4.4.1.1 Results

In line with the inclusion criteria in the b-CFS experiment, only correct answers were included in the analysis, and the cutoff of the accuracy rate was also 75% (Franken *et al*., 2011; Slivac *et al*., 2021). Inclusion criteria were identical to those in the b-CFS, except that the cutoff for extreme outliers was 10 seconds based on scatter plots of the correct answers. The accuracy rate to verify similarities and differences of video-video pairs obtained from monolingual Chinese (M = 91.23%, SD = 0.09) and English (M = 95.35%, SD = 0.05) speakers was quite high. Three Chinese and four English participants were removed due to a low accuracy rate (< 75%). Thus, a total of 73 (6%) data were removed.

Reaction times (RT) in the VV verification experiment were recorded from the onset of the target video to the time of button press. The RTs were log-transformed for a statistical analysis to reduce skewness in a nonnormal distribution, which was based on a standard procedure illustrated in a related study measuring RTs to track high-level visual processing of motion events (e.g., Sakarias, 2019).

The RTs observed in monolingual Chinese and English speakers were generally in line with the predictions. In Table 4.9, the average time taken to make decisions in the *full mismatch* condition was shorter than in the *full match* condition in both monolingual groups. Additionally, the differences in RT between control conditions (Chinese: 1962 ms, English: 1502 ms) were greater than those between the critical conditions (Chinese: 325 ms, English: 82 ms) regardless of the languages spoken. However, surprisingly, and contrary to predictions, the difference in RTs between the two critical conditions was observed to be smaller in the English speakers, rather than the Chinese speakers.

Table 4.9 *Mean and SD of the reaction time (ms) observed in the Chinese and English monolinguals under four conditions from the VV verification experiment*

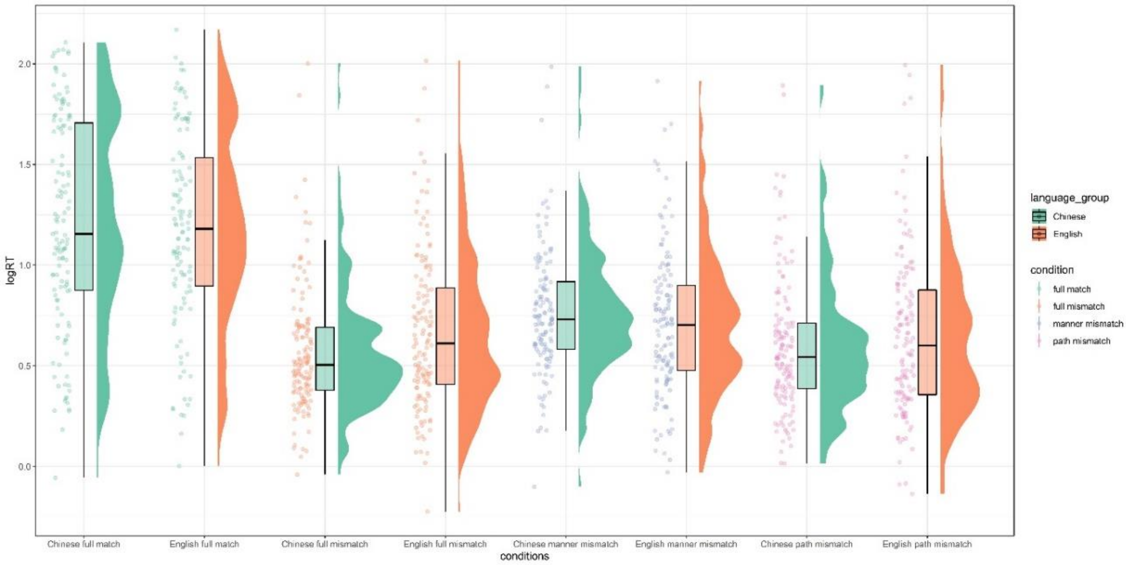
	Mean	*SD*
Chinese		
Full match	3825	1923
Manner mismatch	2265	881
Path mismatch	1940	847
Full mismatch	1863	815
English		
Full match	3644	1707
Manner mismatch	2174	921
Path mismatch	2092	1071
Full mismatch	2124	1031

To further examine the statistical differences between these two monolingual groups, linear mixed-effects regression was used through the lme4 package (Bates *et al*., 2015) in R (R Core Team, 2020). The regression models included *Language group* (English and Chinese) and *condition* (full match, manner mismatch, path mismatch, and full mismatch) as dummy-coded fixed effects, and the reference level was the *full match* condition. *Participant* and *Item* were included as random effects, with logarithmic transformed RTs used as the dependent variables. A significant main effect of *condition* was observed after comparing the full model (Appendix 5: LMM model 30) with reduced models (Appendix 5: LMM model 31, 32, 33) ($\chi2$ (6) = 55.51, p < .001), indicating RTs in *full match* were different from the overall RTs obtained in the rest of the three conditions in both language groups. No significant effect of *language group* was observed ($\chi2$ (4) = 5.77, p = .21) , and there was no interaction between *language group* and *condition* ($\chi2$ (3) = 5.75, p = .12) .

To further explore the statistical differences between conditions within each language group, a post-hoc test was conducted using the *emmeans* package in R (Lenth, 2021). This analysis provided comprehensive comparisons for both language groups. In the English-speaking group, reactions time in the *manner mismatch* condition was not significantly longer compared to those used under the *path mismatch* condition (β = 0.05, *SE* = 0.07, *t* = 0.74, *p* = .46, raw RT β = 951 ms). In contrast, in the Chinese-speaking group, a significant difference was observed between the time taken in the mismatch of the *manner mismatch* and *path mismatch* conditions (β = 0.18, *SE* = 0.09, *t* = 2.02, *p* = .05, raw RT β = 835 ms). Furthermore, in both language groups, RTs in the *full match* condition were significantly longer compared to all three conditions. Specifically, in the English speakers, the RTs in *full match* condition were significantly faster than those in the *full mismatch* (β = 0.53, *SE* = 0.09, *t* = 5.97, $p < .001$, raw RT β = 589 ms), *manner mismatch* (β = 0.50, *SE* = 0.09, *t* = 5.63, $p < .001$, raw RT β = 607 ms), and *path mismatch* (β = 0.55, *SE* = 0.09, *t* = 6.12, $p < .001$, raw RT β = 757 ms) conditions, respectively. In the Chinese speakers, RTs in the *full match* condition were also significantly slower than those in *full mismatch* (β = 0.69, *SE* = 0.11, *t* = 6.38, $p < .001$, raw RT β = 502 ms), *manner mismatch* (β = 0.48, *SE* = 0.11, *t* = 4.48, $p < .001$, raw RT β = 619 ms), and *path mismatch* (β = 0.66, *SE* = 0.11, *t* = 6.16, $p < .001$, raw RT β = 517 ms) conditions, respectively.

In Figure 4.8, it can be seen that the Chinese speakers appear to react faster than the English speakers in the *full match* and *path mismatch* conditions, and to examine the statistical differences between two language groups under each condition, a similar post-hoc analysis was performed with an identical package. The Chinese and English speakers were not significantly different from each other under any of the conditions, including *full match* (β = 0.05, *SE* = 0.10, *t* = 0.58, *p* = .56, raw RT β = 951 ms), *full mismatch* (β = -0.10, *SE* = 0.07, *t* = -1.38, *p* = .17, raw RT β = -905 ms), *manner mismatch* (β = 0.08, *SE* = 0.07, *t* = 1.11, *p* = .27, raw RT β = 923 ms), and *path mismatch* (β = -0.06, *SE* = 0.07, *t* = -0.82, *p* = .42, raw RT β = -942 ms).

Figure 4.8 *Logged RTs from the VV verification experiment in monolingual Chinese and English groups*



4.4.1.2 Discussion

The video-video verification times indicate that language-specific differences between Chinese and English influence the higher-level cognitive patterns of processing *manner* and *path* components in caused motion events, and such effects emerged even with no overt linguistic labels. This is in line with the language-specific evidence reported in Montero-Melis & Bylund (2017), where their Spanish and Swedish participants exhibited distinct cognitive patterns in judging target stimuli after freely viewing the source stimuli. Moreover, the distinct verification results obtained from Chinese and English speakers illustrate a complex picture with respect to the processing *manner* and *path components* in motion event cognition. Specifically, these RT patterns did not correspond to motion features in satellite and equipollently-framed, i.e., bigger differences between *manner* and *path mismatch* conditions in English than Chinese speakers (Ji, 2017). In contrast, the gap between verifying stimuli in the same *manner* and *path* was bigger in Chinese speakers than in English speakers.

One possible reason for the significant asymmetric differences between the *manner* and *path* components in Chinese was the mixed influences of general event perception and linguistic-specific motion event effects. Specifically, the language-specific effects on memory retrieval of motion events were covered by the pervasive spatial influence in motion cognition (Radvansky & Zacks, 2014), resulting in absent cross-linguistic differences in non-verbal experiments examining motion memory (Papafragou *et al*., 2002; Genarri *et al*., 2002). Therefore, even though among the three mismatch conditions, both the Chinese and English speakers spent relatively more time in verifying stimuli with different *manner* than those with different paths due to the universal saliency of *path* in motion cognition (Talmy, 2000). The English speakers reacted relatively faster to identify video pairs with different *manner* components than Chinese speakers due to high saliency in English motion (Slobin, 2006; Feist & Ferez, 2013), consistent with those observed in the b-CFS experiment. Such a *manner* advantage potentially diminishes the significant gap between the processing *manner* and *path* salient components in the current VV verification experiment.

Furthermore, in contrast to the processing patterns observed in the b-CFS experiment, the current VV verification experiment captures higher-level cognitive processes, which arguably emerged at a later stage (Trueswell & Papafragou, 2010). It is reasonable to expect the longest RTs in the *full match* condition in both Chinese and English speakers, as participants were instructed to verify similarities or differences between the source and target stimuli, and according to the situation model theory proposed in Zwaan & Radvansky (1998), sameness is judged after having to detect all the possible differences. Instead of pressing the "No" button rapidly after noticing one potential difference, judging the sameness therefore requires a longer time period to exclude all the possible differences.

To further examine whether the current weak language-specific effects observed in Chinese and English speakers were due to the pervasive influenced by universal saliency of *path* component in motion cognition, a sentence-video verification experiment was conducted. Through replacing the prime videos by English and Chinese sentences, language-specific encoding of *manner* and *path* in Chinese and English may be highlighted, thus suppressing the bias from the universal *path* saliency in motion cognition.

4.4.2 Chinese learners of L2 English

4.4.2.1 Overall results

Following the analysis procedure of RTs obtained from the L2 learners in the b-CFS experiment, L2 learners in the current VV verification experiment were also compared with two monolingual (i.e., English and Chinese) groups to identify potential conceptual transfer, after this, potential factors (such as *daily frequency of L2 use*, *language context*, *onset of L2 acquisition*, *L2 proficiency*, and *length of time spent in an L2-speaking country*) that may influence such conceptual transfer (Pavlenko, 2011) are analysed. Detailed analyses and results are explained further in the following sections.

The inclusion criteria were identical to those used in the monolingual English and Chinese speakers explained in previous sections regarding the VV verification experiment. Specifically, the average accuracy rate in L2 learners recruited in York was 93.75% (*SD* = 0.05). Three participants were eliminated due to the low accuracy rate (< 75%) in verifying the correct video-video pair in the VV verification experiment. Two extreme data (>10 s) were removed based on the scatter plot. Thus, a total of 44 (7%) data were removed.

Table 4.10 *Mean and SD of the reaction time (ms) observed in L2 learners under four conditions from the VV verification experiment*

	Mean	*SD*
Full match	3627	1739
Manner mismatch	2240	976
Path mismatch	1953	858
Full mismatch	2048	1025

Table 4.10 shows that the average RTs observed in the L2 learners mirror those found in the monolingual groups. Specifically, the L2 learners exhibited a greater difference in RT between the two control (1579 ms) and two critical conditions (287 ms). To further examine the statistical differences between two monolingual groups and the L2 group, linear mixed effect regression was constructed using the lme4 package (Bates *et al*., 2015) in R (R Core Team, 2020). In the maximally converged regression

model (Appendix 5: LMM model 34), *condition* (full match, full mismatch, manner mismatch, and path mismatch) and *Language group* (monolingual English, Chinese speakers, and Chinese learners of L2 English) were dummy coded as fixed effects. *Full match* condition and *L2 learners* were the reference levels, and *participant* and *item* were the random effects.

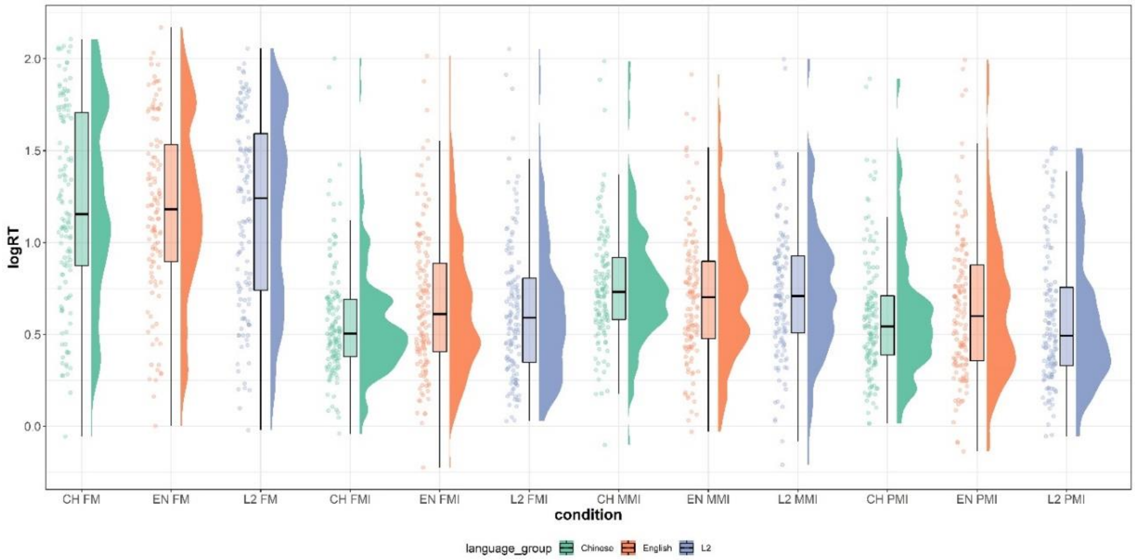
By comparing the full model with three reduced models (Appendix 5: LMM model 35, 36, 37), a significant main effect of *condition* emerged ($\chi 2$ (9) = 66.65, $p < .001$), demonstrating that the *full match* condition was overall significantly different from the other three conditions (i.e., *full mismatch*, *manner mismatch*, and *path mismatch*); no significant interaction was found between the *language group* and *condition* ($\chi 2$ (6) = 7.53, $p = .27$), nor was there a significant main effect of *language group* ($\chi 2$ (8) = 7.67, $p = .47$). This demonstrates that the L2 learners were overall similar to the two monolingual groups, and RTs in the *full match* condition did not interact with other conditions between any of the two language groups.

To further examine the statistical differences between conditions within the L2 group, a post-hoc test was performed using the *emmeans* package in R (Lenth, 2021). RTs in the *full match* condition were observed to be significantly slower than in all the other three conditions, including *full mismatch* ($\beta = 0.54$, $SE = 0.08$, $t = 6.76$, $p < .001$, raw RT $\beta = 583$ ms), *manner mismatch* ($\beta = 0.43$, $SE = 0.08$, $t = 5.36$, $p < .001$, raw RT $\beta = 651$ ms) and *path mismatch* ($\beta = 0.58$, $SE = 0.08$, $t = 7.24$, $p < .001$, raw RT $\beta = 560$ ms). Additionally, the differences in RTs taken to process stimuli between the *manner* and *path mismatch* conditions were also found marginally significant ($\beta = 0.14$, $SE = 0.08$, $t = 1.86$, $p = .07$, raw RT $\beta = 869$ ms).

In Figure 4.9, the L2 learners appears to resemble the English speakers in the *full match* and *manner mismatch* conditions but share similar patterns with the Chinese speakers in the *full mismatch* and *path mismatc*h conditions. To examine the statistical differences between the language groups within each condition, the same package was used to perform a post-hoc analysis in R. However, no significant differences were observed in the results. The L2 leaners were not significantly different from either the

Chinese or English speakers, including *full match* (Chinese: β = -0.07, SE = 0.07, t = -0.99, p = .32, raw RT β = -932 ms; English: β = -0.01, SE = 0.07, t = -0.20, p = .84, raw RT β = -990 ms), *full mismatch* (Chinese: β = 0.06, SE = 0.07, t= 0.94, p = .35, raw RT β = 942 ms; English: β = -0.03, SE = 0.06, t = 0.52, p = .60, raw RT β = -970 ms), *manner mismatch* (Chinese: β = -0.04, SE = 0.07, t = -0.54, p = .59, raw RT β = -961ms; English: β = 0.04, SE = 0.07, t = 0.54, p = .59, raw RT β = 961 ms), and *path mismatch* (Chinese: β = -0.00, SE = 0.07, t = -0.58, p = 1.00, raw RT β = 1000 ms; English: β = -0.06, SE = 0.06, t = -0.91, p = .36, raw RT β = -942 ms).

Figure 4.9 *Logged RTs from the VV verification experiment in monolingual groups (English and Chinese) and L2 group*



Note: CH = Chinese speakers, EN = English speakers, L2 = Chinese learners of L2 English, FM = full match, FMI = full mismatch, MMI = manner mismatch, PMI = path mismatch. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

4.4.2.2 Overall discussion

In line with the findings obtained in the two monolingual groups, the L2 learners appeared to be affected by the pervasive bias of the universal salience of the *path* component in the cognition of motion events. If so, this explains why they also spent the longest time to verify video pairs with identical *manner* and *path*. In terms of the differences between the *manner* and the *path* encoding within the L2 group, these participants were not significantly slower to verify stimuli with identical path than those with identical manner. This L2-like pattern indicates a cognitive shift from an L1 to an L2-based pattern (Bylund & Athanasopoulos, 2014b). This is not consistent with evidence obtained in Ji (2017), who only observed a pattern of L2-like RTs in advanced, but not intermediate L2 learners, while according to the mean IELTS scores provided by L2 learners in the current study, most were classified as intermediate (more details see **Section 4.4.2.4**).

4.4.2.3 Predictor for the cognitive outcomes in L2 learners

This section provides analysis of factors that potentially affect the conceptual transfer observed in L2 learners in the VV verification experiment. These predictors include language context, time in the UK, daily speaking of English, daily writing of English, onset of learning English as L2, and L2 proficiency. Detailed analysis and results are presented in the following.

4.4.2.3.1 Language context results

Table 4.11 *Mean and SD of the reaction time (ms) observed in the L2 learners in separate language contexts under four conditions from the VV verification experiment*

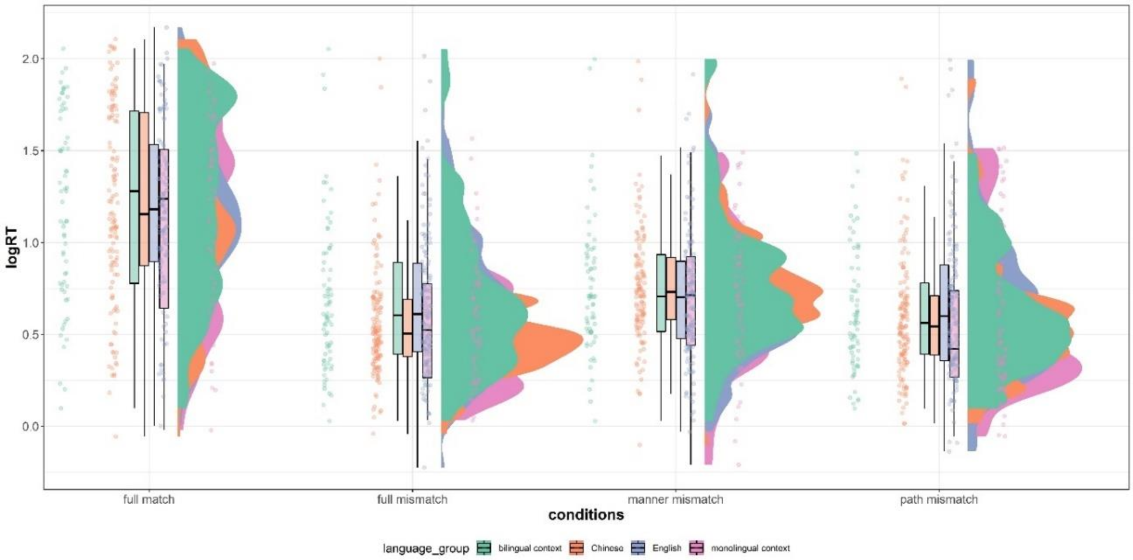
Language group	*mean*	*SD*
Bilingual context		
Full match	3765	1782
Manner mismatch	2300	1099
Path mismatch	1953	703
Full mismatch	2197	1210
Monolingual context		
Full match	3474	1693
Manner mismatch	2174	826
Path mismatch	1953	1000
full mismatch	1883	745

Table 4.11 shows that although the L2 learners in both groups spent longer in the *manner* than the *path mismatch* condition, the difference between these two

conditions was found to be bigger in the L2 learners tested in the bilingual context (374 ms), compared to those tested in monolingual context (221 ms). To directly examine these statistical differences, the *manner mismatch* condition and *L2 learners* tested in *monolingual context* were set as reference levels in a maximally converged LMM (Appendix 5: LMM model 38). A significant main effect of *condition* emerged comparing the full model with three reduced models (Appendix 5: LMM model 39, 40, 41) ($\chi2$ (12) = 60.83, $p < .001$) after, suggesting that the RTs taken in the *manner mismatch* condition overall significantly differed from the other three conditions; while no significant main effect of the *language group* ($\chi2$ (12) = 8.85, $p = .72$) or significant interaction between *language group* and *condition* was found ($\chi2$ (9) = 7.51, $p = .58$), indicating that the L2 learners were not overall significantly different from the two monolingual groups, and the *manner mismatch* did not interact significantly with other conditions between any two language groups.

To further investigate statistical differences between conditions within each L2 group, a post-hoc test was performed using the *emmeans* package in R (Lenth, 2021), which showed all possible comparisons for each L2 group. L2 learners in both groups shared similar patterns in RTs when verifying the video-video pairs. Specifically, RTs in the *manner mismatch* were significantly shorter than in the *full match* condition (bilingual: β = -0.40, *SE* = 0.11, t = -3.65, $p < .001$, raw RT β = -690 ms; monolingual: β = -0.43, *SE* = 0.11, t = -3.80, $p < .001$, raw RT β = -651 ms) and longer than those in the *path mismatch* condition (bilingual: β = 0.15, *SE* = 0.08, t = 1.83, $p = .07$, raw RT β = 861 ms; monolingual: β = 0.15, *SE* = 0.08, t = 1.71, $p = .09$, raw RT β = 861 ms), but they only reach marginal significance. Furthermore, both L2 groups took significantly more time in the *full match* in comparison to all other two conditions, including *full mismatch* (bilingual: β = 0.49, *SE* = 0.11, t = 4.39, $p < .001$, raw RT β = 613 ms; monolingual: β = 0.57, *SE* = 0.12, t = 4.95, $p < .001$, raw RT β = 566 ms) and *path mismatch* conditions (bilingual: β = 0.55, *SE* = 0.11, t = 5.01, $p < .001$, raw RT β = 577 ms; monolingual: β = 0.58, *SE* = 0.11, t = 5.06, $p < .001$, raw RT β = 560 ms).

Figure 4.10 *Logged RTs from the VV verification experiment in monolingual groups (English and Chinese) and L2 groups tested in separate languages*



Note: The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

As shown in Figure 4.10, L2 learners tested in the bilingual context seem to differ more from those tested in the monolingual context in the *path mismatch* condition compared to the other three conditions. Using the same package, an additional post-hoc analysis was conducted for exploring the statistical differences between language groups within each condition. No significant effects emerged between any possible comparisons.

4.4.2.3.2 Language context discussion

No significant main effect of language context on cognitive transfer in the L2 learners was observed in the VV verification experiment. Specifically, the L2 learners resembled the English speakers regardless of the languages used before the non-verbal experiment. This is goes against the prediction, which assumed that participants would be influenced by immediate language use in a high-level categorical experiment. One possible explanation is that current the L2 learners were those who had acquired the L2 at an early age (Lai *et al*., 2014; Wang & Li., 2011b); and, according to Table 4.13, the average AO of L2 learners in the current study was 8.23, suggesting the L2 learners in the current VV verification experiment were early bilinguals. For instance, in Wang &

Li (2021b), their bilinguals were early bilinguals, and who did not exhibit different processing patterns even though they were examined in different language contexts. The L2 learners in the present experiment shouldnot be able to switch between their L1 and L2 based on the languages tested, especially given that, both resembled the monolingual English speakers which did not suggest a lack of L2 proficiency. An alternative interpretation of this absent effects of language context might be that the language-specific effects observed in the current VV verification experiments is strong and profound, not affected by the immediate language use, as predicted by the LRH.

4.4.2.4 Other predictors for the cognitive outcomes in L2 learners

Table 4.12 illustrates the *mean* and *SD* for each potential predictor that may influence the performance of L2 in the VV verification experiment, i.e., RTs obtained to verify the (mis)match between two videos. To directly explore the influence of these factors, using the same method, a maximally LMMs (Appendix 5: LMM model 42) was converged which included all these factors as fixed effects, with the *manner mismatch* condition set as the reference level.

Table 4.12 *Mean and SD of potential predictors recorded in L2 learners from the VV verification experiment*

Measure	*Mean* (*SD*)
Time in the UK (months)	5.73 (4.01)
Daily speaking of English (%)	31.04 (18.51)
Daily writing of English (%)	45.42 (25.16)
Onset of learning English as L2	8.23 (2.27)
Oxford Placement Test score (maximum 100)	63.44 (10.15)
IELTS score	6.41 (0.59)

Results of the full model are presented in Table 4.12. None of the predictors were discovered to significantly affect the RTs obtained from the L2 learners. Therefore, no further analysis regarding these predictors was conducted.

Table 4.13 *Statistical results of potential predictors recorded in L2 learners from the VV verification experiment*

Factor	*Estimate*	*SE*	*t*	*p*
Time in the UK (months)	0.01	0.01	0.80	0.43
Daily speaking of English (%)	-0.003	0.02	-1.16	0.25
Daily writing of English (%)	0.002	0.002	1.03	0.31
Onset of learning English as L2 (AO)	-0.0001	0.02	-0.01	0.99
Oxford Placement Test score (maximum 100)	0.002	0.005	0.49	0.63
IELTS score	-0.05	0.08	-0.60	0.55

To further examine the effects of interaction between *condition* and each predictor, individual models (e.g., *logRT ~ 1 + condition * IELTS + (1 | participant) + (1 | item)*) with each of the predictor as fixed effects were constructed and converged, and differences in RTs between the *manner mismatch* and *full mismatch* conditions were found to exhibit a significant different under the interaction of IELTS score (β = 0.13, *SE* = 0.64, *t* = 1.97, *p* = .05, raw RT β = 878 ms). To further examine this interaction, a detailed analysis of effects of IELTS score between different conditions was conducted.

4.4.2.4.1 The effects of L2 English proficiency results

To operationalise L2 English proficiency, the L2 learners were divided into two sub-groups based on the criteria established by their IELTS score. Specifically, L2 learners who scored above 6.5 on the IELTS score were classified as advanced learners (*N* = 26), while the remaining L2 learners were classified as intermediate learners (*N* = 22).

As shown in Table 4.14, both advanced and intermediate L2 learners spent less time on the *manner mismatch* than the *path mismatch* condition. To directly examine the effects of language-specific differences, the *manner mismatch* condition and *intermediate L2 learners* were set as reference levels in a maximally converged LMMs (Appendix 5: LMM model 43). By comparing the full model with three reduced models (Appendix 5: LMM model 44, 45, 46), a main significant effect of *condition* emerged

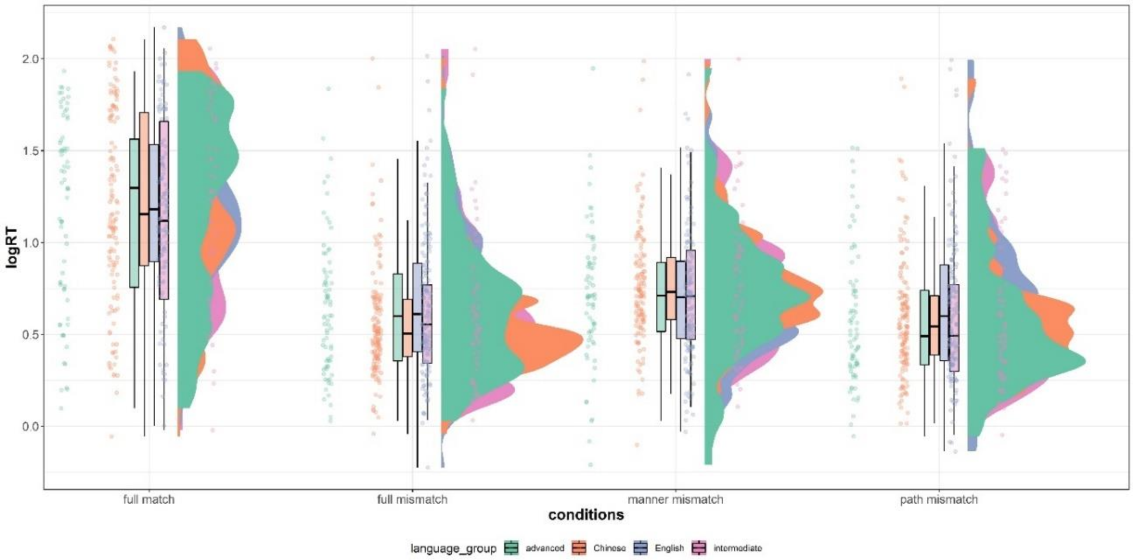
($\chi 2$ (12) = 74.87, p < .001), indicating that RTs in the *manner mismatch* condition were overall significant different from the other three conditions. Moreover, a significant interaction between *language group* and *condition* was observed ($\chi 2$ (9) = 15.17, p = .09) albeit only marginal. No main effect of *language group* was found ($\chi 2$ (12) = 15.39, p = .22), suggesting that the L2 intermediate learners performed overall similarly to the other three language groups.

Table 4.14 *Mean and SD of the reaction time (ms) used by upper and lower intermediate L2 speakers under four conditions from the VV verification experiment*

	Mean	*SD*
Advanced L2 learners		
Full match	3645	1666
Manner mismatch	2188	918
Path mismatch	1920	831
Full mismatch	2046	914
Intermediate L2 learners		
Full match	3606	1836
Manner mismatch	2299	1042
Path mismatch	1990	892
Full mismatch	2049	1146

To further investigate the differences between conditions within each L2 group, a post-hoc test was conducted through using the *emmeans* package in R (Lenth, 2021), which displayed all possible comparisons for each L2 group. In intermediate L2 learners, RTs between the *manner* and *path mismatch* conditions were significantly different (β = 0.17, *SE* = 0.08, t = 2.02 p = .05, raw RT β = 844 ms). No significant difference between *manner* and *path mismatch* was found in the advanced L2 learners (β = 0.12, *SE* = 0.08, t = 1.54, p = .13, raw RT β = 887 ms). Learners in both L2 groups were found to react significantly more slowly in the *full match* condition compared to the other three condition, including *full mismatch* (intermediate: β = 0.53, *SE* = 0.08, t = 6.32, p < .001, raw RT β = 589 ms; advanced: β = -0.55, *SE* = 0.08, t = 6.77, p < .001, raw RT β = -577 ms), *manner mismatch* (intermediate: $\beta = 0.38$, $SE = 0.08$, $t = 4.52$, $p < .001$, raw RT β = 684 ms; advanced: $\beta = 0.48$, $SE = 0.08$, $t = 5.87$, $p < .001$, raw RT β = 619 ms), and *path mismatch* conditions (intermediate: $\beta = 0.55$, $SE = 0.08$, $t = 6.57$, $p < .001$, raw RT β = 577 ms; advanced: $\beta = 0.60$, $SE = 0.08$, $t = 7.46$, $p < .001$, raw RT β = 549 ms).

Figure 4. 11 *Logged RTs from the VV verification experiment in monolingual groups (English and Chinese), intermediate and advanced L2 learners*



Note: Box plots show the medians, and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

As shown in Figure 4.11, advanced and intermediate L2 learners showed similar patterns in the *manner mismatch, path mismatch* conditions, but performed relatively differently in the *full match* and *full mismatch* conditions. Using the same package, an additional post-hoc analysis was conducted to explore the differences between the language groups within each condition. No significant differences were observed between any possible in this analysis.

4.4.2.4.2 The effects of L2 English proficiency discussion

L2 proficiency was found to affect the degrees of cognitive transfer in L2 learners in the VV verification experiment. Verification times in advanced L2 learners and intermediate L2 learners resembled those of the monolingual English and Chinese speakers, respectively, even though the main significantly effect of L2 proficiency

(IELTS score) was not found in the L2 learners. Specifically, intermediate L2 learners took significantly more time to verify stimuli with an alternate *manner* component compared to those with an alternate *path*, whereas although a similar pattern was numerically obtained in the advanced L2 learners, it did not reach statistical significance. This is in line with the evidence observed in Ji (2017), who also found that only advanced L2 learners patterned akin to the English speakers.

4.5 The SV verification experiment

4.5.1 Monolingual English and Chinese speakers

4.5.1.1 Results

In line with the inclusion criteria in the b-CFS and VV verification experiments, only correct answers were included in the analysis, and the cutoff of the accuracy rate was also 75% (Franken *et al*., 2011; Slivac *et al*., 2021) except that the cut-off for extreme outliers was 15 seconds based on scatter plots of the correct answers. Both Chinese (*M* = 90.36%, *SD* = 0.08) and English (*M* = 88.13%, *SD* = 0.05) reached a very high accuracy rate in verifying the similarities and differences of sentence-videos pairs. The cut-off point of the accuracy rate was 75% (Franken *et al*., 2011; Slivac *et al*., 2021). Two English monolinguals were removed, one had a low accuracy rate, and the other reported to fluently know a second language. Thus, a total of 78 (6%) data was removed.

Reaction times (RTs) in the present SV experiment were recorded from video onset to the time of button press. The RTs were log-transformed for a statistical analysis to reduce skewness in a nonnormal distribution, which was based on a standard procedure illustrated in a related study measuring RTs to track high-level visual processing of motion events (e.g., Sakarias, 2019).

The results were consistent with the predictions. In Table 4.15, both the Chinese and the English speakers took less time in the *full mismatch* than the *full match* conditions when the manipulation was not related to language-specific features. In contrast, the differences in average RTs between the *manner mismatch* and *path* *mismatch* conditions were greater in the monolingual English (731 ms) than in the Chinese group (150 ms).

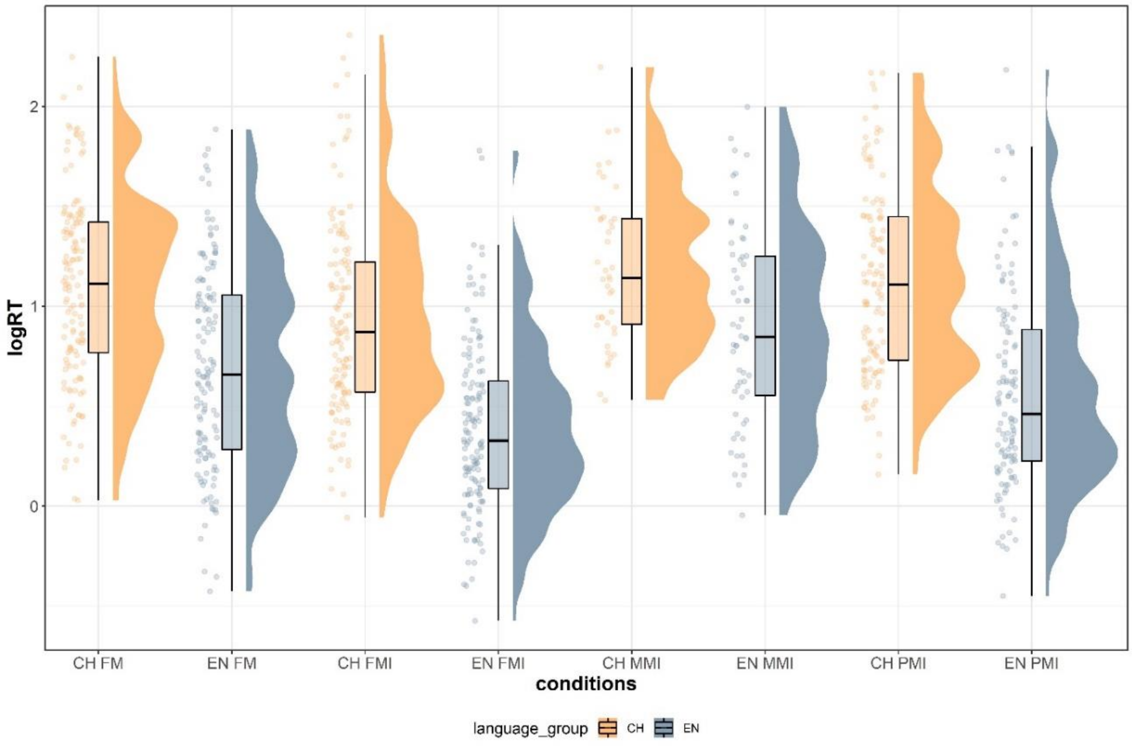
Table 4.15 *Mean and SD of the reaction time (ms) used by Chinese and English monolinguals under four conditions from the SV verification experiment*

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

Moreover, Figure 4.12 illustrates that the Chinese speakers exhibited similar RTs in their detection of stimuli in the *manner* and *path mismatch* conditions, while the English speakers had longer RTs in the *manner mismatch* compared to the *path mismatch* condition. To statistically examine the effects of *Language Group*, *Condition*, and their *Interaction*, linear mixed effects regression via the lme4 package (Bates *et al*., 2015) in R (R Core Team, 2020). The regression models included *Language group* (English and Chinese) and *condition* (full match, manner mismatch, path mismatch, and full mismatch) as dummy-coded fixed effects, and the reference level was *full match* condition. *Participant* and *Item* were included as random effects, with logarithmic transformed RTs utilized as the dependent variables.

Two significant main effects emerged after comparing the full model (Appendix 5: LMM model 47) with three reduced models(Appendix 5: LMM model 48, 49, 50), including a significant main effect of *Language group* ($\chi2$ (4) = 44.89, p < .001), indicating that the monolingual Chinese speakers were overall significantly different from the monolingual English speakers, and a significant main effect of the *condition* ($\chi2$ (6) = 19.36, p = .004), demonstrating that the RTs in the *full match* condition were different from the overall RTs obtained in the other three conditions. There was no significant interaction between the *language group* and the *condition* ($\chi2$ (6) = 19.36, p = .004), suggesting that the difference in RTs between conditions do not vary depending on the different language groups.

Figure 4.12 *Logged RTs from the SV verification experiment in monolingual Chinese and English groups*



Note: CH = Chinese speakers, EN = English speakers, FM = full match, FMI = full mismatch, MMI = manner mismatch, PMI = path mismatch. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

To further explore the specific differences between conditions within each language group, a post-hoc test was conducted using the *emmeans* package in R (Lenth, 2021). This analysis provided comprehensive comparisons for both language groups. In the English-speaking group, RTs in the *manner mismatch* condition were significantly longer compared to those used in the *path mismatch* condition (β = 0.26, SE = 0.13, t = 2.08, p = .04, raw RT β = 771 ms). In contrast, in the Chinese-speaking group, no significant differences were observed between the time taken in the *manner mismatch* and the *path mismatch* conditions (β = 0.06, *SE* = 0.10, *t* = 0.64, *p* = .52, raw RT β = 942 ms). Furthermore, in both language groups, RTs in the *full mismatch* (FMI) condition were significantly shorter compared to the other three conditions. Specifically, in the English speakers, the RTs in *full mismatch* condition were significantly faster than those in the *full match* (FM) (β = -0.32, *SE* = 0.11, *t* = -2.91, *p* = .01, raw RT β = -726 ms), *manner mismatch* (MMI) (β = -0.49, *SE* = 0.12, *t* = -4.00, *p* = .0002, raw RT β = -613 ms) and *path mismatch* (PMI) conditions (β = -0.23, *SE* = 0.11, *t* = -2.13, *p* = .04, raw RT β = -795 ms). In the Chinese speakers, RTs in the *full mismatch* condition were also significantly faster than those in *full match* (FM) (β = -0.18, *SE* = 0.08, *t* = -2.19, *p* = .03, raw RT β = -835 ms), *manner mismatch* (MMI) (β = -0.28, *SE* = 0.10, *t* = -2.81, *p* = .001, raw RT β = -756 ms) and *path mismatch* (PMI) conditions (β = -0.21, *SE* = 0.08, *t* = -2.66, *p* = .01, raw RT β = -811 ms).

4.5.1.2 Discussion

The sentence-video verification times revealed language-specific effects on motion event cognition between monolingual Chinese and English speakers. Specifically, English speakers spent significantly more time verifying stimuli with different *manner* compared to those with different *path* components in the caused motion. In contrast, the Chinese speakers exhibited similar amounts of time for processing these two components. This is in line with the results of reaction times obtained in Ji (2017) and Wang & Li (2011a), despite the fact that the current experiments examined differences rather than the sameness. That is, the current verification experiment requires participants to identify a different *manner* (but same *path*, e.g., *man carrying suitcase into room* vs. *man pulling suitcase into room*) whereas those previous studies required participants to select an alternate with same *manner* (but different *manner*, e.g., *man carrying suitcase into room* vs. *man carrying suitcase out of room*).

As explained in the results in the VV experiment, *manner* encoding is relatively more salient in English, and the amplification of *manner* through a sentence prime resulted in a quicker assessment of a different *path*, which was owe to the fact that the target stimuli have same *manner* component when the task required to judge based on similarity. The cross-linguistic disparity in the speed of verifying stimuli with matched *manner* being faster in English than in Mandarin speakers aligns with previous research investigating high-level processing of caused motion event (Ji, 2017). This effect can be attributed to the rapid and linguistically induced up-regulation of *manner* labels, which facilitates the high-level cognitive process of verifying sentence-video pairs with matched manipulation (Lupyan, 2012).

Moreover, in line with those discovered in the VV verification experiment, both English and Chinse speakers spent more time to verify sentence-video pairs with identical information; however, such effects were not stronger than the language-specific effects observed in the two critical conditions. This indicates a boost effect driven by linguistic labels (Lupyan, 2008).

4.5.2 Chinese learners of L2 English

Following the analysis of RTs obtained by L2 learners in the b-CFS and VV verification experiments. L2 learners were also compared with two monolingual groups to identify a potential conceptual transfer, followed by analyses of potential factors (such as *daily frequency of L2 use*, *language context*, *onset of L2 acquisition*, *L2 proficiency*, and *length of time spent in an L2-speaking country*) that may influence such conceptual transfer (Pavlenko, 2011). Detailed analyses will be explained further in the following sections.

4.5.2.1 Overall results

The inclusion criteria were identical to those used in the monolingual English and Chinese speakers explained in previous sections about the SV experiment. Specifically, the average accuracy rate in L2 learners recruited in York was 86.22% (*SD* = 0.07). Four participants were eliminated due to their low accuracy rates (< 75%) in verifying the correct sentence-video pair in the SV experiment. Two extremes (>15 ms) were removed based on the scatter plot. Thus, a total of 20 (3%) data were removed.

Table 4.16 illustrates the average RTs observed in L2 learners, demonstrating a pattern similar to that observed in the monolingual English speakers. Specifically, the L2 learners exhibited slower time (985 ms) in verifying the stimuli in the *manner mismatch* compared to the *path mismatch* condition.

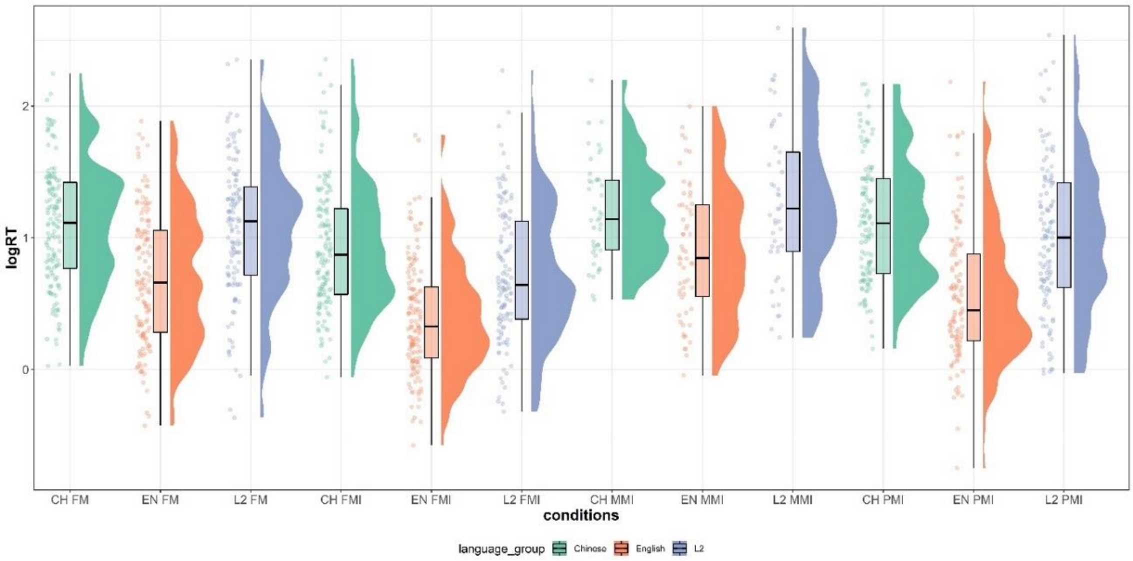
Table 4.16 *Mean and SD of the reaction time (ms) observed in L2 learners under four conditions from the SV verification experiment*

	Mean	*SD*
Full match	3294	1730
Manner mismatch	4176	2650
Path mismatch	3218	2073
Full mismatch	2347	1372

In line with the statistical analysis in the two monolingual groups, the analysis of the Chinese learners of L2 English speakers was also carried out using the linear mixed-effects regression via the package lme4 package (Bates *et al*., 2015) in R (R Core Team, 2020). In the maximally converged regression model (Appendix 5: LMM model 51), *condition* (full match, full mismatch, manner mismatch, and path mismatch) and *Language group* (monolingual English, Chinese speakers, and Chinese learners of L2 English) were dummy coded as the fixed effects. The *Manner mismatch* condition and *L2 learners* were the reference levels, and *Participant* and *Item* were the random effects.

A significant main effect of *Language group* was revealed through comparisons of the full model with three reduced models (Appendix 5: LMM model 52, 53, 54) indicating that the L2 learners were different from the other two language groups overall ($\chi 2$ (8) = 66.35, p < .001). Specifically, L2 learners were overall significantly slower than English speakers (β = -0.44, *SE* = 0.10, t = -4.42, p < .001; raw RT β = -644 ms) but comparable to Chinese speakers (β = -0.08, *SE* = 0.10, t = -0.80, p = .42; raw RT β = -923 ms) in all four conditions. Additionally, a significant main effect of *condition* was also found ($\chi 2$ (9) = 31.62, p < .001), indicating the time taken in the *manner mismatch* condition was found to be different from the other three conditions overall. Furthermore, two significant interactions were observed. These include a significant interaction between *condition* (manner mismatch vs. full match) and *Language group* (L2 vs. Chinese) (β = 0.25, *SE* = 0.10, *t* = 2.43, *p* = .02; raw RT β = 779 ms) and a significant interaction between *condition* (manner mismatch vs. path mismatch) and *Language group* (L2 vs Chinese) (β = 0.21, *SE* = 0.10, *t* = 2.00, *p* = .048; raw RT β = 811 ms).

Figure 4.13 *logged RTs recorded in monolingual speakers and L2 learners in the SV verification experiment*



Note: CH = Chinese speakers, EN = English speakers, L2 = Chinese learners of L2 English, FM = full match, FMI = full mismatch, MMI = manner mismatch, PMI = path mismatch. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

As shown in Figure 4.13, despite the fact that the overall pattern of RT observed in L2 learners was more consistent with those in the monolingual Chinese speakers, the divergence between the *manner* and *path mismatch* conditions reflects a pattern mirrored the English speakers. To further explore the statistical difference within the L2 group, a post-hoc test was conducted using the *emmeans* package in R (Lenth, 2021). RTs in the *manner mismatch* were significantly different from those in the *path mismatch* condition (β = 0.27, *SE* = 0.11, *t* = 2.48, *p* = .01, raw RT β = 763 ms). Furthermore, RTs in the *full mismatch* condition were significantly quicker than in the *full match* condition (β = -0.34, *SE* = 0.09, t = -3.63, p = .001, raw RT β = -712 ms). Furthermore, in line with the two monolingual groups, the L2 learners also reacted the fastest in the *full mismatch* condition in all four conditions, including the two critical conditions: the *manner mismatch* (β = -0.28, *SE* = 0.11, t = -2.61, p = .01, raw RT β = -756 ms), *path mismatch* conditions (β = -0.22, *SE* = 0.09, t = -2.50, p = .02, raw RT β = -803 ms).

An additional post-hoc test was conducted to further investigate the statistical differences between the L2 learners and the two monolingual groups within each condition. In the *full mismatch* condition, the L2 learners were significantly slower than English speakers (β = 0.36, *SE* = 0.07, t = 5.06, p < .001, raw RT β = 698 ms) and significantly faster than the Chinese speakers (β = -0.16, *SE* = 0.07, t = -2.30, p = .02, raw RT β = -852 ms). Whereas in the other three conditions, they were consistently significantly different from English speakers but mirrored the Chinese monolinguals. Specifically, in the *manner mismatch* condition, L2 learners were significantly slower than English speakers (β = 0.44, *SE* = 0.10, t = 4.32, p < .001, raw RT β = 644 ms), but resembled the Chinese speakers (β = 0.08, *SE* = 0.11, t = 0.78, p = .43, raw RT β = 923 ms). In the *path mismatch* condition, the L2 learners were significantly slower than the English speakers (β = 0.44, *SE* = 0.08, t = 5.61, p < .001, raw RT β = 644 ms), but comparable to the Chinese monolinguals (β = -0.12, *SE* = 0.08, t = -1.62, p = .11, raw RT β = -887 ms). In the *full match* condition, the L2 learners were also significantly slower than the English speakers (β = 0.38, *SE* = 0.08, t = 4.81, p < .001, raw RT β = 684 ms), but patterned with the Chinese speakers (β = -0.01, *SE* = 0.08, t = -0.18, p = .858, raw RT β = -990 ms).

4.5.2.2 Overall discussion

The sentence-video verification times in the L2 learners illustrate a cognitive restructuring of L1 and L2 concepts (Pavlenko, 2011). Specifically, the overall RT patterns across all four resembled the Chinese speakers, but RTs between the *manner* and *path mismatch* conditions mirrored the English speakers, thus reflecting a potential conceptual transfer to L2-like pattern in the current experiment. This is inconsistent with evidence found in earlier studies that examined higher-level processing of *manner* and *path* components, which reflected an L2-like RT pattern discovered only in advanced L2 learners (Ji, 2017). However, the current sentence-video verification experiment employed overt linguistic labels which may enhance the language-specific effects of motion in L2 learners, thus resulting in a different outcome (Lupyan, 2008). To further examine the influence of overt linguistic labels, the effects of language context are analysed below. The language contexts were monolingual context which all prime sentences were presented in Chinese, and bilingual context that half prime sentences were English (including 12 critical items and 6 filler), and the other half (18 filler) was in Chinese.

4.5.2.3 Predictor for the cognitive outcomes in L2 learners

4.5.2.3.1 Language context results

Table 4.17 *Mean and SD of the reaction time (ms) observed in L2 learners in separate language contexts under four conditions from the SV verification experiment*

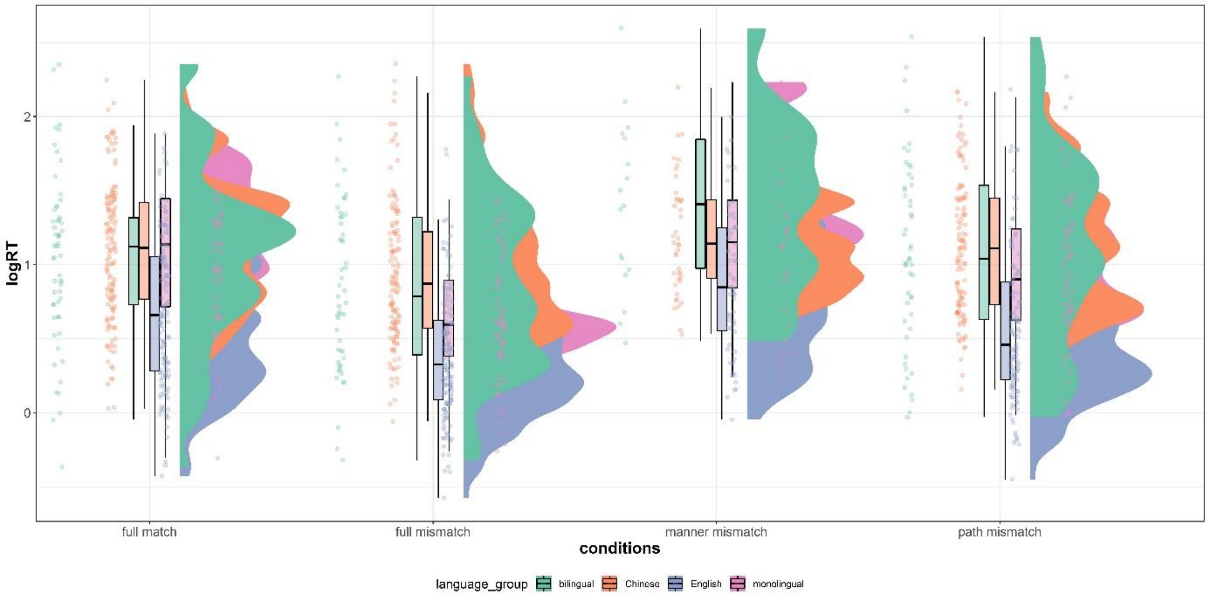
Language group	*Mean*	*SD*
Bilingual context		
Full match	3307	2019
Manner mismatch	4733	2968
Path mismatch	3610	2462
Full mismatch	2780	1756
Monolingual context		
Full match	3281	1420
Manner mismatch	3838	2431
Path mismatch	2925	1688
full mismatch	2022	873

Table 4.17 illustrates that although the L2 learners spent more time in the *manner* than the *path mismatch* conditions, the difference between the two was found to be greater (210 ms) in the L2 learners tested in the bilingual context, compared to those tested in the monolingual context. To directly examine the effects of language-specific

differences, the *manner mismatch* condition and the L2 learners tested in *monolingual context* were set as reference levels in a maximally converged LMM (Appendix: model 55). Three significant effects emerged after comparing the full model with three reduced models (Appendix: model 56, 57, 58), including a main effect of *language group* ($\chi2$ (12) = 76.72, p < .001), suggesting that the L2 learners tested in the *monolingual context* were overall different from the other three language groups; a main effect of the *condition* ($\chi2$ (12) = 42.15, p = .02), suggesting that the time taken in the *manner mismatch* condition was overall different from the those in other conditions; and a main effect of interaction between *language group* and *condition* ($\chi2$ (9) = 18.67, p = .03), demonstrating that the difference between conditions may vary between different language groups. Furthermore, the L2 learners tested in the *monolingual contex*t were found to interact significantly with *Chinese speakers* in two comparisons, including *manner mismatch* and *path mismatch* (β = 0.22, SE = 0.11, t = 2.02, p = .04, raw RT β = 803 ms), and *manner mismatch* and *full mismatch* (β = 0.29, SE = 0.11, t = 2.60, p = .01, raw RT β = 748 ms). However, no significant difference between the two L2 groups were observed regardless of the language context (β = 0.14 SE = 0.14, t = 0.95, p = .34, raw RT β = 869 ms), indicating that these two L2 groups shared similar patterns of overall RTs across all conditions.

To further investigate the differences between conditions within each L2 group, a post-hoc test was conducted through using the *emmeans* package in R (Lenth, 2021), which displayed all possible comparisons for each L2 group. In the *bilingual context*, RTs in the *manner mismatch* were significantly longer than the three remaining conditions, including the *path match* condition (β = 0.29, SE = 0.13, t = 2.16, p = .03, raw RT β = 748 ms), *full mismatch* condition (β = 0.51, SE = 0.13, t = 2.85, p = .0001, raw RT β = 600 ms), and *full match* condition (β = 0.30, SE = 0.13, t = 2.23, p = .03, raw RT β = 741 ms). Furthermore, the RTs in the *full mismatch* were also significantly faster than those in the *full match* condition (β = -0.22, SE = 0.10, t = -2.09, p = .04, raw RT β = -803 ms). In contrast, the L2 learners tested in the *monolingual context* spent significantly more time in *manner mismatch* condition, when compared to the *path* *mismatch* (β = 0.29, *SE* = 0.12, *t* = 2.48, *p* = .01, raw RT β = 748 ms) and *full mismatch* conditions (β = 0.58, *SE* = 0.12, *t* = 4.97, *p* < .001, raw RT β = 560 ms), but not the *full match* condition (β = 0.15, *SE* = 0.12, *t* = 1.25, *p* = .21, raw RT β = 861 ms). Furthermore, they were also significantly faster in the *full mismatch* than in *full match* condition (β = -0.43, *SE* = 0.10, *t* = -4.30, *p* < .001, raw RT β = -651 ms).

Figure 4.14 *Logged RTs recorded in L2 learners in a monolingual and bilingual context from the SV verification experiment*



Note: bilingual = L2 learners tested in a bilingual context, monolingual = L2 learners tested in a monolingual context. The box plots show the medians and the box demonstrates the 50% of the logged RTs. The violin and rain-cloud plots show the distribution patterns of the data in each condition per group.

In Figure 4.14, L2 learners tested in the bilingual context appear to differ from those tested in the monolingual context more in the *manner mismatch* compared to the other three conditions. Using the same package, an additional post-hoc analysis was conducted for exploring the statistical differences between language groups within each condition. In the *manner mismatch* condition, both L2 groups were significantly slower than Chinese speakers (monolingual: β = -0.20, *SE* = 0.10, *t* = -2.05, *p* = .04, raw RT β = -819 ms; bilingual: β = -0.36, *SE* = 0.10, *t* = -3.74, *p* < .001, raw RT β = -698 ms), but comparable to English speakers (monolingual: β = 0.38, *SE* = 0.10, *t* = 3.84, *p* < .001, raw RT β = 684 ms; bilingual: β = 0.21, *SE* = 0.10, *t* = 2.17, *p* = .03, raw RT β = 811 ms). In the *path mismatch* condition, the L2 learners in the monolingual context were significantly slower than the English speakers (β = 0.37, *SE* = 0.09, t = 4.10, p < .001, raw RT β = 691 ms) but faster than the Chinese speakers (β = -0.19, *SE* = 0.09, t = -2.13, p = .03, raw RT β = -827 ms), while those in the bilingual context were significantly slower than the English speakers (β = 0.50, *SE* = 0.09, t = 5.31, p < .001, raw RT β = 607 ms), but comparable to the Chinese speakers (β = -0.05, *SE* = 0.09, t = -0.56, p = .57, raw RT β = -951 ms). In the *full match* condition, both L2 groups were significantly slower than the English speakers (monolingual: β = 0.39, *SE* = 0.09, t = 4.32, p < .001, raw RT β = 677 ms; bilingual: β = 0.37, *SE* = 0.09, t = 4.09, p < .001, raw RT β = 691 ms), but comparable to the the Chinese speakers (monolingual: β = -0.01, *SE* = 0.09, t = -0.15, p = .88, raw RT β = -990 ms; bilingual: β = -0.03, *SE* = 0.09, t = -0.28, p = .78, raw RT β = -970 ms). In the *full mismatch* condition, the L2 tested in monolingual mode were significantly faster (β = -0.25, *SE* = 0.09, t = -2.86, p = .005, raw RT β = -779 ms) than the Chinese speakers, and slower than the English speakers (β = 0.27, *SE* = 0.09, t = 3.09, p = .002, raw RT β = 763 ms), while those test in the bilingual context were significantly slower than English speakers (β = 0.47, *SE* = 0.09, t = 5.04, p < .001, raw RT β = 625 ms), but comparable than Chinese speakers (β = -0.05, *SE* = 0.09, t = -0.56, p = .58, raw RT β = -951 ms).

4.5.2.3.2 Language context discussion

No significant main effect of language context was found in the L2 learners examined in the SV verification experiment, since they were both slower to verify sentence-video pairs with *manner* and *path-manipulated* alterations. This is inconsistent with the predictions. However, the L2 learners tested in the monolingual context were significantly faster than the Chinese speakers in the *path mismatch* condition, whereas those tested in the bilingual context were comparable to the Chinese speakers in the same condition. This seems contradictory to findings in the previous literature related to the effects of language context (e.g., Barner *et al*., 2009), specifically, that cognitive process should be consistent with the language-specific patterns corresponding to the language context.

However, this may in fact be understandable. Recall that both the L2 learners tested in the monolingual and the bilingual context indeed exhibited an English-like pattern in the two critical conditions. Specifically, they both spent significantly more time verifying target videos with *manner* manipulated information in comparison to those with *path* manipulated information. The reason that L2 learners tested in the bilingual mode resembled the Chinese speakers in the *path mismatch* condition may be because the temporal distance between the two critical conditions were slightly bigger (1123 ms) than those tested in the monolingual context (913 ms). Specifically, the equipollently-framed Chinese *manner* and *path* share similar salience, thus resulting in similar RTs in conditions with *manner* and *path* manipulation. Whereas, satellite-framed English has a relatively higher *manner*, thus RTs in conditions with *manner* manipulation would be different from those with *path* manipulation. This is in line with the results obtained from the pilot study, and consistent to the language-specific predictions for the overall RT patterns expected in the Chinese and English – speaking groups. Overall, this indicates that overt linguistic labels might influence the non-verbal performance, but in a limited way.

4.5.2.4 Other predictors for the cognitive outcomes in L2 learners

Table 4.18 *Mean and SD of potential predictors recorded in L2 learners from the SV verification experiment*

Measure	*Mean* (*SD*)
Time in the UK (months)	15.63 (17.56)
Daily speaking of English (%)	28.48 (18.74)
Daily writing of English (%)	59.48 (28.90)
Onset of learning English as L2	8.94 (3.07)
Oxford Placement Test score (maximum 100)	60.92 (12.07)
IELTS score	6.54 (0.71)

Table 4.18 illustrates the *mean* and *SD* for each potential predictor that may influence L2 performance, i.e., RTs obtained to verify the (mis)match between sentences and videos in SV. To directly explore the influence of these factors, using the

same method, a maximally LMMs (Appendix: model 59) was converged which included all these factors as fixed effects, with the *manner mismatch* condition set as the reference level.

The results of the full model are presented in Table 4.19. To sum up, none of the predictors were found to significantly affect the RTs obtained from L2 learners. No significant effects of interaction between the *condition* and these predictors were discovered either. Therefore, no further analysis regarding the predictors was conducted.

Table 4.19 *Statistical results of potential predictors recorded in L2 learners from the SV verification experiment*

Factor	*Estimate*	*SE*	*t*	*p*
Time in the UK (months)	-0.001	0.003	-0.45	0.66
Daily speaking of English (%)	-0.37	0.27	-1.33	0.19
Daily writing of English (%)	0.25	0.18	1.38	0.18
Onset of learning English as L2 (AO)	-0.01	0.02	-0.45	0.66
Oxford Placement Test score (maximum 100)	-0.002	0.01	-0.33	0.74
IELTS score	-0.04	0.08	-0.44	0.66

To sum up, this Chapter illustrates the detailed analysis and results of each specific experiment conducted in the current study. This including the three main experiments: main b-CFS, sentence-video verification, and the video-video verification experiments, and two supplement experiments: b-CFS pilot and b-CFS control. The general results are in line with the predictions, which is that monolingual Chinese and English speakers exhibit distinct RT patterns across all three main and b-CFS pilot experiments, and the mask was found to effectively suppress the dynamic motion stimuli. However, the language-specific differences vary across the three main experiments, which provides intriguing evidence for understanding the motion event cognition in Chinese and English. Detailed discussions regarding each specific result are presented in the following chapter.

Chapter 5: Discussion

The current study is the first exploration to provide systematic evidence in support of the linguistic relativity hypothesis (LRH, Whorf, 1956). Specifically, the current findings show that the cognitive processes of monolingual Chinese and English speakers, and Chinese learners of L2 English, are affected by the distinct way motion events are expressed/described in each language. In line with the words from Lucy (2016), 'the patterns of thinking affected include attention and perception, similarity judgements, and classification, short- and long-term memory, and learning and reasoning. And these effects are evidenced in everyday experience, specialized contexts, or ideational traditions' (p.57). This influence appears to emerge from an early stage of visual processing when motion labels are not able to be recruited (i.e., the b-CFS experiment), to the later stage when semantic analysis requires assistance from overt (i.e., the sentence-video verification experiment) or covert (i.e., the video-video verification experiment) linguistic labels.

In this chapter, the findings obtained from monolingual Chinese and English speakers in each experiment are first discussed in sequence. This is followed by discussions of the evidence observed in L2 learners. The focus turns to the specific contributions related to motion event typology. Subsequently, potential contributions beyond the effects of language on cognition are presented. Finally, several limitations of the current study are set out.

5.1 Cross-linguistic evidence in monolingual Chinese and English speakers

The findings observed in monolingual English and Chinese speakers support the idea that cross-linguistic differences in motion lexicalisation influence motion event cognition consistently (Casasanto, 2016; Lucy, 1992b; Bylund & Athanasopoulos, 2015) from an early stage of visual processing that detects low-level sensory signals to a later stage of visual processing which involves high-level semantic recognition. Such differences were also found in Chinese L2 English, demonstrating various cognitive restructuring in these three experiments. Specifically, satellite-framed English speakers exhibited distinct processing patterns compared to equipollently-framed Chinese speakers across all three non-verbal experiments: b-CFS experiment, video-video verification experiment, and sentence-video verification experiment. Moreover, the L2 learners were observed to undergo certain cognitive/ conceptual restructuring in all three experiments, which might further provide evidence to support the relativist view.

In general, Chinese and English speakers exhibited distinct processing patterns for *manner* and *path* encoding at all three levels; this aligns with previous research exploring how *manner* and *path* encodings vary between Mandarin/Cantonese and English speakers (Ji, 2017; Wang & Li, 2021). These two studies shared similar results and consistently discovered cross-linguistic differences in similarity judgment tasks. However, only evidence observed in the b-CFS, and sentence-video verification experiments reflected a similar result specifically corresponding to the findings obtained from the monolingual English speakers in the two studies mentioned above. That is, English speakers used different amounts of time to process *manner* and *path* manipulated stimuli, whereas Chinese speakers used similar amounts of time to process those two manipulated components reflected in the motion animations. The inconsistency was discovered in the current video-video verification experiment, in which variation in *manner* and *path* coding patterns was found in the Chinese-speaking group, not the expected English-speaking group.

Furthermore, the current study confirms that English and Chinese speakers encode the *manner* and *path* differently at different stages of the visual process, respectively. Recall that previous language-specific effects of motion lexicalisation were found to be highly mediated by specific experimental designs. The current study provides potential explanations for such findings and unveils the dynamic interaction between language-specific effects and universal influence on motion events. Specifically, how the English *manner* and universal *path* competed with passively increased and decreased saliency due to different encoding stages. This includes a highlighted saliency of English *manner* in the b-CFS, and SV verification experiments, and a devalued one in the VV verification experiment.

The most prominent effect of English *manner* was discovered in the b-CFS experiment, which examined the low-level detection of motion signals, thus excluding the possibility of participants' being able to use implicit or explicit language. Then, a similarly strong effect of the English *manner* was found in the SV verification experiment, which explored the high-level semantic decisions of motion components, especially with the assistance of relevant verbal cues. Finally, a dramatically devalued effect was identified in the VV verification experiment when the linguistic representation of motion was only able to be recruited implicitly (non-verbal cues).

As well as the comparison between monolingual Chinese and English speakers, the evidence discovered in the Chinese learners of the L2 English study adds an extra layer of support for the relativist view, and also further confirms that there appear to be language-specific differences of processing *manner* and *path* at different stages. In the following sections, these contributions to the current study are illustrated in more detail.

5.1.1 Low-level advantage of the English manner

The first and main contribution to the current study is reflected in the findings of the b-CFS experiment. Employing this low-level detection paradigm successfully appears to address the notorious circularity problem that researchers have noted when attempting to examine the LRH (Casasanto, 2006). The possibility of using language sub-vocally/implicitly arguably interferes with the potential to discover language-specific effects that could the LRH. This is because the strongest evidence in support of the LRH would be pervasive language-specific effects that are not only found in linguistic description, but also in tasks involving pure non-verbal/nonlinguistic cognitive processes (Lucy, 1992; Bylund & Athanasopoulos, 2015). In the current study, the low-level speed advantage observed in the English speakers in the b-CFS experiment aligns with results obtained from earlier studies utilising low-level detection methods (Athanasopoulos & Casaponsa, 2020; Kamenetski *et al*., 2022): that is the language-specific effects emerge at early stage of visual processing. The b-CFS paradigm can capture cross-linguistic differences at a lower-level processing, where utilising language is not possible (Jiang *et al*., 2007). The basic rationale is that the flickering mask in the b-CFS expands the early time window of automaticity by suppressing the detection of a presented stimuli. This, therefore, offers an ability to exclude the bias of linguistic involvement during non-verbal experiments. Consequently, cross-linguistic differences induced by specific linguistic features related to the expression of *manner* and *path* were discovered among the monolingual English and Chinese speakers, and the L2 learners without conscious language involvement, demonstrating that language-specific effects of motion events are deep and resilient, and appear to emerge as soon as the detection of an incoming perceptual signal becomes available. This aligns with the claim of Casasanto *et al*. that, ‘language can shape primitive, low-level mental processes’ (2004, p.575).

The interpretation of this finding is in line with the framework of predictive processing by Luypan & Clark, which sets out the underlying mechanism between low-level visual detection and high-level semantic salience of motion events, that is, ‘language begins to take on a surprisingly central role in cognition by providing a uniquely focused and flexible means of constructing predictions against which sensory signals can be evaluated’ (2015, p.279). In our case, the incoming visual signal of motion evens is mediated by the downward prediction formed by the variation of motion expressions in English and Chinese.

Specifically, the observed speed advantage of detecting *manner-manipulated* stimuli in monolingual English speakers reflects the greater weight of *manner* (in comparison to *path*) in English motion events; in contrast, the similar time taken to detect *manner* and *path-manipulated* stimuli found in the monolingual Chinese speakers is arguably due to the equal weight of *manner* and *path* in Chinese motion events. Without conscious linguistic involvement, English speakers detected the corners of the stimuli with mismatched *manner* information significantly faster, owing to the involvement of increased relevance of *manner* in their expectation of the incoming visual stimuli. Such facilitation of the highlighted *manner* was not observed in the Chinese speakers’ performance, whose expectations toward the *manner* and *path-based* information were formed similarly due to their equal weight in Chinese. This speed advantage of the English *manner* continued to show in the L2 learners, who also spent less time in the *manner* mismatch condition (6741 ms) than in the *path* mismatch (7234 ms) although this difference (493 ms) did not reach statistical significance.

5.1.2 High-level disadvantage of verifying the English manner

The second contribution in the current study is provided by the results of the sentence-video (SV) verification experiments, which further support the LRH by showing consistent cross-linguistic differences in high-level processing of motion events among the monolingual English and Chinese speakers, and the L2 learners under the employment of identical target stimuli (i.e., motion animations). The findings in the sentence-video verification experiment were in line with the predictions and confirmed that English and Chinese speakers processed the *manner* and *path* distinctly at a higher level: verifying sentence-video pairs.

The interpretation of this high-level finding is consistent with the proposal of situation models (Zwaan & Radvansky, 1998), as explained in Chapter 3. The general rationale is that information stored in long-term memory is formed by consistent input from the short-term memory, and the formation of the short-term memory consists of editing the current model based on the incoming information. In the current SV verification experiment, the sentences served as language cues for participants to retrieve motion-related information stored in long-term memory (i.e., equipollently-framed Chinese motion and satellite-framed English motion) which formed a current model related to motion events. The similarities and differences between target videos and prime sentences within each motion pair as blocks to update the current motion event model, in order to complete the integrated process in short-term memory. As predicted by the LRH, monolingual Chinese and English speakers stored motion events in distinct ways due to the variations of expressing *manner* and *path*, that is, a greater weight of English *manner* and equal weighted Chinese *manner* when compared with the *path* component in each language, respectively. Consequently, the English encoded stimuli pairs with *path* and *manner* variation differently, while the Chinese participants did it in similar ways.

However, unlike the speed advantage observed in the *manner mismatch* condition in the b-CFS experiment, a speed disadvantage was identified in the current SV verification experiment. This may seem atypical, but it is crucial to bear in mind that b-CFS examines low-level detection abilities, so the earlier speed advantage was induced by the ability to spot contrasting features. In contrast, the verification procedure examined in the current SV verification experiment involves high-level semantic analysis. Specifically, the facilitation effect with matching motion animation within the sentence-video pairs emerged after key motion features had been activated by the proceeding prime sentences. Since *manner* is more prominent than *path* in English, this arguably leads to the faster recognition of *manner* sameness, thus demonstrating a speed advantage under the *path mismatch* condition.

Similarly to those discussed in the results obtained in the b-CFS experiment, the L2 learners examined in the current SV verification experiment also illustrated a consistent pattern. Specifically, they spent more time in the *manner mismatch* than the *path* mismatch condition, and this temporal difference resembled that found in the monolingual English speakers.

These results are in line with the cross-linguistic differences of processing *manner* and *path* reported in Ji (2017), a study whose experimental tasks also involved a high-level semantic analysis. Specifically, their monolingual English speakers also spent less time in selecting an alternative motion animation with the same *manner* (with a different *path*) than those with the same *path* (with a different *manner*) in a similarity judgement task, while their Chinese participants showed similar time durations regardless of the *manner* or *path*. Ji (2017) argued that such results were due to the different underlying mechanisms for processing the *manner* and *path* between English and Chinese speakers, that is, English speakers process the *manner* and *path* in sequence, whereas Chinese speakers encode the two in parallel. The current study could not confirm this, as it needs to be further examined in a well-designed study that focusses on whether Chinese and English speakers encode *manner* and *path* simultaneously and sequentially, respectively. Furthermore, although the current study also examined the high-level semantic analysis of *manner* and *path* between Chinese and English speakers, the current SV verification experiment focused on the evaluation of differences, while the similarity judgement task conducted in Ji (2017) mainly studied the sameness.

Alternatively, the longer time taken to verify motion animation with manipulated information could also be consistent with the ‘reversed Whorfian effect’ proposed by Papafragou *et al*. (2008). As introduced in the Chapter 2, this study recorded the eye-movements of English (satellite-framed) and Greek (verb-framed) speakers before and during viewing motion event stimuli before engaging in a subsequent non-verbal recognition task. The predictions were that English speakers should pay more attention on the *manner* regions since *manner* is more salient than *path* in English, while the Greek speakers should focus more on the *path* regions due to the higher saliency of *path* (compared to *manner*) in Greek (Slobin, 2006). The results showed an opposite pattern between English and Greek speakers. Specifically, in the English-speaking group, a late fixation was observed more in the *path* regions when they were preparing to engage in a subsequent memory task. After more study of the *path* information, English speakers performed better than Greek speakers in recognising objects reflecting the *path* endpoint. The Greek speakers were found to observe the *manner* regions more. The authors interpreted such cross-linguistic differences as an outcome of thinking-for-speaking (Slobin, 1996), since the participants might recruit language sub-vocally to achieve better performance in the later memory task. This conclusion acknowledged the effects of language on thought but under the condition of linguistic mediation (i.e., they might use language sub-vocally to achieve better performance in the subsequent memory task). That is, such cross-linguistic differences discovered before the recognition task might be due to the silent language used during task preparation, thus contradictory to the relativistic view.

However, the current study provides additional evidence to interpret such *path*-attention effects as evidence predicted by the LRH. Specifically, according to the predictive processing account (Lupyan & Clark, 2015), the reasons for the late fixation of the *path* regions observed in the English speakers in Papafragou and colleagues' work might be due to the inspection to complete a subsequent memory task. To successfully judge the target motion animation presented in the subsequent recognition task, details of motion events need to be evaluated beforehand. Since the English *manner* is prominent rather than the *path*, English speakers needed to make more efforts to identify the specific *path*, thus fixating more on the corresponding regions.

In the current SV verification experiment, participants were given longer (self-paced) to evaluate specific details of the target motion animation, *path* manipulated stimuli were quicker to be assessed due to its less saliency in English. Specifically, to further explain this in line with the predictive processing account, English speakers verified the upcoming motion animation with expectations formed by the proceeding sentence cues. To make semantic decisions between the target motion animation (e.g., man pulling suitcase into room) and the earlier sentence cues (e.g., man carrying suitcase into room) and, detailed inspections were undertaken during observation of the target stimuli. During this process, the less salient English *path* guided participants to spend less time evaluating *path*-based stimuli (e.g., man carrying suitcase out of room). Consequently, English speakers spent longer verifying motion videos reflecting an inconsistent *manner* component.

5.1.3 High-level advantage of verifying the Chinese path

The third contribution in the current study is reflected in the results of the video-video (VV) verification experiment, which continues to provide evidence to support the LRH by presenting a relatively subtle but consistent cross-linguistic difference of high-level processing of motion events among monolingual English, and Chinese speakers, and L2 learners in non-verbal experiments without overt linguistic labels. The findings in the video-video verification experiment were consistent with the early prediction in a general way, that is, monolingual Chinese and English speakers should exhibit distinct processing patterns, although the specific differences between processing *manner* and *path* within each language group were contrary to the predictions. That is, a significant temporal difference between verifying stimuli with *manner* and *path* manipulation was found in the Chinese speakers, instead of the predicted English speakers. Following this, it is very likely that language-specific effects emerged at a higher cognitive level, even without recruiting language labels: verifying video-video pairs.

The interpretation of this finding is also consistent with the proposal of situation models (Zwaan & Radvansky, 1998). Specifically, *manner* and *path* were activated as prime cues in the first video and functioned distinctly in the subsequent evaluation procedure due to specific language-specific preferences. The specific processing patterns of *manner* and *path* within each language group continued to be consistent with the predictive processing account (Lupyan & Clark, 2015), that is, the language-specific expectations orientated the time taken to verify video-video pairs.

In terms of the results within each language group, these specific cognitive processes of verifying video-video pairs between *manner* and *path* went against those that were predicted in accordance with the motion typology in the equipollently-framed Chinese and satellite-framed English. Specifically, asymmetric reaction times taken to process the *manner* and *path* should be observed in monolingual English speakers because of the distinct weight of these two components expressed in English, but the current VV verification observed such imbalanced processing patterns in monolingual Chinese speakers. More specifically, Chinese speakers used significantly less time to verify target motion videos with different *manner*s than those with different *path*s. In contrast, even though English speakers also took longer to evaluate video pairs with different *manner*s, the comparison did not reach statistical significance.

This finding can be interpreted as an outcome of increased *path* salience in Chinese. Although both Chinese and English speakers spent slightly less time to verify video pairs in the *path mismatch* than *manner mismatch* conditions, however, this temporal distance between prime and target video in the Chinese-speaking group reach statistical significance. This might seem odd at first glance. The Chinese speakers exhibited a speed advantage in *path mismatch* and *full mismatch*, not under the *manner mismatch* and *full mismatch* conditions, although these differences were not statistically significant. Recall that Chinese speakers were significantly slower than English speakers under all conditions when language-specific effects were boosted by overt linguistic labels in the sentence-video verification experiment. It is the faster assessment in the *path mismatch* condition, which inflated the temporal gap between *manner* and *path mismatch* conditions, thus reaching statistical significance.

Before Slobin (2006) classified Chinese as an equipollently-framed language, earlier motion event typology regarded it as a verb-framed language since it also can use main verb to describe *path*, and *manner* is optional in this case (Tamly, 2000). Consequently, the *path* becomes more prominent for Chinese compared to English. Moreover, equipollently-framed features were mainly identified in motion expressions using *BA* construction in Chinese (Tusun, 2023) which contains verb compound construction expressing *manner* and *path* equally in two verbs (e.g., *nan2ren2 BA3 xiang1zi0 ban1 jin4 wu1*, '*man BA suitcase carrying entering room'). Without highlighting by *BA* linguistic labels, the same motion videos can be described by using the *ZHE* construction in Chinese, which also uses two verbs to describe the *manner* and *path* both in verbs, but this structure encodes the *path* in the main verb (e.g., *nan2ren2 ban1 ZHE0 xiang1zi0 jin4 wu1*, 'man carrying ZHE suitcase entering room'). Encoding *path* in the main verb is often classified as the verb-framed language (Tamly, 2000; Ji & Hohenstein, 2014a). Consequently, in tasks that involve memory retrieval mediated merely by a linguistic prompt, universal *path* salience appears to emerge. Specifically, with the relatively greater weight of the Chinese *path* due to the absent equipollent-framed linguistic prime, Chinese were faster to verify the *path mismatch* than in the *manner mismatch* condition.

However, if this explanation holds and the language-specific effects were simply weakened by the missing overt linguistic labels, both Chinese and English speakers should show a significant temporal distance between processing *manner* and *path* manipulations, and the highlighted Chinese *path* should exhibit a greater influence on this baseline. Moreover, the Chinese speakers should react significantly more slowly in verifying the *path* manipulated stimuli, as reflected in the SV verification experiment. One possibility is that the cognitive processes to verify video stimuli with and without overt linguistic labels are distinct from each other. This is in line with the findings of Lupyan & Thompson-Schill (2014), which investigated the effects of verbal (e.g., a word ‘cow’) and nonverbal labels (e.g., a sound ‘mooing’) on conceptual activation. They found that although the use of linguistic labels led to a stronger effect on visual perceptions than those of non-linguistic labels, the specific effects were complicated due to distinct underlying mechanisms in different domains. For example, in the domain of colour, specific colour labels facilitated the recognition of the corresponding colour patches, and these effects remained effective even though overt linguistic labels (Lupyan *et al*., 2020). However, in the domain of motion, Chinese speakers might exhibit distinct cognitive patterns due to alternative ways to describe the same motion event, such as the *BA* and *ZHE* construction, which both expresses *manner* and *path* in verbs, but the former encodes *manner* in the main verb (e.g., *nan2ren2 BA3 xiang1zi0 ban1 jin4 wu1*, ‘*man BA suitcase carrying entering room’), while the latter encodes *path* in the main verb (e.g., *nan2ren2 ban1 ZHE0 xiang1zi0 jin4 wu1*, ‘man carrying ZHE suitcase entering room’). Similar processing patterns might be observed consistently regardless of whether involving overt linguistic when pair the satellite English with another language without controversial motion expressions, such as Spanish (verb-framed) that only encodes *path* (e.g., *saliendo*) in the main verb (e.g., *Una marmota saliendo de una cueva (corriendo),* ‘a marmot exiting from a cave running’).

Alternatively, the unexpected lack of statistically significant temporal distance between verifying the *manner* and *path* manipulation in the English speakers might be due to the higher salience of English *manner*. The rationale for processing *manner* and *path* salient components in a memory task without overt linguistic labels remains the same, that is, the more salient the stimulus the faster it will be to verify (Lupyan & Clark, 2015). Consequently, significantly faster verification times are expected to be observed in the *path mismatch* condition due to the universal spatial information taking control in the VV verification experiment. However, the saliency of English *manner* remains, and result in a relatively faster evaluation times in the English speakers than those in the Chinese speakers. Consequently, the difference between the time taken in the two critical conditions did not reach statistical significance arguably due to the shortened temporal distance between the *manner* and *path* manipulation, reflecting the deep influence of the English *manner*.

It is unknown which of these two interpretations is more likely until further evidence is provided in future studies with explicit designs for examining these specific interpretations. However, the findings obtained from the L2 learners in the VV verification experiment might have brought to light more of the details of the underlying mechanisms involved in the verification of motion animations without overt linguistic labels between Chinese and English speakers, and thus on the processing of motion events themselves.

5.1.4 The effects of the English manner on visual perception

Various processing speeds observed between monolingual Chinese and English speakers within the comparison of *manner*- and *path-manipulated* stimuli across the three experiments in the current study suggest an intriguing picture of language-specific effects on visual perception extending to the domain of motion. This includes cross-linguistic differences reflected in a low-level visual detection process without involving conscious non-verbal judgement, a high-level visual processing driven by conscious decision-making behaviour, and an additional high-level visual processing enhanced by the use of overt linguistic labels. Specifically, the monolingual English group was faster at detecting the *manner* manipulated stimuli in the b-CFS experiment, but slower in doing so for the same stimuli in the sentence-video verification experiment, and these speed differences disappeared when the same *manner* manipulated videos were assessed in the video-video verification experiment.

Within the predictive processing framework (Lupyan & Clark, 2015), the English speakers were faster to break through the mask suppression and detect in a *manner* (e.g., man pulling suitcase into room) than a *path* (e.g., man carrying suitcase out of room) manipulated stimulus because of the sharper contrast detected against the default expectations formed by the earlier sequential photo (e.g., man carrying suitcase into room). Signals with more contrast (e.g., alteration in *manner* for English speakers is more contrast than those in *path* due to a higher saliency English *manner*) are less predictive input that are processed up to the visual hierarchy, thus triggering participants to respond more quickly in updating their subsequent predictions. However, the situation is restructured when proceeding at a later stage of visual perception. Specifically, English speakers took more time to verify the *manner* than *path* manipulated stimuli in the sentence-video verification experiment, since, after quickly detecting the high-contrast English *manner*, high-level cognitive processing requires further estimations of the visual input (lupyan, 2008). Specifically, a clear prediction was formed by the proceeding sentence prime (e.g., man carrying suitcase into room); to verify the incoming video stimuli (e.g., *manner* manipulation: man pulling suitcase into room, or *path* manipulation: man carrying suitcase out of room), participants then needed to identify the specific details presented in the target videos and then map it with the early described motion scenarios (e.g., man carrying suitcase into room). The more highly salient English *manner* competed with more possible expectations induced by various linguistic labels for the same nonlinguistic features, thus slowing down the verification process. In contrast, the rough estimation shortened the time taken to verify the differences in the video pairs in the VV verification task.

5.1.5 The battle between manner and path saliency

Another interesting finding observed in the current study was that there were rather profound language effects induced by the English *manner* in the b-CFS and SV verification experiments. This is in line with the salience hypothesis (Slobin, 2006), which assumes languages can be classified based on the degree of *manner* salience in the motion expression, for instance, English is a *high-manner-salient* language since it has a rich lexicon to express *manner* with minor difference (e.g., hop, jump), whereas Spanish is a *low-manner-salient* language since it can use one word to describe a certain range of *manner*s (e.g., *saltar* can refer to both ‘hop’ and ‘jump’). As indicated by Feist and Férez (2013), “higher codability of *manner* of motion correlates with improved memory for *manner*” (p. 398). However, similar mismatch detection advantage was not found in the Chinese monolinguals, as it took them similar amounts of time to process all types of mismatches (i.e., *manner* mismatch, *path* mismatch, and full mismatch).

Following this line of reasoning, high-*manner*-salient English is expected to induce a distinct cognitive process compared to Chinese due to its relatively lower *manner* salience, and this is indeed supported by the evidence obtained from the monolingual English and Chinese speakers in the b-CFS and SV verification experiments in the current study. Specifically, the English speakers detected *manner* manipulated stimuli significantly faster than all the other three conditions, reflecting a pervasive effect induced by the salient English *manner*. This continued to be reflected in the SV verification experiment, in which the English speakers spent the longest time verifying manipulated stimuli in a *manner* that was more similar to the other three conditions. In contrast, the Chinese speakers did not resemble this pattern in those two experiments.

However, the prominence of English *manner* appears to diminish when the universal salience of *path* becomes dominant, as in the VV verification experiment. The influence of the salient English *manner* was attenuated, although an overt linguistic indication of motion details was introduced in the experiment instructions. Consequently, the language-specific evidence expected to be observed in the examination of the lexicalization of motion events was severely impaired due to the pervasive universal *path*. This is in line with earlier studies, for example, participants chose motion alternates with the same *path* regardless of the languages they speak (‘English and Spanish’ in Gennari *et al*., 2002; ‘Chinese and English’ in Ji, 2017) in non-verbal similarity judgment tasks. Such impairments became worse when participants were required to retrieve motion details from memory (Papfragou *et al*., 2002; Papafragou *et al*., 2008), even with the assistance of overt linguistic labels (Gennari *et al*., 2002).

The current study might resolve the absent language-specific effects under the pervasive influence of spatial information in motion event cognition by observing a surprised language-specific effect reflecting in the *path* manipulated stimuli obtained from the Chinese-speaking group. Without overt motion labels, visual inspection was drawn by the dynamic spatial information in motion events, thus paving the way for the manifestation of *path* saliency (Tamly, 2000). With an option to silently describe motion events by using the *ZHE* construction, Chinese speakers exhibit distinct patterns between processing *path* and *manner* manipulated stimuli due to an increased salience of Chinese *path*. In contrast, the salience of English *manner* was suppressed by the increased salience of *path*, and thus struggled to exert its influences as expected. However, it is unknown whether Chinese speakers indeed utilised the *ZHE* construction, and whether this in fact affects the verification times in the VV verification experiment. If this is the case, Chinese speakers in Ji (2017) would also have exhibited a similar advantage for Chinese *path* in their non-verbal similarity judgement task due to the availability of the *ZHE* construction, which was not what they observed. In the following section, the evidence from the L2 learners in the current study is discussed, and this could provide further evidence regarding the issues raised in the current discussion.

5.2 Processing *manner* and *path* in L2 learners

Consistent with the specific processing patterns observed in the two monolingual groups, the evidence obtained from the L2 learners provides further evidence to support the LRH, since L2 learners observed across all three non-verbal experiments seem to undergo certain cognitive/ conceptual transfer of encoding motion events (Athanasopoulos, 2012). Specifically, L2 learners in the b-CFS experiment exhibit a distinct pattern to detect low-level motion signals compared with those in the two monolingual groups, thus indicating a possible cognitive convergence; L2 learners in the video-video verification experiment exhibit an English-like pattern between encoding *manner* and *path* in non-verbal experiments without overt linguistic labels, suggesting a potential conceptual transfer to a L2 English pattern in motion event cognition; L2 learners in the sentence-video verification experiment exhibit an English-like pattern between the critical conditions, but an overall pattern that mirrored the native speakers, thus also indicating a possible cognitive transfer of encoding motion event cognition in L2 learners. That is, learning a new language might indeed alter underlying cognition as reflected in the potential cognitive restructuring observed in the three non-verbal experiments in the current study. More importantly, the L2 findings illustrate a detailed mechanism for possible cognitive restructuring in motion events, specifically, how the two language-specific concepts interact with each other and what factors might play an important role in the mediation of the cognitive processes. Specifically, the restructuring of L2 cognitive processes reflected in the detection times recorded in the b-CFS experiment focusses on the low-level automatic processing of motion events, thus demonstrating how underlying motion event cognition evolves due to the restructuring of motion expressions without being biased by potential linguistic involvement (Athanasopoulos & Bylund, 2023).

Furthermore, the L2 findings in the current study add an extra layer to our previous assumptions as regards the tension between salient *manner* and the universal *path* in motion events, and thus may shed light on unveiling the mystery of the mixed results observed in previous studies examining the motion lexicalisation (Bylund & Athanasopoulos, 2014). Additionally, exploring the potential cognitive restructuring in L2 learners whose source and target languages shared both similarities and differences will also help to fill the gap in our understanding of the influences of linguistic similarities (Wang & Li, 2021b; Ringbom, 2007), since the majority studies concerning L2 motion events have been undertaken with learners who have contrasting source and target languages (e.g., satellite-framed English and verb-framed Spanish). The rationale and interpretation of the specific speed differences found in L2 learners are consistent with those explained in the monolingual Chinese and English speakers. For instance, variation in detection and verification times mediated by specific factors (e.g., English *manner*) indicates an underlying change of the corresponding language-specific cognition, according to the predictive processing account (Luypan & Clark, 2015). With these cross-linguistic findings as baselines, our discussion of the evidence of L2 focusses on the potential cognitive restructuring and the influential factors influencing each specific cognitive outcome.

5.2.1 Low-level cognitive restructuring in L2 learners

The RT results of L2 learners observed in the L2 learners demonstrate an L1 influence on the cognitive restructuring (Pavlenko, 2011), and even appears to occur when conscious language assistance is impossible. This conclusion was based on two main findings. First, the overall pattern of processing stimuli across all four conditions in L2 learners was significantly different from both monolingual Chinese and English speakers, reflecting an in-between unique cognitive behaviour that seems to fit in the convergence process that is merged with both L1 and L2 patterns. However, when focussing on the two critical conditions, the L2 learners resembled the Chinese speakers, and although they spent less time detecting the suppressed animation in a different *manner* (e.g., man pulling suitcase into room) compared to the sequential prime (e.g., man carrying suitcase into room), but unlike the English speakers, this speed advantage did not reach statistical significance. In short, the cognitive processes exhibited by L2 learners in the b-CFS reflect a potential change towards native English patterns, but with constraints imposed by their L1-based cognition. This is consistent with the claim of the process of transfer of L1 concepts, which is that cognitive outcomes display “the influence of the L1 on the L2, refers to cases where speakers’ L2 performance is guided by L1 linguistic categories, frames of reference or preference” (Pavlenko, 2011, p. 246).

Additionally, the influence of the learners’ L1 was discovered in the differences between the L2 learners and two monolingual-speaking groups within each specific condition. Specifically, the in-between cognitive pattern obtained from the L2 learners was only discovered in the *manner mismatch* condition. A cognitive shift to the L2-based cognitive pattern was observed in all the other three conditions: the L2 learners resembled the English speakers and were significantly faster than the Chinese speakers to break through the suppressed mask and detect the corners of the appearing stimuli. This finding can be interpreted by the resilient English *manner* (Slobin, 2006). Unlike the typological prediction between equipollently-framed Chinese and satellite-framed English for encoding the *manner* and *path* L2, learners should exhibit potential changes similarly in processing both *manner* and *path* manipulated stimuli. In other words, the L2 learners should not only exhibit significantly different pattern to detect stimuli with *manner* manipulation (e.g., man pulling suitcase into room), but also detect those with *path* manipulation differently when compared with the two monolingual groups.

In addition, other factors might also play a role in affecting cognitive restructuring in L2 learners. In the follow-up analyses of the potential predictors in the present b-CFS experiment, only L2 proficiency (IELTS score) was found to exert a significant effect on the detection speed of the suppressed stimuli. To further explore the influences of L2 proficiency on the low-level detection speed, L2 learners were divided into two subgroups based on the IELTS scores: intermediate (IELTS score < 6.5) and advanced L2 (IELTS score > = 6.5) group. No overall difference was found between these two L2 groups. However, both groups were discovered to significantly faster than Chinese but slower than English speakers in detection stimuli with *manner* manipulated stimuli. This indicates that even advanced L2 learners struggled to develop cognitive patterns that mirror the L2 native speakers restructuring.

Specifically, the high-level L2 learners mirrored the English monolinguals in detecting stimuli with *path* manipulation, while in this specific condition, the low-level L2 learners resembled the Chinese monolinguals. However, this pattern was only found in the *path* mismatch condition, and the L2 learners were distinct from both monolinguals in detecting stimuli in the *manner* mismatch condition regardless of the different proficiency levels. This provides a further potential explanation for the inconsistency of L2 results within the L2 groups in the current b-CFS compared to those reported in Wang & Li (2021b) and Ji (2017). In their studies, English-Cantonese bilinguals and advanced Chinese learners of L2 English exhibited a cognitive shift to L2-based cognitive pattern when selecting both alternates with *manner* and *path* manipulation.

One potential reason for the L1-like pattern observed in advanced L2 learners in the current b-CFS study when detecting the *manner* manipulation might be that the overall proficiency level in the L2 group is not high enough to achieve an English-like cognitive pattern, considering that the evidence claimed to discover English-like pattern in earlier studies was observed only in advanced L2 learners (Ji, 2017) and bilingual speakers (Wang & Li, 2021a). If an IELTS score equal to or above 6.5 is equivalent to a proficient user based on the Common European Framework (CEFR) (Take IELTS, n.d.), and the L2 proficiency in L2 learners (IELTS: *mean* = 6.46) recruited in the current b-CFS experiment was lower than the advanced (IELTS > = 6.5) on average.

Alternatively, the current b-CFS experiment examined the low-level detection of the caused motion events, whereas those in Ji (2017) and Wang & li (2021b) examined the semantic analysis of the caused motion events. Their discovery of an English-like pattern of encoding both *manner* and *path* might be due to the potential linguistic recruitment during non-verbal experiments, that is, their L2 learners/ bilingual speakers might use L2 English sub-vocally to guide their encoding pattern exhibit an English-like way. More details are discussed in the two verification experiments, when a similar higher semantic level was the focus to examine the conceptual transfer in L2 learners.

Cultural immersion was also examined by comparing L2 learners who lived in China and York with equivalent L2 proficiency, since cognitive shift is also often related to the fact of whether living in a L2-speaking country (Cook *et al*., 2006; Athanasopoulos *et al*., 2009). The RT results obtained from the L2 learners in the current b-CFS experiment confirmed that cultural immersion did affect cognitive restructuring in L2 learners. Specifically, L2 learners who lived in China were significantly slower to detect stimuli than L2 learners lived in York regardless of the information manipulated in the stimuli. Moreover, similar to the L2 learners in York, those L2 learners in China were also faster to detect stimuli with *manner* manipulated information than those with *path* manipulated information, and this difference did not reach statistical differences. When focussing on the results in each condition, L2 learners in China were significantly slower than the three language groups (L2 learners in York, English speakers, and Chinese speakers) in detecting stimuli in the two critical conditions. Their RT patterns changed in the two control conditions, that is, they were significantly faster than L2 learners in York and English speakers but resembled the Chinese speakers.

These L2 learners in China seem to experience a cognitive transfer that is changing towards the L2 native speaker, and instead of being facilitate by such cognitive transfer, they were struggling! This finding might be related to L1 attrition in Ma & Vanek (2024), who also discovered that Chinese speakers who teach English in China performed worse than Chinese speakers who teach other subjects in a lexical comprehension and video description task. Even strikingly, the distinct patterns obtained from L2 learners in China between control and critical conditions indicate that such L1 attrition might be very likely related to motion cognition.

Language context was examined in the current study because previous literature discovered that L2 cognitive restructuring might also be mediated by immediate language use before undertaking non-verbal experiments (Bylund & Athanasopoulos, 2014). The RT results discovered in the L2 learners in the current b-CFS experiment showed that the non-verbal detection behaviour observed in the L2 learners was not affected by the immediate language use overall. However, interestingly, although no significant differences were found between the *manner* and *path* mismatch conditions in either of the L2 groups, they did demonstrate a tendency towards the language-specific pattern corresponding to the language used before the b-CFS experiment. That is, those tested in Chinese took slightly longer in the *manner mismatch* than the *path* mismatch condition, in contrast, those tested in English spent less time in the *manner* mismatch compared to time taken in the *path* mismatch condition.

There are a few potential reasons for this result. First, this may be reasonable because the current b-CFS experiment examines low-level processing, which is assumed to be driven by the underlying mechanism formed by long-term memory (Jiang *et al*., 2007). In this view, this paradigm successfully suppresses the pressure of the higher-level bias. Second, this might also be due to the lower proficiency of L2 learners in the present b-CFS experiment. The flexible switch between the L1 and English-like cognitive pattern requires fully acquired linguistic knowledge and is often observed in bilinguals with balanced L1 and L2 proficiency (Athanasopoulos *et al*., 2015). This is not the case for L2 learners in the current b-CFS experiment.

Third, the pattern of results might be due to the age of onset (AO) at which L2 learners acquired the English as a second language. Despite the ongoing debates on the definition of early versus late L2 learners/ bilinguals in the literature (Singleton & Lesniewska, 2021), if the L2 learners (AO, *mean* = 8.56, *SD* = 2.67) in the present b-CFS experiment are defined as early L2 learners based on the puberty age (e.g., age from 12 to 14) for differentiating early (e.g., before puberty age) and late (e.g., after puberty age) L2 learners (Bylund, 2009; Singleton & Lesniewska, 2021; cf. Lai *et al*., 2014), then the absent effects of language context might be reasonable. Since previous studies found that only late L2 learners were able to switch between L1 and L2-based concepts. For instance, in Kamenetski *et al*. (2022), who also detected low-level motion signals in an ERP study, only those L2 learners who had been exposed to English as an L2 before the age of 4 exhibited an English-like cognitive pattern. Similarly, the bilingual Cantonese-English examined in Wang & Li (2021b) were also bilinguals who started to acquire the L2 at an early age (that is, around 4 years old).

Furthermore, another factor that could have potential influence on L2 detection speed in the current b-CFS experiment was found, the daily use of L2 written English of L2, since an interaction between this factor and RT taken was found to exhibit a marginally significant result. This is reasonable, as this factor is also related to L2 proficiency (Pavlenko, 2011). No significant differences were found between L2 learners who used less and more L2 written English when compared these L2 groups in the four conditions. However, these two L2 groups did exhibit different patterns of detection speed in the *path mismatch* condition. Specifically, those who used more written L2 English resembled the English speakers, while those who used less L2 written English patterned with Chinese speakers.

This finding is consistent with the overall cognitive pattern observed in the group of L2 learners as a whole, who resembled English speakers in the *path mismatch* condition but were different from English speakers in the *manner mismatch* condition.

That is, the frequency of using an L2 in daily life might indeed influence cognitive restructuring in motion events, but only in the feature that requires less effort to acquire. Another reason might be that the L2 learners were not proficient enough, and cognitive restructuring might have been observed in the *manner mismatch* condition if we had tested near-native L2 speakers.

In sum, the current b-CFS experiment captured sensitive cognitive difference obtained from the L2 learners and displayed a detailed image to understand the specific processing of cognitive restructuring in motion event cognition. Firstly, without the help of using language verbally or non-verbally, even advanced L2 learners did not fully resemble the English speakers. Secondly, cultural immersion is vital for cognitive restructuring in motion events (Ji, 2017; Athanasopoulos, 2007). Specifically, during the processing of evolving towards an English-like cognitive pattern, L2 learners who live in China might experience an L1 attrition first before fully acquiring the L2 pattern. Thirdly, language context might influence the L2 performance in non-verbal experiments, but our b-CFS paradigm maximally suppressed such bias. Lastly, using L2 written English more might encourage L2 learners to achieve a cognitive restructuring and this happens to the motion features required less effort to acquire.

5.2.2 High-level cognitive transfer in the VV verification experiment

The RT results of L2 learners observed in the VV verification experiment suggest a potential cognitive transfer (Pavlenko, 2011), and this was only found in the advanced L2 learners. Specifically, the overall time taken to verify video-video (e.g., prime video: man carrying suitcase into room) pairs observed in the L2 learners was not significantly different from the two monolingual groups. Time taken in the *manner* mismatch (e.g., man pulling suitcase into room) condition was not significantly different from those found in the *path* mismatch (e.g., man carrying suitcase out of room) condition. This may be surprising from the first glance but becomes reasonable considering that the two monolingual groups were not different from each other in the overall RT patterns. However, the monolingual groups did spend significantly less time verifying target videos with *path* manipulated information arguably due to the pervasive effects driven by the salient Chinese *path* observed specifically in the video-video verification experiment. In L2 learners, even though they also spent less time in the *path mismatch* compared to those in the *manner mismatch* condition, the shorter RTs were not found to be significantly different from those observed in the *manner mismatch* condition.

Moreover, it is important to note that in the current VV verification experiment, the cross-linguistic difference between Chinese and English was not consistent with the typological differences in accordance with what has been reported for equipollently-framed and satellite-framed languages (Ji, 2017). That is, the encoding of *manner* and *path* should be significantly different from each in English speakers due to the higher salient English *manner* when expression motion events, whereas the encoding patterns between these two components should be similar due to the equal saliency of expressing Chinese *manner* and *path*. The salience of Chinese *path* might be highlighted by a possible use of the *ZHE* construction, which typically expresses motion events more likely to follow the typology as a verb-framed language. The cognitive restructuring (compared to the two monolingual groups) in the L2 learners suggested a transformation from higher salience in *path* to a higher salience in *manner*, which corresponds to the idea that "later learned languages encode categories, perspectives or frames of reference absent in the L1" (p.247), as defined by Pavlenko (2011) to identify a cognitive transfer. These findings are consistent with part of the L2 evidence found in Ji (2017), for example, who also found an English-like pattern in the L2 learners in a non-verbal similarity judgement task, and which was specific to those of advanced proficiency.

Immediate language use (operationalised in the current experiment as experimental context) was not found to significantly affect the RT taken to verify target motion animations in L2 learners. The language context is designed to be achieved by the communication between the experimenter and the participants, since the current VV verification experiments examined the motion events in non-verbal context. Specifically, in the monolingual context, the L2 learners were only allowed to speak

Chinese to the experimenter in the communication before performing the non-verbal experiments in the language lab; additionally, the instruction before engaging in the VV verification experiments was also presented in Chinese. In contrast, those in the bilingual context were required to use Chinese for the communication before engaging in the VV verification experiment, and they had to switch to English when they were in the lab, including the instruction and other verbal communication. The results of language context are not consistent to the predictions, since high-level semantic analysis should be affected by immediate language use without measurements to prevent participants from recruiting languages sub-vocally (Gennari *et al*., 2002). Moreover, to prompt the potential language-specific effects on the examination of motion lexicalisation (Papafragou & Selimis, 2010), participants were reminded to focus on the motion details for verifying the video-video pairs. Furthermore, the L2 learners recruited in the current VV verification experiment were early (AO: *mean* = 8.23, *SD* = 2.27) L2 learners (Singleton & Lesniewska, 2021, Bylund, 2009; cf. Lai *et al*., 2014), and only late bilingual or L2 learners have been more often found to exhibit a flexible cognitive shift between L1- and L2-based motion event processing.

However, like the potential reason set out above to explain the earlier b-CFS findings, the L2 learners in the present video-video verification experiment were also mixed intermediate and advanced L2 learners based on their average IELTS score (6.41). An IELTS score equal to or above 6.5 is equivalent to a proficient user according to the Common European Framework (CEFR) (Take IELTS, n.d.). These studies discovered effects of language context on non-verbal experiments, often studied bilinguals with balanced L1 and L2 proficiency (Barner *et al*., 2009; Sachs & Coley, 2006). Thus, cognitive shifts induced by immediate language use may not occur unless L2 proficiency is high enough (Bylund & Athanasopoulos, 2014).

A marginally significant effect of the interaction between verification times and L2 proficiency (IELTS scores) was found, indicating a potential influence of L2 proficiency on the time taken to verify video-video pairs. Following the same procedure of splitting the L2 learners based on their L2 proficiency, the intermediate and advanced

L2 learners were found to vary in temporal distance between verifying a target video with information manipulated in the *manner* and *path.* Specifically, advanced L2 learners resembled the English speakers, and did not exhibit a significant difference between processing stimuli in *manner* and *path mismatch* conditions. In contrast, intermediate L2 learners were found to verify stimuli significantly faster in the *path mismatch* than those in the *manner mismatch* condition.

No significant differences between the language groups were found in each specific condition, which is consistent with what was found for the two monolingual groups. This might be due to the universal salience of spatial information in the dynamic processing of motion events, which may have suppressed the language-specific effects induced by the salient English *manner,* thus prompting the influence of *path* salience instead (Talmy, 1991; Radvansky & Zacks, 2014).

5.2.3 High-level cognitive transfer in the SV verification experiment

RTs obtained from L2 learners in the current sentence-video (SV) verification experiment illustrates a cognitive transfer (Pavlenko, 2011; Athanasopoulos, 2011). Specifically, the overall verification times taken to process stimuli across all four conditions aligned with those discovered in the Chinese speakers but were significantly slower than English speakers. However, for critical conditions, the L2 learners were found to spend significantly more time verifying the target video with information manipulated in a *manner* based on the motion description presented in the earlier prime sentence, thus reflecting adjusted weights from the *manner* and *path*, which were transformed from the equivalent weights between the Chinese *manner* and *path*.

Furthermore, when the L2 learners were compared with the two monolingual groups in each condition, they were found to be slightly faster than the Chinese speakers in all conditions (i.e., *full match*, *path mismatch*, and *full mismatch*) except the *manner mismatch* condition. Specifically, instead of encoding stimuli manipulated in a *manner* faster than the Chinese speakers, and patterning more similar to the performance of the English speakers, the L2 learners was in the opposite direction, and they spent the longest time in the *manner mismatch* condition, which is even slower than the verification times taken in monolingual Chinese speakers, given that the Chinese speakers were slower than the English speakers under all four conditions. This longest time spent in the *manner mismatch* condition stretches the temporal distance between verifying stimuli with *manner* and *path* manipulated information, even reached statistical significance, and thus resembled the English speakers. Also, this longest RT time found in the *manner mismatch* conditions is also consistent with the English speakers, who spent the longest time in the same condition. It seems that with the assistance of overt linguistic labels (sentence primes), the L2 learners exerted an English-like pattern, but it may not yet be internalized, otherwise they should also be faster to encode *manner* manipulated information in comparison to the Chinese speakers, reflecting potential cognitive transfer (Pavlenko, 2011).

Language mode (context) was not found to affect the verification times observed in L2 learners in the current SV verification experiment, indicating that these participants performed similarly on the target videos regardless of the languages used in the sentence primes. However, differences between groups emerged within each condition. Although the L2 learners showed similar patterns in verifying the stimuli with *manner* manipulation regardless of the tested languages, they did vary in the *path* manipulated conditions. More unexpectedly, this is inconsistent to previous studies (Gennari *et al*., 2002; Wang & Li, 2021a; Papafragou & Selimis, 2010), that is, participants engaged in the non-verbal experiments are mediated by overt linguistic labels. The current SV verification study employed sentence primes as overt linguistic labels to activate relevant motion cognition, and L2 learners should exhibit conceptual patterns more like English monolinguals when presented with English sentence and resemble the Chinese monolinguals when the sentences were in Chinese. Instead, the L2 learners tested in the monolingual context were significantly faster than the Chinese speakers but slower than the English for the *path* manipulated stimuli, whereas those tested in the bilingual context were significantly slower than the English speakers but comparable to the Chinese speakers in processing the same *path* salient stimuli.

Recall that in the experimental design of the SV experiment, since there were overt linguistic labels, the language mode was implemented in the sentence primes. That is, the L2 learners in the monolingual context viewed Chinese sentences only, while the L2 learners in the bilingual context viewed all the critical stimuli in English, and 6 more filler sentences in English, while the rest of the sentences were in Chinese. Perhaps it was the increased cognitive load caused by the need to switch between two languages that slowed the L2 learners down, and this is only found in the stimuli with *path* manipulation is because Chinese and English *path* are equally salient for L2 learners. The overt linguistic labels could have pushed the participants to switch between English and Chinese *path*, which may have required extra cognitive effort to map *path* with equal linguistic representations in Chinese and English (Athanasopoulos *et al*., 2015), thus resulting in slower verification times. Whereas in the *manner* manipulated condition, *manner* is less salient in Chinese, thus involving less competition in semantic analysis.

None of the other predictors was found to significantly affect the verification times obtained from the L2 learners, including L2 proficiency, daily speaking and writing of L2 English, age of onset of learning English as L2, and time spent in UK.

5.2.4 Cognitive restructuring vary at different cognitive levels

Different cognitive processing patterns were found in L2 learners across the three non-verbal experiments conducted in the current study. Specifically, an L1 influence was observed in the detection times in the b-CFS experiment, cognitive transfer was evidenced in the verification times taken to evaluate motion videos in L2 learners in the VV verification experiment, and cognitive restructuring was identified in the SV verification experiment.

Firstly, such distinct cognitive outcomes might be due to the variation of L2 proficiency levels examined in these three experiments. As discussed in Park *et al*. (2022) and Athanasopoulos (2007), L2 cognitive restructuring is mediated by the L2 proficiency, and the higher the L2 proficiency, the more likely learners to display target-like cognitive processing patterns (Pavlenko, 2011). The L2 learners recruited in the current three experiments had different levels of proficiency. Specifically, the average IELTS scores reported in L2 learners in the b-CFS experiment, sentence-video verification, and video-video verification experiments were 6.46 (*SD* = 0.71), 6.54 (*SD* = 0.71), and 6.41 (*SD* = 0.59), respectively. Those L2 learners examined in the b-CFS and VV verification experiments seem to be similar, as their average IELTS were both below 6.5. Those recruited in the SV verification experiment were above 6.5, which is equivalent to advanced levels according to the Common European Framework (CEFR) (take IELTS, n.d.). However, in contrast to the influence of L2 proficiency, L2 learners tested in the b-CFS and VV verification experiments exhibited distinct cognitive processes: L1 transfer and L2 transfer.

It seems that instead of being affected by level of proficiency, these specific cognitive processes might be due to the variation levels examined in these specific experiments. For instance, the b-CFS experiment captures the low-level visual detection of the motion signals (Jiang *et al*., 2007), and it examined language-specific effects without being biased by recent language use. In contrast, the VV verification experiment was unable to exclude such bias. Therefore, although these two groups of L2 learners had similar L2 proficiency, their cognitive outcomes were distinct when examined between the advanced and intermediate L2 groups. Specifically, when conscious linguistic recruitment is possible, the L2 learners appeared to be more likely to behave like the L2 native speakers. More importantly, the advanced L2 learners mirrored the English speakers, while the intermediate learners patterned with the Chinese monolinguals when verifying stimuli between *manner* and *path* manipulated information. However, when linguistic assistance was excluded, their cognitive behaviour remained in the L1 pattern. As reflected in both b-CFS, the advanced and intermediate L2 learners mirrored the Chinese speakers. Therefore, the L2 learners examined in the VV verification experiment might have exhibited a cognitive trasnfer driven by surface linguistic representations. In sum, during the process of learning a new language, particularly, learning to express motion events in English, L2 learners might be able to internally think like the L2 native speakers, but not without effort.

Fundamentally, their cognitive pattern still mirrored the native L1 speakers. Future studies with explicit designs, especially with the same group of L2 learners examined at different cognitive levels, would need to be carried out to further investigate the value of this speculation.

Another reason for having distinct cognitive outcomes in the b-CFS and VV verification experiment might be the different motion features examined. The language-specific effects detected in the b-CFS were consistent with the typological differences in describing motion events between an equipollently-framed Chinese and satellite-framed English, that is, the Chinese speakers detected motion stimuli with *manner* and *path* manipulated information similarly, but the English were more attentive to those with *manner* manipulated information. In contrast, with the increased possibility of high-level bias, a rather unique situation was created in the VV verification experiment. As explained in Section **5.2.4.** with the increase of the universal salient spatial information in dynamic motion events, Chinese speakers might be encouraged to use *ZHE* construction to complete the verification task. The *ZHE* construction is regarded as a representation of verb-framed language (Tamly, 2000; Ji & Hohenstein, 2014a), thus spotlighting the salience of *path*. This was exactly what was discovered in the VV verification experiment. Moreover, unlike the *manner* component, *path* is not optional in either Chinese or English, thus it might be more easily acquired for the L2 learners. Specifically, L2 learners tested in the b-CFS were not observed to change to an English-like cognitive pattern even when their level of L2 proficiency was high, while L2 learners examined in VV verification demonstrated an L1-like pattern in their intermediate group and a English-like pattern in their advanced group.

Interestingly, such *manner* resistance observed in the b-CFS experiment became an assistance in the SV experiment. Specifically, the L2 learners examined in the SV verification experiments may have been guided toward the cognitive patterns specific to the typology of an equipollently-framed or a satellite-framed language, due to the overt linguistic labels displayed in the sentence primes. These L2 learners mirrored English speakers in the two critical conditions but remained akin to the Chinese speakers in the overall processing preference. That is, L2 learners were found to spend the longest time in the *manner mismatch* compared to all other conditions. *Manner* salience appears to be stronger in the L2 learners than in those exhibited by the English speakers, as the time taken to verify sentence-video pairs in the English speakers was only significantly longer than the *path* and *full mismatch*, but not in the *full match* condition.

As discussed in Section **5.1.1.**, in the b-CFS experiment, the advantage of the English *manner* was also evident in the shortest detection times obtained in the *manner mismatch* condition. However, the interpretation is different. The strong salient *manner* discovered in L2 learners can be interpreted as evidence of knowledge that has not (yet) been internalised. For example, L2 learners might be able to describe motion events in English without a complete transformation to the thought similar to the target language. This is another reason for the need to undertake more non-verbal experiments in order to understand the potential effects of language on thought (Lucy, 1997; Athanasopoulos, 2012; Bylund & Athanasopoulos, 2014; Levinson, 2012). In the SV verification experiment, linguistic labels are available to complete the semantic analysis, and this was further confirmed by the effects of language context. It was the L2 learners tested in the bilingual context who exhibited stronger English-like patterns when the sentence primes were presented in English.

Furthermore, the different cognitive outcomes of the L2 learners examined in the SV and VV verification experiments were also evident when comparing the verification times taken for target videos with and without linguistic labels. Regarding predictors that might affect the cognitive processing patterns in L2 learners, the effects of language context (or immediate language use in the VV verification experiment) varied across the two verification experiments. Consistent with the findings of previous studies (Wang & Li, 2021b; Athanasopoulos *et al*., 2015), only the late L2 learners/bilingual speakers appeared to exhibit a flexible switch between L1 and L2-based cognitive patterns, and this seems to only be achieved when participants reach a relatively high level of L2 proficiency. In the current study, if participants in both L2 groups recruited in the SV verification experiment were regarded as late L2 learners (AO, *mean* = 8.94, *SD* = 3.07) and VV (AO, *mean* = 8.23, *SD* = 2.27), only the advanced (IELTS = 6.54) L2 learners tested in the SV exhibited language-specific patterns corresponding to the specific language used in the prime sentences, while those who were less proficient did not differentiate in performance when tested in either language although they were also late L2 learners.

5.3 The pervasive language effects of motion events

5.3.1 Unparallel thinking and speaking

Previous studies examining L2 cognitive behaviour undertaken to investigate the effects of language on thought support the thinking-for-speaking (TFS) hypothesis (Slobin, 1996), which claims that the linguistic influence on thought is limited to the period of actively using the language. The findings of the present study reject this assumption since language-specific differences between monolingual Chinese and English speakers were found during lower-level motion encoding when conscious language recruitment is impossible. Furthermore, via a comparison across experiments examining different levels of motion encoding, it is evident that when the effect of language was boosted by overt linguistic labels, even intermediate L2 learners were able to perform like the speakers of the target language. However, in the absence of verbal assistance, L2 learners remained in the L1 pattern, although their L2 proficiency was at advanced levels. With a more complicated situation in the domain of motion events, when motion-related cognition was activated earlier before engaging in a non-verbal experiment (Papafragou & Selimis, 2010; Wang & Li, 2021a, 2021b), the advanced and intermediate L2 learners diverged from each other, and their performance resembled either the target or source language, respectively.

Furthermore, the findings obtained from the monolingual speakers in the two verification experiments also partially reject the predictions of the TFS, since the degrees of linguistic involvement might exert different influences on the cognitive activity and thus blur the identification of the language-related effects. As reflected in the distinct verification patterns obtained from the two non-verbal experiments, both overt and covert linguistic labels indeed exerted a boosting impact on motion perception, and the specific processes diverged dramatically, as predicted by the label-feedback hypothesis (Lupyan, 2012). That is, in the SV verification experiment, according to this account, the overt linguistic labels (prime sentences) activated the internal motion event cognition formed by long-term use of motion description in specific languages (e.g., equipollently-framed Chinese and satellite-framed English), the verification procedure of the subsequent video motion continues functions to up or down regulate the current motion perceptions build be the preceding motion labels. Furthermore, label-feedback also asserts that although linguistic labels exert robust effects compared to non-linguistic representations they function in different and complex ways (Luypan *et al*., 2020), in the authors' own words 'in a task requiring people to remember an item's exact position or colour, the finding that people's memory is affected by categories (both linguistic and nonlinguistic) can be usefully modelled by merging continuous perceptual representations with more categorical (discrete) conceptual/linguistic ones, with the original perceptual representation left intact' (p,939). Specifically, the video primes employed in the VV verification experiment not only activated the language-specific (*manner* and *path* in equipollently-framed Chinese and satellite-framed English) motion cognition, but also activated the other categorical activation (e.g., dinosaur in the current *full mismatch* condition) in a caused motion scenario.

5.3.2 The overall slower reaction times in the Chinese speakers

One of the most important contributions of the current L2 findings was to provide further evidence for language-specific effects, but this is unlikely to be correlated with motion events. In the cross-linguistic evidence obtained in the two monolingual groups, Chinese speakers were significantly slower than English speakers across all four conditions (including the control conditions) in both the b-CFS and SV verification experiment, but not in the SV verification experiment.

The slower overall RTs have also been observed in other related studies (Vanek & Zhang, 2023; Ji, 2017). For instance, in the non-verbal similarity judgement experiment conducted in Ji (2017), the average times spent by the Chinese speakers were also consistently longer in processing *manner* (Chinese: 2553 ms; English: 1883 ms) and *path* (Chinese: 2557 ms; English: 2093 ms) alternated stimuli than their English speakers. However, in another non-verbal similarity judgement experiment conducted in Wang & Li (2021a), their Chinese speakers showed comparable overall RT patterns compared to English, Japanese, Cantonese (Chinese) – English bilinguals, and Cantonese-English-Japanese multilingual speakers. One potential reason might be the cultural difference between Chinese and English speakers. Specifically, the Chinese speakers in Ji (2017), Vanek & Zhang (2023) and the current study were all recruited in mainland China, and the Cantonese speakers in Wang & Li (2021b) were recruited in Hongkong, China. Furthermore, in another L2 group in our current b-CFS experiment, who also participated in the experiment in China, an overall slower RT pattern compared to the L2 learners tested in York with equivalent proficiency levels. The possible reasons for this might be the environment of learning English as a second language and their familiarity of engaging in language experiments. For example, Chinese monolinguals in Ji (2017), Vanek & Zhang (2023) and the current study were school students who had never been exposed to natural language learning environments, while those in Wang & Li (2021b) might start to learn English in a more natural scenario, considering the multilingual environment in Hongkong, China.

However, this assumption is not supported by the evidence observed in the current VV verification experiment. Specifically, all three language groups (i.e., monolingual English and Chinese, L2 speakers) examined in this non-verbal verification experiment exhibited similar overall reaction times regardless of the locations in which they were tested. This overall cross-linguistic differences disappear when the task does not increase cognitive load, for instance, Ji (2017) requires participants to make judgements under a verbal interference task, the current b-CFS experiment requires a slight self-adjustment for the experimental equipment, the unfamiliarity of these procedures might result different overall RT patterns between Chinese speakers in China and English speakers in the UK.

This overall language-specific difference was further confirmed by the evidence obtained from L2 learners in the b-CFS and SV verification experiments. Specifically, in the b-CFS experiment, the overall detection times observed in L2 learners across the four conditions were significantly faster than those in the Chinese speakers, but significantly slower than English speakers. In the SV verification experiment, the overall verification times recorded in the L2 group were also slower than those recorded in English speakers, but faster in the Chinese speakers. However, only the difference between L2 learners and English speakers was found to be statistically significant in the SV verification experiment. Furthermore, in the results in each condition, the L2 learners tested in the b-CFS experiment were only significantly different from the two monolingual groups when detecting stimuli with *manner* manipulated information but mirrored English native speakers in detecting stimuli with *path-manipulated* information. In contrast, the L2 learners in the SV experiment resembled the Chinese speakers in verifying the video stimuli with both *manner* and *path* manipulated information.

However, this overall language-specific difference observed in RTs might also affect the specific cognitive processes found in L2 learners. Specifically, although L2 learners in the b-CFS patterned with English speakers, their significantly slower detection times observed in the *manner* manipulated condition seems to compress the temporal distance between the detection of the two critical conditions, that is, no significant difference between detecting stimuli in the *manner* and *path* mismatch condition, thus reflecting an L1-like cognitive pattern between the two critical conditions. In contrast, the L2 learners examined in the SV verification experiment mirrored the Chinese speakers in overall RTs across four conditions, but the temporal distance between the two critical conditions resembled the English speakers. That is, they verified the target videos in the *manner* and *path* mismatch conditions in significantly different ways, thus displaying an English-like cognitive pattern. If the overall RT differences are not related to motion event processing, our interpretation of the cross-linguistic differences would only focus on the temporal distance between the two critical conditions, as predicted by the equipollently-framed Chinese and satellite-framed English (Tamly, 2000). Consequently, alternative interpretations for the L2 results obtained from all three non-verbal experiments would be that the L2 learners in the b-CFS did not exhibit a different cognitive pattern compared to the English speakers, and the L2 learners in the SV and VV experiment would both illustrate a cognitive process prone to an English-like cognitive pattern, suggesting a potential conceptual transfer.

However, this simplified interpretation would lose the observation of specific details within the processes of cognitive restructuring observed in the L2 group. For example, in Papafragou *et al*. (2008), even though Greek (verb-framed) and English (satellite-framed) speakers did perform uniformly in a non-verbal recognition task, their eye movements indicate cross-linguistic differences as predicted by the LRH, but in a typologically reversed way (Tamly, 2000). That is, English and Greek speakers tended to inspect regions that are opposite to the salient components highlighted by the motion expressions in their specific languages, that is, English speakers should pay more attention on the *manner* region due to the higher salient English *manner* (compared to *path*) while Greek speakers should focus more on the *path* region due to the relatively higher salient of *path* in Greek. Including the overall cross-linguistic differences as reference lines might also reveal the dynamic relationship between the salient *manner* and universal spatial information in non-verbal experiments focussing on the motion lexicalisation with overt or covert linguistic involvement, especially in the L2 motion studies, when the detailed cognitive restructuring are in demand to understand the gradual and linear process of cognitive transfer (Pavlenko, 2011; Jarvis & Pavlenko, 2010).

5.3.3 The resilient English manner

In line with the previous studies using online paradigms to capture the language-specific details suppressed by the uniformed *path* preference in motion cognition observed in non-verbal experiments (Ji, 2017; Wang & Li, 2021b; Papafragou *et al*., 2008), the present SV verification experiment also adds more supporting evidence to the cross-linguistic differences predicted by the distinct motion expressions. Furthermore, the current study includes L2 learners in the participant pool and examines the same stimuli at the low and high-level of cognitive behaviour, thus providing an alternative inference of the ‘missing’ language-specific effects in earlier motion event studies (Papafragou *et al*., 2002; Gennari *et al*., 2002). That is, language-specific effects induced by motion lexicalisation are evident only when linguistic recruitment is available. The critical issue in previous non-verbal experiments exploring *manner* and *path* encoding is that spatial information attracted attention naturally in an unfolding motion scenario.

Unlike other domains with static stimuli, such as colour (Winawer *et al*., 2007; Gilbert *et al*., 2016), or object (Gilbert *et al*., 2008), which illustrates relatively less information in the stimuli, thus making it more likely to find language-specific differences. In the domain of motion, language-specific and universal effects seem to emerge in dynamic ways in encoding the *manner* and *path.* Specifically, *path* is preferred across language groups in non-verbal experiments. For example, in the similarity judgment task (Ji, 2017), even though language-specific differences were not captured by semantic decisions, but the time used to make these decisions diverged in language-specific ways. That is, Chinese speakers used a similar amount of time to select alternations with *manner* manipulation compared to those with *path* manipulation, whereas it took English speakers less time to make decisions for *manner* manipulation than those with *path* manipulation. Consistently, the evidence obtained from the current study indicates that *manner* and *path* compete during perceptual processing, and even though *path* is shared across languages, they might vary as regards the degrees of the saliency.

The current interpretations of the dynamic interaction between the English *manner* and universal *path* start from the overall RT patterns obtained from all participants in the two control conditions (*full match* and *full mismatch*), which sets the reference lines for the low-level detection and high-level semantic analysis and illustrates the specific differences between different levels of *manner* and *path* encoding. Specifically, low-level detection should preclude nonlinguistic biases with specific manipulation in the stimuli (Flecken *et al*., 2015). In the present study, this is confirmed by similar detection patterns between the two control conditions. As regards the higher level of semantic analysis, this nonlinguistic bias emerges between the control conditions, so the participants, regardless of the languages they speak, reacted significantly faster in the full mismatch condition (stimuli with language-irrelevant manipulation) than those in the full match condition (stimuli with no manipulation) in the SV verification experiment. The temporal distance between the two control conditions expanded maximally in the VV verification experiment, when language-specific effects were not highlighted enough.

Following this argument, the findings in the current study create a more complete picture of the saliency competition between *manner* and *path*. Firstly, the salient English *manner* emerged at an early stage of visual processing, that is, at the detection stage. This is the period of automatic processing that was induced by deep memory formation over a long period of time, such as in the description of motion events. With imbalanced salience between *manner* and *path*, English speakers appear to be more sensitive to the *manner* manipulation at this stage, reflecting a more salient *manner* in English. This is consistent with its typological categorisation as being verb-framed which uses main verbs to describe *manner* (Tamly, 1985). Whereas Chinese speakers reacted similarly between the *manner* and *path* manipulations, which corresponds to the equally weighted *manner* and *path* in the equipollently-framed Chinese. Moreover, the fact that *manner* is salient is also evident in the findings in the L2 learners. Specifically, cognitive transfer was observed in all the other conditions except in the *manner* manipulated condition (Pavlenko, 2011).

Moving forward to a later stage of the higher-level motion encoding, for instance, when linguistic recruitment is obligatory, as designed in the current SV verification experiment, universal spatial influence and language-specific effects appeared to compete and exerted comparable effects in decisions involving semantic analysis. This was reflected in the similar temporal distance to verify motion manipulation within each pair of nonlinguistic (i.e., the control conditions: *full match*, 'man carrying suitcase into room'; *full mismatch*, 'man carrying dinosaur into room') and linguistic (i.e., critical conditions: *manner mismatch*, 'man pulling suitcase into room'; *path mismatch*, 'man carrying suitcase out of room') pair in the English-speaking group, where language-specific effects emerged. The Chinese behaved similarly in the control conditions since the discrimination of an object manipulation (e.g., suitcase vs. dinosaur) did not vary between Chinese and English speakers. This pattern was also found in the L2 group. More importantly, at this stage, with the facilitation of overt linguistic labels, English *manner* continued to function well despite the semantic noise evident in the control conditions (i.e., English speakers verify the target video with *manner* manipulation significantly different from those with *path* manipulation, and they also processed stimuli in the *full mismatch* significantly different from those in the *full match* condition). This is further supported by the findings observed in the L2 learners, who showed an overall L1-like pattern under the experimental conditions, except in the *manner* mismatch condition. Aided by overt *manner* labels, cognitive transfer to English-like motion cognition is evident through *manner-manipulated* stimuli (Pavlenko, 2011).

It appears that the salience of English *manner* is found when the language-specific event motion typology is highlighted by verbalised motion expressions (i.e., in the SV verification experiment), or when the universal spatial information in motion events is suppressed (i.e., in the b-CFS experiment). The situation changed dramatically when the task required high-level semantic evaluation between two motion animations, at which point spatial information became predominant. As illustrated in the current VV verification experiment, the temporal distance between the two control conditions is the largest, but unlike the findings observed in the SV verification experiment, those between the two critical conditions (i.e., *manner* mismatch and *path* mismatch) conditions in the English-speaking group were smaller when the prime sentences were replaced by motion videos in the VV verification experiment. The Chinese-speaking group showed a similar pattern under the control conditions; however, the RT differences between the two critical conditions were greater and reached statistical significance. Together with the findings obtained from the L2 learners in the VV verification experiment (that is, they mirrored the English speakers, exhibit numeric differences between processing *manner* and the *path* manipulation but did not reach statistical significance), thus, the final interpretation seems more prone to mediation from the English *manner*.

Specifically, in line with the two monolingual groups, L2 learners also exhibited a large temporal gap between the two control conditions, further confirmed the high-level semantic analysis examined in the VV verification experiment. In terms of critical conditions, the L2 learners mirrored English speakers: their temporal distance between verifying *manner* and *path* manipulations was small and did not reach statistical significance, reflected in monolingual Chinese speakers. More importantly, their verification times were not affected by the distinct language contexts used before the non-verbal experiment. This rejected our previous assumption, which assumed that the monolingual Chinese speakers were able to use *ZHE* construction (e.g., encode *path* in main verb) sub-vocally, thus generating a distinct RT pattern between processing *manner* and *path-manipulated* stimuli in the VV verification experiment. The L2 findings in the VV verification experiment fill the gap in our earlier discussions regarding the dynamic interaction between the salience of *manner* and *path*, supporting the notion of a pervasive influence of the salient *manner*.

In conclusion, when all the pieces are considered together, it seems that the language-specific effects of motion event processing emerge at an early stage of visual processing (Athanasopoulos & Casaponsa, 2020; Kamenetski *et al*., 2022), and continue to influence processing at a later stage with the assistance of linguistic labels (Gennari *et al*., 2002) and linguistic prompts (Papafragou & Selimis, 2010), reflected by the distinct but consistent processing patterns of salient English *manner* reflected in the three non-verbal experiments in the current study.

5.4 Wider contributions

5.4.1 Using b-CFS for the examination of moving stimuli

Another contribution of the current study is methodological. It is the first study to employ the b-CFS task to examine the processing of real-life (not schematic) dynamic stimuli with complex information. Previous research has attempted to utilise b-CFS for examining schematic stimuli, but limited to simple spatial information, such as directions (Hong, 2015). Motion events contain multiple types of information, such as directions, actions, subject, objects, and even background. One major concern was the efficiency of the suppressing function induced by the mask: specifically, that it may not be effective for such complex stimuli. Conversely, an alternative concern might be that the stimuli were too complex to be detected through the suppressing mask. That is, the balance between the mask and the stimuli was essential. Following the brief instructions provided by previous studies using b-CFS (Pournaghdali & Schwartz, 2020), which indicates that the colour and flickering rate should be consistent with the stimuli. In the present study, the stimuli comprised colourful real-life motion animation, and we designed a colourful mask and set the flickering rate at 10 hz. By running a control experiment, our mask was found to effectively suppress dynamic stimuli. This provides evidence for the efficacy of this method for studies in other semantic domains, particularly those that investigate the relationship between language and cognition. Furthermore, the success of applying the b-CFS in motion event cognition also suggests that perhaps complex stimuli are in fact better options for visual detection. In particular, to examine the effects of language on thought, the saliency of language effects is prominent when high-level noise is removed by the b-CFS experiment, which is consistent with the predictions proposed by the LRH.

5.4.2 Is b-CFS better?

The B-CFS is flexible enough to be manipulated according to different research purposes. In studies using other low-level detection approaches, such as EEG and ERG, the process in non-verbal experiments is often not carefully controlled for inner speech. For example, Flecken *et al*., (2015) recorded the brain activity of participants while performing a video-picture matching task. This approach captures the onset of semantic analysis regardless of whether it is related to language or not, filtering the effects of language based on the brain activity recorded during semantic analysis. The process of making these judgments involved high-level cognitive analysis. In contrast, in the b-CFS experiment in the current study, the task was to detect the position of the stimuli appearing, which involves low-level visual recognition. Consequently, although EEG records the brain activity of using language to make semantic decisions, the process of making such judgement does not exclude the use of inner speech. We argue that the b-CFS maximally reduces the possibility of using inner speech even during non-verbal experiments. On the other hand, b-CFS remains a powerful tool for examining the relationship between language and cognition, particularly when focusing on low-level processing. Most importantly, future studies employing this method should carefully consider the validity of the mask used to suppress their experimental stimuli.

However, b-CFS is not suitable for all studies investigating the relationship between language and cognition, such as those looking at specific language effects based on cross-linguistic differences reflected in the ‘language brain’ (i.e., lateralisation). For example, Gilbert *et al*. (2016) and Athanasopoulos & Casaponsa (2020) claim to support the LRH by discovering cross-linguistic differences specific to the right visual field (representing the left brain), indicating a language-related effect. In the present study, b-CFS also separated the visual fields, and the stimuli were placed in the dominant visual field. Therefore, if cross-linguistic differences were discovered when the b-CFS mask is presented to the left eye, it suggests that a language effect emerged. However, the results in the present study illustrated a reverse scenario. That is, language-specific differences were only found when stimuli were presented to the left visual field, which is contrary to the evidence obtained from previous laterization studies. One possible interpretation is that the b-CFS is not suitable for research related to laterization, particularly with dynamic stimuli in the domain of motion event processing. According to the feedback obtained from a brief follow-up question for the confirmation of effectiveness of the mask, some participants claimed that the stimulus animation was moving from one direction to the opposite (e.g., from the left corner to the right corner). Thus, although the mask and the stimuli were presented to separated eyes, the dynamic stimuli might move across the two visual fields, thus diminishing the effects of the laterization. Perhaps such ‘language brain’ (i.e., lateralisation) effects might be able to be witnessed in other studies that use static stimuli, such as those investigating ‘colour’ or ‘objects’.

5.4.3 Verbal interference task is not an option

The results obtained from participants with distinct language backgrounds in the three non-verbal experiments in the current study indicate that the influence of language is pervasive but also very flexible and sensitive. Particularly, when it is examined in non-verbal experiments that require high-level semantic analysis. Inappropriate measurement of preventing participants from using inner speech might eliminate any linguistic effects, confirming the danger of drawing conclusions about the effects of language on thought in non-verbal experiments created by implementing verbal interference tasks.

As discussed in earlier chapters, the rationale of the verbal interference task in a number of studies has been to use verbalisation to cut off the phonological loop formed by the inner speech during non-verbal experiments (Baddeley, 1992), thus preventing participants from using the language silently. However, the effect of verbal language labels can be highly flexible and extremely sensitive (Lupyan & Ward, 2013). For instance, Athanasopoulos *et al*. (2015) discovered that this verbal interference can be transformed into verbal assistance under certain circumstances. Specifically, its bilingual speakers were found to exhibit language-specific patterns corresponding to the language not used in the verbal interference task. In other words, letting Chinese – English bilinguals use Chinese in the verbal interference task (e.g., repeat numbers in Chinese) might block the phonological loop created by Chinese, but not the English which might be also employed during the non-verbal experiments. The potential use of English sub-vocally might be the reason that these bilinguals exhibit an English-like pattern (Wang & Li, 2021b).

In the domain of motion, verbal interference brings about more concerns. The major issue of examining the motion event processing demonstrated in earlier literature (Gennari *et al*., 2002; Papafragou *et al*., 2002; Trueswell & Papafragou, 2010) was the conflict between language-specific effects and the universal spatial influence rooted in dynamic motion cognition (Talmy, 2000; Radvansky & Zacks, 2014), and this is more evident in studies examining the *manner* and *path* (Ji, 2017; Wang & Li, 2021b; Papafragou *et al*., 2008). The language-specific effects lie in the *manner* and *path* encoding required to be activated by linguistic labels, including a verbal interference task in the non-verbal experiment, which might completely block the language effects required linguistic activation to exhibit non-verbal language-specific effects. This is confirmed by the results obtained in the present study. The absence of cross-linguistic differences in results obtained from non-verbal experiments conducted simultaneously with verbal interference tasks might indeed indicate an impairment of the research design, rather than nonexistence of the language effects. Therefore, a better approach for investigating the question of whether language effects persist without linguistic involvement should be those that detect low-level language signals (Thierry, 2016).

5.4.4 Underestimated high-level semantic analysis

Evidence obtained from non-verbal experiments has been criticised for being highly biased by the possible use of inner speech when claiming support for the LRH in the previous literature. This is due to the lack of direct evidence of cross-linguistic differences without the assistance of covert language use. The current study provides supporting evidence for the LRH as well, but the contribution is not because of the findings from the low-level detection experiment (i.e., b-CFS experiment) alone, given that it maximally excluded the conscious recruitment of language during non-verbal experiments. It was the consistent cross-linguistic differences obtained from participants in three language groups across all three non-verbal experiments, demonstrating the pervasive language-specific effects, thus arguably supporting the LRH (Athanasopoulos, 2011).

It is important to highlight that the cross-linguistic differences were observed in the two verification experiments as well, since they were major pieces of evidence illustrating the dynamic effects induced by the Chinese and English motion expressions. Specifically, the evidence found in the two verification experiments provides the valid pieces for the motion event processing puzzle. That is, although cross-linguistic differences were evident in the b-CFS experiment, the specific distinction between *manner* and *path* encoding in English speakers was not consistent with those in Ji (2017) and Wang & Li (2021b). The persistent cross-linguistic differences found in the VV and SV verification experiments increase the degrees of validity of the language-specific effects concluded in the b-CFS experiment.

Furthermore, the current study is also the first to apply a novel verification task in research concerning motion event processing (Zwaan *et al.*, 2002). The mixed findings in previous studies examining the motion lexicalization demonstrates that the language-specific differences are evident under specific conditions (see Chapter 2 for a detailed discussion). Generally speaking, in non-verbal experiments examining high-level semantic analysis, language-specific effects were only evident in similarity judgement tasks where language has been activated. No cross-linguistic differences were found in studies involving memory retrieval (Gennari *et al.*, 2002; Papafragou *et al.*, 2002; Engemann *et al.*, 2015). However, the findings obtained from the current verification experiments indicate that motion expressions influence motion memory, but it is easily affected by other high-level factors, such as general cognitive load (Gernnari *et al.*, 2002) and universal spatial preference in motion (Engmann *et al.*, 2015). The current verification paradigm examines motion memory and provides updated evidence for the specific effects induced by overt/covert linguistic labels.

5.5 Limitations

The current study also has several limitations due to various reasons, such as the challenges of conducting face-to-face experiments during the pandemic. As a result, part of the SV and VV verification experiments were moved to an online platform. Furthermore, limitations stemming from the lack of funding led to a restricted number of experimental stimuli. However, these limitations did not significantly affect the experimental results. Other potential limitations, such as the recruitment of L2 learners, the design of the stimuli and the equipment used for the b-CFS, may have introduced minor differences, but did not significantly affect overall results. Details of each limitation are explained below.

One limitation of the current study concerns L2 learner participants. Recall that L2 proficiency was not found to significantly affect the main detection or reaction times recorded in the b-CFS and verification experiments, respectively. It could be due to the limited variation of L2 proficiency in current L2 learners. Also, only IELTS scores were found to influence the RT results in L2 learners, despite also examining the OPT scores. One possible reason for this discrepancy is that the OPT used in the current study primarily assessed grammatical knowledge, whereas the IELTS provides a broader evaluation of language proficiency, encompassing speaking, writing, listening, and reading skills. Moreover, recall that in the b-CFS experiment, even advanced L2 learners did not fully resemble the (L2) English speakers in detecting the *manner* manipulation. It might be that the L2 learners do not completely mirror the English monolinguals since there are two language systems actively operating, and even though they can verbally describe in an English-like way.

The second limitation is the creation of suppression masks. For example, Pournaghdali & Schwartz (2020) claim that the efficiency of the suppression function exerted by the mandarin mask in b-CFS should be maintain maximally by creating unique masks for each unique stimulus, since colours, shapes might all influence the suppression power of the mask. In the current study, the motion event stimuli were 3D animation presenting real-life scenarios (cf. schematic spatial stimuli in Hong *et al*., 2015), and this is the first time that stimuli have been formed by complex information in this task. Unique masks were not created because in the difficulty of measuring the suppression efficiency of each unique mask. The masks used in the present study were examined in a separate control experiment, and the results confirmed that the efficiency was enough to suppress the motion stimuli for a relatively longer time. Within this time period, language-specific differences were observed. This might indeed be due to the saliency-induced by motion expressions in Chinese and English. Language effects were observed to emerge at an early stage of visual perception (Athanasopoulos & Casaponsa), and the b-CFS experiment extended the time frames of the earlier stage of visual detection and thus captured the language-specific differences.

The third limitation relates to the motion animation used in the present study. First, they were newly designed based on those used in previous studies (Ji, 2017; Wang & Li, 2021b), and this could have resulted inconsistent results compared to what has been reported in the previous literature. Moreover, only caused motion events were utilised in the animation design, which might bring concerns for the application of the current results to the domain of motion event processing in general. Another potential limitation might be that these newly designed stimuli were not examined in verbal description tasks, even though the SV verification task was intended to provide reference evidence for having overt linguistic labels for the same stimuli. Future studies might include a verbal description task subsequently when the examination was utilising novel stimuli and only non-verbal experiments were involved.

Another potential issue relates to the absence of language effects predicted by laterization. As discussed in earlier sections, this could have been caused by a combination of factors, including the use of the b-CFS paradigm and the dynamic characteristics of the motion events themselves, leading to conflicting results. Furthermore, it is possible that the uneven distribution of the data set presented to the left and right eyes contributed to this outcome, since the majority (70%) of participants had the right eye dominant. This imbalance in the presentation of the data to the "language brain" may have resulted in the absence of language-specific effects. In future studies, it could be beneficial to recruit an equal number of participants with right eye dominance and left eye dominance to mitigate this potential bias.

In sum, in this chapter, all the results obtained from the current study have been discussed. The interpretation of these findings focusses on the pervasive language-specific effects predicted by the LRH, motion typology, the salient English *manner* and universal spatial influence, cognitive restructuring in L2 learners, language effects and visual perception, and language-specific effects at different cognitive levels. Subsequently, topics related to the specific contributions to the application of the novel paradigms and the potential limitations of the present study and possible suggestions for future studies were discussed. The following chapter sets out the final conclusions of the current study.

Chapter 6: Conclusion

To understand the pervasive language-specific effects predicted by the linguistic relativity hypothesis in the domain of motion events, the current study employed three novel non-verbal experiments, including the b-CFS experiment, sentence-video verification, and video-video verification experiments, to examine the encoding of manner and path among monolingual Chinese and English, and L2 (native Chinese) speakers. These three experiments focus on different cognitive levels, demonstrating consistent and dynamic language-specific effects in motion event processing. Specifically, monolingual English and Chinese speakers were observed to exhibit distinct detection and verification patterns in all three experiments. Furthermore, these cross-linguistic differences vary in the specific experiments, reflecting a dynamic interaction between the influence of the uniformed spatial information in motion events and language-specific effects that lie in the motion components.

In sum, monolingual English speakers spent less time detecting the manner than the path manipulation in the b-CFS experiment, reflecting that language-specific effects emerge at an early stage in the cognitive processing of motion events. This asymmetric reaction time pattern between manner and path was observed, but in the opposite direction, in the sentence-video (SV) verification experiment, demonstrating consistent language-specific effects at a higher cognitive level. In contrast, the Chinese speakers performed similarly when detecting and verifying the manner and path manipulations in the b-CFS and SV verification experiments, respectively. However, this cross-linguistic difference was reversed in the video-video (VV) verification experiment, illustrating different language-specific effects also at a higher cognitive level, but without assistance from overt linguistic labels. Specifically, the Chinese speakers took significantly different amounts of time verifying video stimuli between the manner and path manipulations, whereas the English speakers exhibited a similar pattern, although this latter comparison was not statistically significant. In terms of the L2 learners, they are exhibited similar cognitive patterns in the SV and VV verification experiments compared to those observed in the b-CFS experiment. That is, advanced L2 learners showed an L1 influence at the lower cognitive level, whereas the intermediate L2 learners appeared to have shifted towards an English-based cognitive pattern at the higher cognitive level. In other words, when language recruitment was impossible, the L2 learners exhibit a cognitive pattern that was distinct from both Chinese and English monolinguals, and they continue to diverge from the monolingual Chinese speakers with the assistance from linguistic labels, or even with nonlinguistic cues.

The evidence obtained from both the L1 and L2 participants across all three experiments illustrates consistent language-specific effects, confirming the predictions proposed by the linguistic relativity hypothesis (Wholf, 1956). This is the first time that this has been found in motion event cognition, specifically in the encoding of motion components. This sheds light on the manner and path encodings in Chinese (equipollently-framed) and English (satellite-framed) at early and later stages of visual perception, aligning with the predictive processing account (Lupyan & Clark, 2015), and demonstrating the role of language on visual processing at distinct cognitive levels. Furthermore, the current study provides updated evidence for understanding the absence of language-specific effects in non-verbal experiments that have required semantic analysis and memory retrieval of motion events. Specifically, the observed language-specific effects were found to be persistent and flexible in the current study. The lack of similar evidence in earlier research may be due to the limitations of the methodological techniques that have been applied in these studies.

Several limitations are evident in the present study, such as the limited motion stimuli, and the current study only used caused motion events, perhaps it might be better to also include voluntary motion events in the future studies; another limitation is that the current study did not include a verbal description task for confirming the distinct motion typologies in Chinese and English expressions regarding the current motion event stimuli. Overall, and despite the limitations, it is hoped that this study will add to the field of motion event cognition and the role of language within it.

Appendices

Appendix 1

Language background information sheet

L2 learners

Please complete all questions relevant to you and mark tick boxes as appropriate.

1. Today's date: ____

2. Participant No.: ___________________________

3. Age: ____

4. Gender: □ Female □ Male

5. Nationality:________________

6. Mother tongue: ____

7. Linguistic dialect (which is the dialect that has most influenced the way you speak?):________________

8. Where do you currently live? Country: ______________________Town:__________________

9. How many years/months have you lived there? _______________________________________

10. EDUCATION:

Highest qualification to date (please specify) ____

11. KNOWLEDGE OF OTHER LANGUAGES

Do you speak any languages other than your mother tongue and English? □ YES □ NO

If YES, please give details:

Language: _____________________________Years of learning: ___________________

Level (tick one): □ beginner □ intermediate □ advanced □ fluent □ bilingual

Language: ______________________________ Years of learning: ___________________

Level (tick one): □ beginner □ intermediate □ advanced □ fluent □ bilingual

Language: ______________________________ Years of learning: ___________________

Level (tick one): □ beginner □ intermediate □ advanced □ fluent □ bilingual

12. At what age did you begin to learn English?_ ____

13. How long have you been studying English? Years: _____________

14. Way of learning English (tick one or more boxes):

□ At school / foreign language school Years: _____________

□ Private English language lessons Years: _____________

□ By exposure to an English-speaking environment Years: _____________

□ Grew up / lived in an English-speaking country. Please specify:____________________

15. Mother's first language__________

Has she lived in a foreign country for over a year?

□ YES □ NO

16. Father's first language__________

Has he lived in a foreign country for over a year?

□ YES □ NO

17. Do you watch English-speaking TV / films without subtitles?

□ YES, hours per week:_______ □NO

18. Do you watch English-speaking TV / films with subtitles?

□ YES, hours per week: ______ □NO

19. Do you listen to English-speaking radio / online programmes?

□ YES, hours per week:_____ □NO

20. Do you read English language books / magazines / newspapers / online texts outside

university?

□ YES, hours per week: ______ □ NO

21. Have you taken any international exam in English?

□ YES □ NO If YES, please give details:

Test (e.g. IELTS) Date when test taken:

Overall grade:

22. In a typical day, what percentage of English/mother tongue/other language(s) do you speak or

write?

Speaking English ______% Mother tongue ___% Other language(s) ____% Total: %

Writing English ______% Mother tongue ___% Other language(s) ____% Total: %

23. How often do you socialise with English native speakers outside of class?

□ ALWAYS (every day) □ OFTEN (3-4x a week)

□ RARELY (3-4 x a month) □NEVER

24. Are you right-handed or left-handed?

□ Right-handed □ Left-handed

Monolinguals

1. Today's date:

____

2. Participant No.: ___________________________

3. Age: ____

4. Gender: ☐ Female ☐ Male

5. Nationality:_____________

6. Mother tongue:

____

7. Where do you currently live? Country:

__________________Town:_______________

8. How many years/months have you lived there?

____

9. EDUCATION:

Highest qualification to date (please specify)

____

10. KNOWLEDGE OF OTHER LANGUAGES

Do you speak any languages other than your mother tongue? ☐ YES ☐ NO

If YES, please give details:

Language: ________________________Years of learning: ________________

Level (tick one): ☐ beginner ☐ intermediate ☐ advanced ☐ fluent ☐ bilingual Language: _________________________ Years of learning: ________________

Level (tick one): ☐ beginner ☐ intermediate ☐ advanced ☐ fluent ☐ bilingual Language: _________________________ Years of learning: ________________

Level (tick one): ☐ beginner ☐ intermediate ☐ advanced ☐ fluent ☐ bilingual

11. Mother's first language_________Has she lived in a foreign country for over a year?

☐ YES ☐ NO

12. Father's first language__________Has he lived in a foreign country for over a year?

☐ YES ☐ NO

24. Are you right-handed or left-handed?

☐ Right-handed ☐ Left-handed

Appendix 2

Video Stimuli list

1. A man is **carrying/pulling** a suitcase **into/out** of a room.
2. A man is **pushing/rolling** a tyre **along/across** a double yellow line.
3. A man is **rolling/carrying** a log **towards/away from** a cabin.
4. A man is **rolling/pushing** a haystack **around/across** a pit.
5. A man is **carrying/rolling** a snowball **up/down** a hill.
6. A man is **pushing/carrying** a giftbox **into/out of** a truck.
7. A man is **carrying/pushing** a wheelchair **towards/away from** a house.
8. A man is **pulling/carrying** a skateboard **away from/towards** a skate park.
9. A man is **pulling/dragging** a toy-car **along/across** a river.
10. A man is **dragging/carrying** a sack **out of/into** a cave.
11. A man is **dragging/pulling** a cart **across/around** a garden.
12. A man is **carrying/dragging** a toy-bear **across/around** a playground.

Appendix 3

Quick placement task

Start of Block: Quick placement task

Participant No. Please fill your participant number in the following box :

____

Q1
PART 1
Water __________ at a temperature of 0°C .

- ○ be freezing (1)
- ○ is freezing (2)
- ○ freezes (3)

Q2 In some countries ___________ dark all the time in winter.

- ○ there is (1)
- ○ is (2)
- ○ it is (3)

Q3 In hot countries people wear light clothes __________ cool.

- ○ for keeping (1)
- ○ to keep (2)
- ○ for to keep (3)

Q4 In Madeira they have __________ weather almost all year.

- the good (1)
- good (2)
- a good (3)

Q5 Most Mediterranean countries are __________ in October than in April.

- more warm (1)
- the more warm (2)
- warmer (3)

Q6 Parts of Australia don't have _________ rain for long periods.

- the (1)
- some (2)
- any (3)

Q7 In the Arctic and Antarctic ________ a lot of snow.

- ○ it is (1)
- ○ there is (2)
- ○ it has (3)

Q8 Climate is very important in _________ people's lives.

- ○ most of (1)
- ○ most (2)
- ○ the most (3)

Q9 Even now there is _________ we can do to control the weather.

- ○ little (1)
- ○ few (2)
- ○ less (3)

Q10 In the future ________ to get a lot of power from the sun and the wind.

- ○ we'll need (1)
- ○ we are needing (2)
- ○ we can need (3)

Q11 For many people the name Pelé still means ________ famous footballer in the world.

- ○ the more (1)
- ○ the most (2)
- ○ most (3)

Q12 Pelé _______ born in 1940.

- ○ had been (1)
- ○ is (2)
- ○ was (3)

Q13 His mother _______ him to become a footballer.

- ○ not want (1)
- ○ wasn't wanting (2)
- ○ didn't want (3)

Q14 But his father ________ practice every day.

- made him to (1)
- made him (2)
- would make him to (3)

Q15 By 1956 he _______ the Brazilian clubs, *Santos*, and had scored in his first game.

- has joined (1)
- joined (2)
- had joined (3)

Q16 In 1957 he _________ for the Brazilian national team.

- has been picked (1)
- was picked (2)
- was picking (3)

Q17 The next World Cup Finals were in 1958 and Pelé was looking forward to ______________.

- ◯ play (1)
- ◯ playing (2)
- ◯ the play (3)

Q18 And _______________ he was injured he helped Brazil to win the final.

- ◯ even though (1)
- ◯ even so (2)
- ◯ in spite of (3)

Q19 Pelé was ____________ brilliant player that he helped Brazil win 3 World Cups.

- ◯ a such (1)
- ◯ such a (2)
- ◯ a so (3)

Q20 He didn't stop __________ for *Santos* till he was 34.

- ○ playing (1)
- ○ to play (2)
- ○ play (3)

Q21 After calling it a day in 1974, he came __________ retirement and played for *New York Cosmos*.

- ○ from (1)
- ○ off (2)
- ○ out of (3)

Q22 _________ the end of his career he had scored over a thousand goals.

- ○ Till (1)
- ○ By (2)
- ○ In (3)

Q23 He then settled for a role ________ a sporting ambassador for Brazil.

- as (1)
- like (2)
- in (3)

Q24 By the end of the 20th Century he had received a great _________ of awards.

- many (1)
- number (2)
- deal (3)

Q25 Though honoured with the title *Athlete of the century*, he will always be remembered __________ .

- as footballer (1)
- as a footballer (2)
- as the footballer (3)

Q26 Football, or soccer as it is sometimes known, ___________ played

- has been (1)
- is being (2)
- was (3)

Q27 for _______ 150 years, but the first World Cup

- above (1)
- over (2)
- more that (3)

Q28 competition _________ held until 1930,

- has not been (1)
- was not (2)
- was not being (3)

Q29 when Uruguay ________ the first professional final.

- could win (1)
- were winning (2)
- won (3)

Q30 Four teams had entered from Europe, but with __________ success.

- a little (1)
- little (2)
- few (3)

Q31 The 1934 World Cup was again won by __________ home team, Italy,

- a (1)
- the (2)
- their (3)

Q32 ___________ went on to win the 1938 final as well. Winning successive

- who (1)
- which (2)
- that (3)

Q33 finals is something that __________ achieved again

- is not (1)
- was not (2)
- has not been (3)

Q34 until Brazil managed __________ in 1958 and 1962.

- them (1)
- these (2)
- it (3)

Q35 If Brazil _________ again in 1966 then the FIFA

- would have won (1)
- would win (2)
- had won (3)

Q36 authorities would have needed to ___________ the original World Cup replaced.

- have (1)
- let (2)
- make (3)

Q37 However, England stopped the Brazilians __________ a third successive win.

- to get (1)
- getting (2)
- get (3)

Q38 In the 1970s the honours were shared ___________ Europe and South America.

- among (1)
- between (2)
- inside (3)

Q39 Argentina succeeded __________ in 1978, but in 1982, in Spain,

- ○ to win (1)
- ○ at winning (2)
- ○ in winning (3)

Q40 they had __________ getting beyond the early stages.

- ○ difficulty in (1)
- ○ difficulties to (2)
- ○ difficulty to (3)

Q41 They won again in Mexico in 1986, ___________ Maradonna

- ○ where (1)
- ○ which (2)
- ○ while (3)

Q42 managed to win ____________ of the games, especially the one

- ○ much (1)
- ○ some (2)
- ○ any (3)

Q43 against England, almost __________ . The 1990s finals were

- by his own (1)
- by himself (2)
- on himself (3)

Q44 dominated by European teams __________ from Brazil's win in the USA in 1994,

- except (1)
- apart (2)
- save (3)

Q45 with the 1998 finals in France again __________ won by the hosts.

- to be (1)
- being (2)
- having (3)

Q46 Throughout the 1990s police in the host countries ___________ kept busy keeping

- was (1)
- were (2)
- have been (3)

Q47 rival fans apart, but __________ to be no such problems when the first

- there was (1)
- there were (2)
- it was (3)

Q48 World Cup Finals of the 21st century took ___________ in Japan and South Korea in 2002.

- part (1)
- place (2)
- hold (3)

Q49 Football's third century ____________ success for a number of footballing nations in

- has seen (1)
- saw (2)
- seeing (3)

Q50 Africa and Asia, who __________ prove to be the teams of the future.

- may well (1)
- may as well (2)
- might as well (3)

Q51 PART 2

Millions of __________ around the world now use the Internet almost every day.

- persons (1)
- people (2)
- peoples (3)

Q52 The majority of children in the UK __________ access to a PC.

- have (1)
- has (2)
- are having (3)

Q53 Learning to use the Internet is not the same ___________ learning traditional skills.

- as (1)
- like (2)
- than (3)

Q54 Most of us start off with email, ____________ is fairly easy to use.

- who (1)
- which (2)
- what (3)

Q55 Children generally find using computers easy, but some adults can't get used __________ with them.

- to work (1)
- to working (2)
- work (3)

Q56 There aren't __________ shortcuts to becoming proficient -- everyone needs training and practice.

- no (1)
- any (2)
- some (3)

Q57 Those who do best are those who also use computers a lot __________ .

- on their own (1)
- by their own (2)
- on themselves (3)

Q58 It's no use ___________ to become an expert just by reading books.

- ○ in trying (1)
- ○ to try (2)
- ○ trying (3)

Q59 There are many who wish they ___________ learning earlier.

- ○ started (1)
- ○ would have started (2)
- ○ had started (3)

Q60 A few unsuccessful learners have resigned themselves to never ___________ how to use the Internet.

- ○ know (1)
- ○ knowing (2)
- ○ known (3)

Q61 Some new users quickly become almost addicted ___________ on line.

- to be (1)
- to being (2)
- be (3)

Q62 Others decide they would just ____________ not have anything to do with computers.

- rather (1)
- prefer (2)
- better (3)

Q63 The trend continues ____________ for computers to get smaller and smaller.

- to be (1)
- be (2)
- by being (3)

Q64 Some companies already have more palmtops ____________ desktops.

- that (1)
- than (2)
- as (3)

Q65 It is thought that we'll have mobile phones as powerful as PCs ___________ the end of the decade.

- till (1)
- by (2)
- in (3)

Q66 **Below is a letter written to the 'advice' column of a daily newspaper. Select the correct answers.**

Dear Marge,

___________ to you because I

- I'm writing (1)
- I will write (2)
- I should write (3)

Q67 __________ what to do I'm twenty-six and a teacher at

- am not knowing (1)
- don't know (2)
- know not (3)

Q68 a primary school in Norwich where __________ for the last five years.

- I'm working (1)
- I've worked (2)
- I work (3)

Q69 When I _________ there for a couple of years, one of the older members of staff

- was (1)
- have been (2)
- had been (3)

Q70 __________ and a new teacher

- would leave (1)
- left (2)
- had been leaving (3)

Q71 __________ appointed to work in the same department as me.

- would be (1)
- became (2)
- was (3)

Q72 We ___________ together with the same classes during her first year

- worked (1)
- have worked (2)
- should work (3)

Q73 and had the ___________ up a good professional

- opportunity for building (1)
- possibilities to build (2)
- chance to build (3)

Q74 relationship. Then, about eighteen months after __________

- she has arrived (1)
- to have arrived (2)
- arriving (3)

Q75 in Norwich, she decided to buy _________ house.

- her own (1)
- herself (2)
- her a (3)

Q76 She was tired of __________ in rented accommodation and wanted a place

- to live (1)
- live (2)
- living (3)

Q77 __________ . At about the same time, I

- by her own (1)
- of her own (2)
- of herself (3)

Q78 __________ notice by the landlord of the flat

- was given (1)
- have been given (2)
- gave (3)

Q79 __________ in

- what I was living (1)
- that I had lived (2)
- I was living (3)

Q80 and she asked me if I _________ to live

- liked (1)
- had liked (2)
- would like (3)

Q81 with her. She __________ me that by the time she

- said (1)
- told (2)
- explained (3)

Q82 _________ the mortgage

- would pay (1)
- would have paid (2)
- had paid (3)

Q83 and the bills ________ wouldn't be

- it (1)
- there (2)
- they (3)

Q84 __________ left to live on. She suggested

- a lot (1)
- many (2)
- few (3)

Q85 __________ share the house and share the costs.

- us to (1)
- we should (2)
- we may (3)

Q86 It seemed like a good idea, so after __________ all the details

- we'd agreed (1)
- we could agree (2)
- we agreed with (3)

Q87 __________ needed to be sorted out, we moved into the new house together.

- what (1)
- that (2)
- who (3)

Q88 At the end of this month _________

- we have lived (1)
- we have been living (2)
- we'll have been living (3)

Q89 together for a year and a half. It's the first time __________ with anybody before, but

- I live (1)
- I'm living (2)
- I've lived (3)

Q90 __________ what would happen. I've fallen in love with her and now she's been offered another job 200 miles away and is going to move. I don't know what to do. Please give me some advice.
Yours in shy desperation,
Steve

- I should guess (1)
- I might have guessed (2)
- I'd have guessed (3)

Q91 **Now choose the correct question tag in the following 10 questions:**

Steve's off to China, ____________ ?

- has he (1)
- hasn't he (2)
- isn't he (3)

Q92 It'll be a year before we see him again, ___________ ?

- won't it (1)
- won't we (2)
- shan't it (3)

Q93 I believe he's given up smoking, _____________ ?

- isn't he (1)
- don't I (2)
- hasn't he (3)

Q94 I'm next to the list to go out there, _____________ ?

- am not I (1)
- are I (2)
- aren't I (3)

Q95 No doubt you'd rather he didn't stay abroad too long, ____________ ?

- shouldn't you (1)
- wouldn't you (2)
- hadn't you (3)

Q96 He's rarely been away for this long before, ____________ ?

- is he (1)
- hasn't he (2)
- has he (3)

Q97 So you think he'll be back before November, __________ ?

- shall he (1)
- will he (2)
- do you (3)

Q98 Nobody's disagreed with the latest proposals, __________ ?

- did he (1)
- has he (2)
- have they (3)

Q99 We'd better not delay reading this any longer, __________ ?

- should we (1)
- did we (2)
- had we (3)

Q100 Now's hardly the time to tell me you didn't need a test at all, _________ ?

- did you (1)
- is it (2)
- isn't it (3)

End of Block: Quick placement task

Appendix 4

List of prime sentences:

English:

1. A man is carrying a suitcase into of a room.
2. A man is pushing a tyre along a double yellow line.
3. A man is rolling a log towards a cabin.
4. A man is rolling a haystack around a pit.
5. A man is carrying a snowball up a hill.
6. A man is pushing a giftbox into a truck.
7. A man is carrying a wheelchair towards a house.
8. A man is pulling a skateboard away from a skate park.
9. A man is pulling a toy-car along a river.
10. A man is dragging a sack out of a cave.
11. A man is dragging a cart across a garden.
12. A man is carrying a toy-bear across a playground.

Chinese:

1. 一个男人把箱子搬进屋子。
2. 一个男人把轮胎沿着黄线推。
3. 一个男人把木桩推去小木屋。
4. 一个男人把干草垛卷着绕坑地。
5. 一个男人把雪球滚上山。
6. 一个男人把箱子推进卡车。
7. 一个男人把轮椅搬去房子。
8. 一个男人把滑板拉离公园。
9. 一个男人把玩具车拉过一条河。
10. 一个男人把麻袋拽出洞穴。
11. 一个男人把推车拽过公园。
12. 一个男人把玩具熊抱着穿过公园。

Appendix 5

LMM models

1. Difference RT ~ 1 + condition * group + (1 + condition | participant) + (1 + condition * group | item)
2. logRT ~1 + mask + (1 + mask | subject) + (1 + mask | item)
3. logRT ~1 + (1 + mask | subject) + (1 + mask | item)
4. logRT ~1 + condition * group + (1 + condition | participant) + (1 | item)
5. logRT ~1 + condition + (1 + condition | participant) + (1 | item)
6. logRT ~1 + group + (1 + condition | participant) + (1 | item)
7. logRT ~1 + condition * group *mask + (1 | participant) + (1 | item)
8. logRT ~1 + condition * group + (1 | participant) + (1 | item)
9. logRT ~1 + condition * group + (1 | participant) + (1 | item)
10. logRT ~1 + condition+ (1 | participant) + (1 | item)
11. logRT ~1 + group + (1 | participant) + (1 | item)
12. logRT ~1 + condition + group + (1 | participant) + (1 | item)
13. logRT ~1 + condition * context + (1 | participant) + (1 | item)
14. logRT ~1 + condition + (1 | participant) + (1 | item)
15. logRT ~1 + group + (1 | participant) + (1 | item)
16. logRT ~1 + condition + group + (1 | participant) + (1 | item)
17. logRT ~1 + condition * group + (1 | participant) + (1 | item)
18. logRT ~1 + condition + (1 | participant) + (1 | item)
19. logRT ~1 + group + (1 | participant) + (1 | item)
20. logRT ~1 + condition + group + (1 | participant) + (1 | item)
21. logRT ~1 + condition + IELTS scores + OPT scores + length of stay in the UK + AO + daily use of speaking English + daily use of writing English + (1 | participant) + (1 | item)
22. logRT ~1 + condition * group + (1 | participant) + (1 | item)
23. logRT ~1 + condition + (1 | participant) + (1 | item)
24. logRT ~1 + group + (1 | participant) + (1 | item)
25. logRT ~1 + condition + group + (1 | participant) + (1 | item)
26. logRT ~1 + condition * group + (1 | participant) + (1 | item)
27. logRT ~1 + condition + (1 | participant) + (1 | item)
28. logRT ~1 + condition + group + (1 | participant) + (1 | item)
29. logRT ~1 + group + (1 | participant) + (1 | item)
30. logRT ~1 + condition * group + (1 + condition | participant) + (1 + group | item)
31. logRT ~1 + condition + (1 + condition | participant) + (1 + group | item)
32. logRT ~1 + group + (1 + condition | participant) + (1 + group | item)

33. logRT ~1 + condition + group + (1 + condition | participant) + (1 + group | item)
34. logRT ~1 + condition * group + (1 | participant) + (1 + group | item)
35. logRT ~1 + group + (1 | participant) + (1 + group | item)
36. logRT ~1 + condition + (1 | participant) + (1 + group | item)
37. logRT ~1 + condition + group + (1 | participant) + (1 + group | item)
38. logRT ~1 + condition * group + (1 + condition | participant) + (1 + group | item)
39. logRT ~1 + group + (1 + condition | participant) + (1 + group | item)
40. logRT ~1 + condition + (1 + condition | participant) + (1 + group | item)
41. logRT ~1 + condition + group + (1 + condition | participant) + (1 + group | item)
42. logRT ~1 + condition + IELTS scores + OPT scores + length of stay in the UK + AO + daily use of speaking English + daily use of writing English + (1 | participant) + (1 | item)
43. logRT ~1 + condition * group + (1 | participant) + (1 | item)
44. logRT ~1 + condition + (1 | participant) + (1 | item)
45. logRT ~1 + group + (1 | participant) + (1 | item)
46. logRT ~1 + condition + group + (1 | participant) + (1 | item)
47. logRT ~1 + condition * group + (1 + condition | participant) + (1 + group | item)
48. logRT ~1 + condition + (1 + condition | participant) + (1 + group | item)
49. logRT ~1 + group + (1 + condition | participant) + (1 + group | item)
50. logRT ~1 + condition + group + (1 + condition | participant) + (1 + group | item)
51. logRT ~1 + condition * group + (1 + condition | participant) + (1 | item)
52. logRT ~1 + condition + (1 + condition | participant) + (1 | item)
53. logRT ~1 + group + (1 + condition | participant) + (1 | item)
54. logRT ~1 + condition + group + (1 + condition | participant) + (1 | item)
55. logRT ~1 + condition * group + (1 | participant) + (1 | item)
56. logRT ~1 + condition + (1 | participant) + (1 | item)
57. logRT ~1 + group + (1 | participant) + (1 | item)
58. logRT ~1 + condition + group + (1 | participant) + (1 | item)
59. logRT ~1 + condition + IELTS scores + OPT scores + length of stay in the UK + AO + daily use of speaking English + daily use of writing English + (1 | participant) + (1 | item)

Appendix 6

Investigating the conceptual changes of motion event cognition in advanced Chinese learners of English: a breaking Continuous Flash Suppression study

Consent Form

Please tick each box if you are happy to take part in this research.

I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve me taking part as described above.	
I understand that participation in this study is voluntary.	
I understand that my data will not be identifiable and the data may be used in publications, presentations and online.	
I confirm that I have read the information about GDPR	
I understand that anonymised data will be made available to the research community on the Open Science Framework website for secondary analyses.	
I understand that I can withdraw my data at any point during data collection and up to one week after data collection.	

NAME____

SIGNATURE____

DATE____

University of York Information on GDPR

Processing personal data

Under the General Data Protection Regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data.

In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1)(e) of the GDPR:

Processing is necessary for the performance of a task carried out in the public interest

Special category data is processed under Article 9 (2) (j):

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research will only be undertaken where ethical approval has been obtained, where there is a clear public interest and where appropriate safeguards have been put in place to protect data.

In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

Protecting and storing personal data

Information that research participants provide will be treated confidentially and shared on a need-to-know basis only. The University is committed to the principle of data protection by design and default and will collect the minimum amount of data necessary for the project. In addition we will anonymise or pseudonymise data wherever possible.

We will put in place appropriate technical and organisational measures to protect your personal data and/or special category data (for example, data may be stored in secure filing cabinets and/or on a password protected computer).

Sharing of data

The default position is that personal data will only be accessible to members of the project team. In some cases, however, the research may be of a collaborative nature and hence the data will be made accessible to others from outside the

University. Information specific to the project will include details of when this is the case, who the 3rd parties are, and what they will do with the data. It is possible that personal data may be shared anonymously with others for secondary research and/or teaching purposes.

Transfer of data internationally

The default position is that data will be stored on University devices and held within the European Economic Area in full compliance with data protection legislation.

However, data may be transferred to the project partners based outside the European Economic Area. Any international transfer will be undertaken in full compliance with the GDPR.

The University has access to cloud storage provided by Google which means that data can be located at any of Google's globally spread data centres. The University has data protection compliant arrangements in place with this provider. For further information see,

https://www.york.ac.uk/it-services/google/policy/privacy/

Your rights in relation to your data

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdrawal. Please note, not all rights apply where data is processed purely for research purposes. For information see, https://www.york.ac.uk/records-management/dp/individualsrights/

Right to complain

If you are unhappy with the way in which your personal data has been handled, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see www.ico.org.uk/concerns

Reference

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

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language*, *60*(2), 270–290. https://doi.org/10.1016/j.jml.2008.10.001

Athanasopoulos, P., & Albright, D. (2016). A perceptual learning approach to the Whorfian hypothesis: supervised classification of motion. *Language Learning*, *66*(3), 666-689. https://doi.org/10.1111/lang.12180

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*(1), 89–96. https://doi.org/10.1017/S1366728905002397

Athanasopoulos, P. (2007). Interaction between grammatical categories and cognition in bilinguals: The role of proficiency, cultural immersion, and language of instruction. *Language and Cognitive Processes*, *22*(5), 689–699. https://doi.org/10.1080/01690960601049347

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: the case of Greek blues, *Bilingualism: Language and Cognition*, *12*(1), 83-95.

https://doi.org/10.1017/S136672890800388X

Athanasopoulos, P. (2011). Cognitive restructuring in bilingualism. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 29-65). Bristol: Multilingual Matters.

Athanasopoulos, P. (2012). Linguistic Relativity and Second Language Acquisition. In *The Encyclopedia of Applied Linguistics*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405198431.wbeal0722

Athanasopoulos, P., & Aveledo, F. (2012). Linguistic relativity and bilingualism. In Altarriba, J., & Isurin, L. (Eds.), *Memory, language, and bilingualism: Theoretical and applied approaches*. (p.236-255). Cambridge: Cambridge University Press

Athanasopoulos, P., & Bylund, E. (2013a). 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. (2013b). The 'thinking' in thinking-for-speaking: Where is it? *Language, Interaction and Acquisition*, *4*(1), 91–100. https://doi.org/10.1075/lia.4.1.05ath

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

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5–6), 393–412.

https://doi.org/10.1080/02643294.2020.1769050

Athanasopoulos, P., Damjanovic, L., Krajciova, A., & Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition*, *14*(1), 9–17. https://doi.org/10.1017/S1366728909990046

Athanasopoulos, P., & Kasai, C. (2008). Language and thought in bilinguals: The case of grammatical number and nonverbal classification preferences. *Applied Psycholinguistics*, *29*(1), 105–123. https://doi.org/10.1017/S0142716408080053

Aveledo, F., & Athanasopoulos, P. (2016). Second language influence on first language motion event encoding and categorization in Spanish-speaking children learning L2 English. *International Journal of Bilingualism*, *20*(4), 403–420. https://doi.org/10.1177/1367006915609235

Aveledo, F., & Athanasopoulos, P. (2023). Bidirectional cross-linguistic influence in motion event conceptualisation in bilingual speakers of Spanish and English. *International Review of Applied Linguistics in Language Teaching*, *61*(1), 13–36. https://doi.org/10.1515/iral-2022-0179

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

Baddeley, A. (1992), Working Memory: The Interface between Memory and Cognition. *Journal of Cognitive Neuroscience, 4*(3): 281–288. https://doi.org/10.1162/jocn.1992.4.3.281

Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition*, *111*(3), 329–344. https://doi.org/10.1016/j.cognition.2009.02.008

Bassetti, B., Clarke, A., & Trenkic, D. (2018). The linguistic transparency of first language calendar terms affects calendar calculations in a second language. *Acta Psychologica*, *186*, 81–89. https://doi.org/10.1016/j.actpsy.2018.04.006

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

Baumann, L., & Valuch, C. (2022). Priming of natural scene categorization during continuous flash suppression. *Consciousness and Cognition*, *104*, 103387. https://doi.org/10.1016/j.concog.2022.103387

Berthele, R., & Stocker, L. (2017). The effect of language mode on motion event descriptions in German–French bilinguals. *Language and Cognition*, *9*(4), 648–676. https://doi.org/10.1017/langcog.2016.34

Bohnemeyer, J. (2020). Linguistic Relativity: From Whorf to Now. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics* (1st ed., pp. 1–33). Wiley. https://doi.org/10.1002/9781118788516.sem013

Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

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

Brown, R. (1976). Reference: In Memorial Tribute to Eric Lenneberg. *Cognition*, *4*(2), 125–153.

Brown, A., & Gullberg, M. (2011). Bidirectional cross-linguistic influence in event conceptualization? Expressions of Path among Japanese learners of English. *Bilingualism: Language and Cognition*, *14*(1), 79–94. https://doi.org/10.1017/S1366728910000064

Bylund, E. (2009). Effects of age of L2 acquisition on L1 event conceptualization patterns*. *Bilingualism: Language and Cognition*, *12*(3), 305–322. https://doi.org/10.1017/S1366728909990137

Bylund, E. (2011a). Segmentation and temporal structuring of events in early Spanish-Swedish bilinguals. *International Journal of Bilingualism*, *15*(1), 56–84. https://doi.org/10.1177/1367006910379259

Bylund, E. (2011b). Language-specific patterns in event conceptualization: Insights from bilingualism. In A. Pavlenko (ed.), *Thinking and speaking in two languages* (pp. 108-142). Clevedon: Multilingual Matters.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does first language maintenance hamper nativelikeness in a second language: A Study of Ultimate Attainment in Early Bilinguals. *Studies in Second Language Acquisition*, *34*(2), 215–241. https://doi.org/10.1017/S0272263112000034

Bylund, E., & Athanasopoulos, P. (2014a). Language and thought in a multilingual context: The case of isiXhosa. *Bilingualism: Language and Cognition*, *17*(2), 431–441. https://doi.org/10.1017/S1366728913000503

Bylund, E., & Athanasopoulos, P. (2014b). Linguistic Relativity in SLA: Toward a New

Research Program. *Language Learning*, *64*(4), 952–985. https://doi.org/10.1111/lang.12080

Bylund, E., & Athanasopoulos, P. (2015). Introduction: Cognition, Motion Events, and SLA. *The Modern Language Journal*, *99*(S1), 1–13. https://doi.org/10.1111/j.1540-4781.2015.12175.x

Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics*, *51*(5). https://doi.org/10.1515/ling-2013-0033

Bylund, E., Hyltenstam, K., & Abrahamsson, N. (2021). Age of acquisition – not bilingualism – is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism: Language and Cognition*, *24*(1), 18–30. https://doi.org/10.1017/S1366728920000188

Bylund, E., & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition*, *14*(1), 47–59. https://doi.org/10.1017/S1366728910000180

Casasanto, D. (2016). Linguistic relativity. In N. Riemer (Ed.), *Routledge handbook of semantics* (pp. 158-174). New York: 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. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Conference Cognitive Science Society* (pp. 575-580). Austin, TX: Cognitive Science Society.

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

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509. https://doi.org/10.1017/S0140525X08004998

Clark, E. V. (2003). Languages and Representations. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind* (pp. 17–24). The MIT Press. https://doi.org/10.7551/mitpress/4117.003.0006

Cook, V. (2003). ‘The changing L1 in the L2 user’s mind. In V. J. Cook (Ed.), *Effects of the Second Language on the First* (pp. 1-18). Clevedon: Multilingual Matters.

Cook, V. (2015). Discussing the Language and Thought of Motion in Second Language Speakers. *The Modern Language Journal*, *99*(S1), 154–164. https://doi.org/10.1111/j.1540-4781.2015.12184.x

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. A. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism*, *10*(2), 137–152. https://doi.org/10.1177/13670069060100020201

Dolscheid, S., Çelik, S., Erkan, H., Küntay, A., & Majid, A. (2020). Space-pitch associations differ in their susceptibility to language. *Cognition*, *196*, 104073. https://doi.org/10.1016/j.cognition.2019.104073

Everett, C. (2013). *Linguistic relativity: Evidence across languages and cognitive domains*. De Gruyter Mouton.

Feinmann, D. (2020). Language and Thought in the Motion Domain: Methodological Considerations and New Empirical Evidence. *Journal of Psycholinguistic Research*, *49*(1), 1–29. https://doi.org/10.1007/s10936-019-09668-5

Feist, & Férez. (2013). Remembering How: Language, Memory, and the Salience of Manner. *Journal of Cognitive Science*, *14*(4), 379–398. https://doi.org/10.17791/JCS.2013.14.4.379

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

Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism*, *15*(4), 466–485. https://doi.org/10.1177/1367006911403062

Filipovic, L. (2018). Speaking in a second language but thinking in the first language: Language specific effects on memory for causation events in English and Spanish. *International Journal of Bilingualism*, 22(2), 180-198.

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., Carroll, M., Weimar, K., & Von Stutterheim, C. (2015). Driving Along the Road or Heading for the Village? Conceptual Differences Underlying Motion Event Encoding in French, German, and FrenchGerman L2 Users. *The Modern Language Journal*.

99(S1), 100-122

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

Forder, L., Taylor, O., Mankin, H., Scott, R. B., & Franklin, A. (2016). Colour Terms Affect Detection of Colour and Colour-Associated Objects Suppressed from Visual Awareness. *PLOS ONE*, *11*(3), e0152212. https://doi.org/10.1371/journal.pone.0152212

Francken, J. C., Kok, P., Hagoort, P., & De Lange, F. P. (2015). The Behavioral and Neural Effects of Language on Motion Perception. *Journal of Cognitive Neuroscience*, *27*(1), 175–184. https://doi.org/10.1162/jocn_a_00682

Gayet, S., Van der Stigchel, S., & Paffen, C. L. E. (2014). Breaking continuous flash suppression: Competing for consciousness on the pre-semantic battlefield. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00460

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

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Gilbert, A., Regier, T., Kay, P., & Ivry, R. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98. https://doi.org/10.1016/j.bandl.2007.06.001

Gleitman, L. R., & Papafragou, A. (2012). New Perspectives on Language and Thought. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (1st ed., pp. 543–568). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0028

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149. https://doi.org/10.1017/S136672899800025X

Gumperz, J. J., & Levinson, S. C. (Eds.). (1996). *Rethinking linguistic relativity*. Cambridge University Press.

Hong, S. W. (2015). Radial bias for orientation and direction of motion modulates access to visual awareness during continuous flash suppression. *Journal of Vision*, *15*(1), 3–3. https://doi.org/10.1167/15.1.3

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62(2), 169–200. https://doi.org/10.1016/S0010-0277(96)00784-6

Jarvis, S. (2007). Theoretical and Methodological Issues in the Investigation of Conceptual Transfer. *Vigo International Journal of Applied Linguistics*, *4*, Article 4.

Jarvis, S., & Pavlenko, A. (2010). *Crosslinguistic influence in language and cognition* (Paperback edition). Routledge.

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 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

Ji, Y., & Hohenstein, J. (2014b). 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. (2018). English and Chinese children's motion event similarity judgments. *Cognitive Linguistics*, *29*(1), 45–76. https://doi.org/10.1515/cog-2016-0151

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

Kamenetski, A., Lai, V. T., & Flecken, M. (2022). Minding the manner: Attention to motion events in Turkish–Dutch early bilinguals. *Language and Cognition*, *14*(3), 456–478. https://doi.org/10.1017/langcog.2022.10

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (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*, 638–653. https://doi.org/10.1037/a0020507

Lai, V. T., Rodriguez, G. G., & Narasimhan, B. (2014). Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition*, *17*(1), 139–152. https://doi.org/10.1017/S1366728913000151

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press.

Levinson, S. C. (2012). Foreword. In J. B. Carroll, S. C. Levinson, & P. Lee (Eds.), *Language, thought, and reality: selected writings of Benjamin Lee Whorf* (2nd ed.) (pp. vii-xxiii). Cambridge: MIT Press.

Levinson, S. C., & Majid, A. (2014). Differential Ineffability and the Senses. *Mind & Language*, *29*(4), 407–427. https://doi.org/10.1111/mila.12057

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

Liao, Y., Flecken, M., Dijkstra, K., & Zwaan, R. A. (2020). Going places in Dutch and mandarin Chinese: Conceptualising the path of motion cross-linguistically. *Language, Cognition and Neuroscience*, *35*(4), 498–520. https://doi.org/10.1080/23273798.2019.1676455

Lucy, J., A. (1992a). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*: Cambridge University Press.

Lucy, J., A. (1992b). *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*: Cambridge University Press.

Lucy, J. A. (1997). Linguistic Relativity. *Annual Review of Anthropology*, *26*(1), 291–312. https://doi.org/10.1146/annurev.anthro.26.1.291

Lucy, J. A., and Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman and S. C. Levinson (Eds.), *Language acquisition and conceptual developmen*t (pp. 257-283). Cambridge: Cambridge University Pres

Lucy, J. A. (2014). Methodological approaches in the study of linguistic relativity: Corpus method and cognitive theory. In L. Filipović & M. Pϋtz (Eds.), *Human Cognitive Processing* (Vol. 44, pp. 17–44). John Benjamins Publishing Company. https://doi.org/10.1075/hcp.44.01luc

Lucy, J., A. (2016). Recent Advances in the Study of Linguistic Relativity in Historical Context: A Critical Assessment. *Language Learning*, 66(3), 487-515.

Lucy, J. A., & Gaskins, S. (2003). Interaction of Language Type and Referent Type in the Development of Nonverbal Classification Preferences. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind* (pp. 465–492). The MIT Press. https://doi.org/10.7551/mitpress/4117.003.0023

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, *108*(2), 566–577. https://doi.org/10.1016/j.cognition.2008.03.009

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review*, *16*(4), 711–718.

https://doi.org/10.3758/PBR.16.4.711

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., Abdel Rahman, R., 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., & 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., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170. https://doi.org/10.1037/a0024904

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

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, *31*(5), 663–672. https://doi.org/10.3758/BF03196106

Malt, B. C., & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition*, *6*(1), 47–67.

https://doi.org/10.1017/S1366728903001020

Mo, L., Xu, G., Kay, P., & Tan, L.-H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences*, *108*(34), 14026–14030. https://doi.org/10.1073/pnas.1111860108

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

Montero-Melis, G., Jaeger, T. F., & Bylund, E. (2016). Thinking Is Modulated by Recent Linguistic Experience: Second Language Priming Affects Perceived Event Similarity. *Language Learning*, *66*(3), 636–665. https://doi.org/10.1111/lang.12172

Moors, P., Hesselmann, G., Wagemans, J., & Van Ee, R. (2017). Continuous Flash Suppression: Stimulus Fractionation rather than Integration. *Trends in Cognitive Sciences*, *21*(10), 719–721. https://doi.org/10.1016/j.tics.2017.06.005

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin, 132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Nedergaard, J. S. K., Wallentin, M., & Lupyan, G. (2023). Verbal interference paradigms: A systematic review investigating the role of language in cognition. *Psychonomic Bulletin & Review, 30*(2), 464-488

Nation, P. (1990). *Teaching and learning vocabulary.* NewYork: Newbury House/Harper Row.

Odlin, T. (2008). Conceptual transfer and meaning extensions. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp.306-340). Routledge.

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 499–508. https://doi.org/10.1037/xhp0000313

Ostarek, M., Joosen, D., Ishag, A., De Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition*, *182*, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017

Paffen, C. L. E., Sahakian, A., Struiksma, M. E., & Van der Stigchel, S. (2021). Unpredictive linguistic verbal cues accelerate congruent visual targets into awareness in a breaking continuous flash suppression paradigm. *Attention, Perception, & Psychophysics*, *83*(5), 2102–2112. https://doi.org/10.3758/s13414-021-02297-y

Panayiotou, A. (2004). Switching Codes, Switching Code: Bilinguals' Emotional Responses in English and Greek. *Journal of Multilingual and Multicultural Development*, *25*(2–3), 124–139. https://doi.org/10.1080/01434630408666525

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

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*(2), 189–219.

https://doi.org/10.1016/S0010-0277(02)00046-X

Papafragou, A., & Selimis, S. (2010). Event categorisation and language: A cross-linguistic study of motion. *Language and Cognitive Processes*, *25*(2), 224–260. https://doi.org/10.1080/01690960903017000

Park, H. I., Jarvis, S., & Kim, J. (2022). Exploring motion event construal: How much attention do speakers of different languages and cultures pay to context? *Lingua*, *265*, 103164. https://doi.org/10.1016/j.lingua.2021.103164

Park, H. I., & Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean–English bilinguals. *Bilingualism: Language and Cognition*, *17*(2), 410–430. https://doi.org/10.1017/S1366728913000400

Pavlenko, A. (2011). *Thinking and speaking in two languages*. Multilingual matters.

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

Phillips, W., & Boroditsky, L. (2003). Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25). https://escholarship.org/uc/item/31t455gf

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Vol. 7529). Penguin UK.

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

Pavlenko, A. (2011). Thinking and Speaking in Two Languages: Overview of the field. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 237-257). Bristol: Multilingual Matters

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

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

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Multilingual Matters.

Sachs, O. S., & Coley, J. D. (2006). 8. Envy and Jealousy in Russian and English: Labeling and Conceptualization of Emotions by Monolinguals and Bilinguals. In A. Pavlenko (Ed.), *Bilingual Minds* (pp. 209–231). Multilingual Matters. https://doi.org/10.21832/9781853598746-010

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. In S. Zufferey & P. Gygax, *The Routledge Handbook of Experimental Linguistics* (1st ed., 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.

Singleton, D., & Leśniewska, J. (2021). The Critical Period Hypothesis for L2 Acquisition: An Unfalsifiable Embarrassment? *Languages*, *6*(3), Article 3. https://doi.org/10.3390/languages6030149

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

Slobin, D. I. (1987). Thinking for speaking. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society*, 13, 435-445.

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

Slobin, D. I. (2000). Verbalised events: A dynamic approach to linguistic relativity and determinism. In S. Niemeier. & R. Dirven. (Eds.), *Evidence for linguistic relativity* (pp. 107-138). Amsterdam, The Netherlands: John Benjamins.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D.Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 157-192). Cambridge, MA: MIT Press.

Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59-81). Amsterdam & Philadelphia, PA:

John Benjamins.

Smith, L. B., & Samuelson, L. (2006). An Attentional Learning Account of the Shape Bias: Reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Developmental Psychology, 42*(6), 1339-1343.

Soroli, E. (2024). How language influences spatial thinking, categorization of motion events, and gaze behavior: A cross-linguistic comparison. *Language and Cognition*, 1–45. https://doi.org/10.1017/langcog.2023.66

Speed, L. J., de Valk, J., Croijmans, I., Huisman, J. L. A., & Majid, A. (2023). Odor-Color Associations Are Not Mediated by Concurrent Verbalization. *Cognitive Science*, *47*(4), e13266. https://doi.org/10.1111/cogs.13266

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *IRAL - International Review of Applied Linguistics in Language Teaching*, *44*(2). https://doi.org/10.1515/IRAL.2006.006

Stein, T., Hebart, M., & Sterzer, P. (2011). Breaking Continuous Flash Suppression: A New Measure of Unconscious Processing during Interocular Suppression? *Frontiers in Human Neuroscience*, *5*. https://www.frontiersin.org/articles/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.

Syndicate, U. C. L. E., & Press, O. U. (2001). *Quick Placement Test*. Oxford University Press.

https://books.google.co.uk/books?id=pj4gnQEACAAJ

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: Cambridge University Press.

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

Thierry, G. (2016). Neurolinguistic Relativity: How Language Flexes Human Perception and Cognition. *Language Learning*, *66*(3), 690–713. https://doi.org/10.1111/lang.12186

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*(11), 4567–4570. https://doi.org/10.1073/pnas.0811155106

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, *63*(1), 64–82. https://doi.org/10.1016/j.jml.2010.02.006

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, 8, 1096–1101.

Vanek, N., & Zhang, H. (2023). On truth and polarity in negation processing: Language-specific effects in non-linguistic contexts. *Frontiers in Psychology*, *14*. https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1244249

Von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012a). How grammaticized concepts shape event conceptualization in language production:

Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, *50*(4). https://doi.org/10.1515/ling-2012-0026

Von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012b). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, *50*(4). https://doi.org/10.1515/ling-2012-0026

Wang, Y., & Li, W. (2021a). 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

Wang, Y., & Li, W. (2021b). Two languages, one mind: The effects of language learning on motion event processing in early Cantonese-English bilinguals. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://escholarship.org/uc/item/6618x8mw

Ward, E. J., & Lupyan, G. (2011). Linguistic penetration of suppressed visual representations. *Journal of Vision*, *11*(11), 322. https://doi.org/10.1167/11.11.322

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

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *WIREs Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104

Wolff, P., & Ventura, T. (2009). When Russians learn English: How the semantics of causation may change. *Bilingualism: Language and Cognition*, *12*(2), 153–176. https://doi.org/10.1017/S1366728909004040

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*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00724

Zlatev, J., & Blomberg, J. (2015). Language may indeed influence thought. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01631

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

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/10.1111/1467-9280.00430