Language control adapts to the immediate but not to the overall language environment during language switching in production and comprehension

Marion Coumel[a], Cong Liu[b], Danijela Trenkic[c,d], Angela de Bruin[a*]

[a]Department of Psychology, University of York, UK

[b]Department of Psychology, Qingdao University, China

[c]Department of Education, University of York, UK

[d] Department of Education, University of Oxford, UK

*Corresponding author:

Angela de Bruin

Department of Psychology

University of York

York YO10 5DD

UK

Phone: +44 (0)1904 322868

Email: angela.debruin@york.ac.uk

Abstract

In dual-language contexts, bilinguals often switch between their languages. How they do this, and how they control their languages during switching, can depend on the nature of the interactional context and the task (comprehension or production). Here, we examined the influence of the immediate and overall language context on language control. First, we examined how language control differs between producing language switches in response to cues, producing switches voluntarily, and comprehending switches. Second, we examined whether language control changes after a change in a bilingual's daily-life overall environment.

To do this, we conducted a longitudinal study with Mandarin-English bilinguals who moved from China (L1-dominant environment) to the UK (bilingual/L2-dominant environment) and with a control group staying in China. Participants completed three tasks twice (seven months apart): cued picture naming (cues indicating language choice), voluntary picture naming (free language choice), and comprehension of spoken words.

Language control differed between the three tasks. Participants showed greater language-switching costs in cued production than during voluntary production and comprehension. Furthermore, only cued production showed that using two languages was more costly than using one (mixing costs). However, we found no evidence that a change in language environment resulted in changes in language control. This suggests a bilingual's language control mechanisms adapt to the immediate context they are communicating in, but are perhaps not shaped as strongly by the overall language environment they live in.

Keywords: Bilingualism, Language Control, Language Production, Language Comprehension, Language Switching, Control Adaptation

1. Introduction

As the world becomes increasingly globalised, more people travel internationally for leisure, work, or studies. For instance, a large number of students move countries (worldwide the estimated number of international students is 6.4 million, Project Atlas, 2022). Many of them live and study in a second language (L2) environment. In many cases, these bilinguals not only use their L2 but also continue to use their first language (L1), often in dual-language contexts in which the two languages can or have to be used interchangeably. Language switching is a frequent phenomenon in such contexts, both in production and comprehension. To fluently switch languages, bilinguals have been argued to use language control to select words in the intended target language, avoid interference from the language not currently in use, and to switch languages when needed or wanted (see Declerck & Philipp, 2015, and Goldrick & Gollan, 2023, for reviews). However, this language control might depend on, and can potentially be shaped by, the immediate context a bilingual is interacting in as well as the more general language environment bilinguals live in. Indeed, bilinguals can differ in how they (have to) switch languages in different contexts. For instance, some contexts require bilinguals to switch languages in response to contextual cues, such as when they have to switch languages in response to a new monolingual conversation partner ("cued switching"). In other circumstances, for instance when talking with a bilingual who speaks the same languages, a bilingual might be able to switch between languages more freely ("voluntary switching").

These interactional contexts can change quickly within a bilingual's day. Within a day, a Mandarin-English bilingual may be able to switch freely with another Mandarin-English bilingual at one moment, while they may afterwards have to switch to English to have a conversation with an English monolingual. However, the general language environment can also change more globally, such as when bilinguals move between countries. For example, a Mandarin-English bilingual might have mostly spent time in non-switch L1 contexts while living in China but is likely to experience more dual-language switching contexts after moving to the UK.

In the current study, we assessed how adaptive a bilingual's language control system is (cf. Green & Abutalebi, 2013) in response to both the immediate context a bilingual is in currently, as well as in response to a larger change in overall language environment. First, we examined whether language control differs between cued production, voluntary production, and comprehension contexts (i.e., role of immediate context). Second, we examined whether changes in the overall daily-life language environment influence language control (i.e., role of overall context). We therefore conducted a study with multiple time points testing bilinguals moving from China (their home environment dominated by their L1-Mandarin) to the UK (a dual-language environment or an environment dominated by L2-English), as well as a control group who stayed in China.

1.1. Language control in different contexts

The Adaptive Control Hypothesis (Green & Abutalebi, 2013) argues that within dual-language contexts, language control during production depends on *how* bilinguals switch languages. It distinguishes between high-control dual-language environments and lower-control dense code-switching contexts in which bilinguals can use their languages more freely. Below, we first review research using cued production tasks to study the first type, followed by voluntary production studies looking at free switching. Finally, language control has also been argued to differ between comprehension and production (e.g., Declerck et al., 2019). However, comparisons so far have focused on cued production versus comprehension, leaving open the question whether any differences are due to production and comprehension differing in general or specifically due to language control being highest during *cued* production. Our study therefore compares cued production, voluntary production, and language comprehension within the same participants. We will use the phrase "immediate context" to refer to the three contexts/tasks studied here.

1.1.1. Cued switching during language production

According to the Adaptive Control Hypothesis, the highest levels of control are required in dual-language contexts where bilinguals need to select their languages and switch between them in response to external cues. In these contexts, bilinguals need to monitor the surroundings to detect cues (e.g., face of an English monolingual), select the language accordingly, monitor conflict and competition between the languages, avoid interference from the language not currently in use (e.g., Mandarin), and adjust language activation or inhibition when switching languages.

Dual-language contexts such as the one described above have been studied through cued production tasks. In these tasks, participants name pictures or digits in the language indicated by a cue (e.g., a face or a country flag). Two key measures of language control can be derived from these tasks (and are also used in voluntary production and comprehension tasks): switching and mixing effects. Starting with switching effects, these are computed by comparing switch trials (current trial requires a different language than the preceding trial) and non-switch trials (same language as preceding trial). Bilinguals often show cued language switching costs: they are slower to name the target on a switch trial than on a non-switch trial (e.g., Costa & Santesteban, 2004; Meuter & Allport, 1999). Furthermore, some studies (e.g., Meuter & Allport, 1999) with unbalanced bilinguals who have a higher L1 than L2 proficiency have reported an asymmetry in switching costs, with larger costs when switching to the L1 than to the L2 (but cf. Gade et al., 2021, suggesting this pattern is not consistently observed across the literature). These asymmetries are often explained through inhibition-related accounts (Green, 1998) arguing that bilinguals inhibit the non-target language, with more inhibition applied over the stronger L1. When switching languages, time is needed to overcome this previously applied inhibition. If the L1 was inhibited more, this can explain larger L1 than L2 switching costs. However, language switching costs can also be explained through increased activation, in particular of the L2. Following this explanation, larger L1 switching costs can result from increased L2 activation persisting into the next trial, making it more time consuming to switch from the L2 to the L1 than vice versa (Philipp et al., 2007).

In addition to the reactive language control associated with the moment of switching, studies have also compared dual-language contexts to single-language contexts where only one language is used. These comparisons show mixing costs, reflecting that bilinguals respond more slowly in dual-language contexts, even when they are not actually switching, compared to single-language contexts (e.g., Christoffels et al., 2007; de Bruin et al., 2018). This is thought to reflect higher attentional and control demands in a cued dual-language environment (cf. Rubin & Meiran, 2005, for a discussion of mixing costs in non-linguistic task switching), which requires cue detection, conflict monitoring, and potentially more proactive language control to balance the activation and inhibition of the L1 and L2 in preparation for dual-language use. These mixing costs too have shown asymmetries, typically with larger L1 than L2 costs (e.g., Christoffels et al., 2007). This suggests bilinguals might also proactively inhibit their L1 and/or increase activation of their L2 in preparation for dual-language use. Where such inhibition or activation is applied too strongly, this can result in overall faster L2 than L1 naming times (referred to as reversed dominance effects, cf. e.g., Goldrick & Gollan, 2023).

1.1.2. Voluntary switching during language production

In contrast, other contexts that also involve the use of multiple languages have been argued to require less language control (Green & Abutalebi, 2013). This might be the case, for instance, when a Mandarin-English bilingual interacts with another Mandarin-English bilingual and can use both languages more freely. Such contexts may allow for opportunistic or cooperative language use, with bilinguals using the words that come to mind fastest regardless of the language they belong to (see e.g., Green & Abutalebi, 2013; Green & Wei, 2014).

To study this type of context, voluntary production tasks ask bilinguals to name each picture in their language of choice. In line with the predictions of the Adaptive Control Hypothesis (Green & Abutalebi, 2013), language production is overall faster in such tasks than in cued production tasks (e.g., de Bruin et al., 2018; Gollan et al., 2014). Furthermore, while cued production tasks consistently show a mixing cost associated with dual-language use, voluntarily using two languages is often not associated with a mixing cost and in some cases even shows a mixing benefit (e.g., de Bruin et al., 2018; de Bruin & Xu, 2023; Gollan & Ferreira, 2009; Grunden et al., 2020). That is, participants can sometimes name words faster in voluntary dual-language contexts than when they have to use *one language* only in single language blocks. This mixing benefit likely reflects benefits associated with opportunistic language use, for instance with bilinguals naming less accessible L2 items in their L1 (see e.g., Gollan & Ferreira, 2009) and more generally with bilinguals using the word that comes to mind fastest, regardless of the language (de Bruin et al., 2018). In addition, the mixing benefit is associated to a lower need for proactive language control when both languages can be used freely. Overall, these findings align with the Adaptive Control Hypothesis (Green & Abutalebi, 2013) in suggesting that voluntary switching production contexts require less language control than cued switching contexts and, in some bilinguals, even less control than in single-language contexts.

Most studies do find significant switching costs in voluntary production tasks (e.g., de Bruin et al., 2018, 2020; Gollan & Ferreira, 2009; Gross & Kaushanskaya, 2015; Jevtović et al., 2020). Some studies find voluntary costs to be comparable to cued costs (e.g., de Bruin et al., 2018) while others find voluntary costs to be smaller than cued costs (e.g., de Bruin & Xu, 2023; Gollan et al., 2014; Jevtović et al., 2020). The presence of voluntary switching costs suggests that, even in this context, some influence from the language preceding a switch trial persists, increases language competition, and potentially the need for language control when switching languages. Such switching costs are also observed in experimental designs looking at switching in context (as opposed to production of individual words, de Bruin & Shiron, 2024) as well as corpus data (Fricke et al., 2016). Indeed, language switching is likely not completely opportunistic or purely driven by lexical access, even in voluntary switching environments. Rather, other factors might also play a role, including the influence of language preferences (e.g., de Bruin & Martin, 2022) and cues present in the environment (e.g., Vaughn-Evans, 2023). In line with this, switching costs are typically only absent when language competition is greatly reduced, either through specific language-choice instructions (Kleinman & Gollan, 2016) or by using concepts that are rare or absent in one of the languages (e.g., Zhu et al., 2022; but cf. Blanco-Elorrieta & Pylkkänen, 2017).

1.1.3. Processing of language switches during comprehension

While the Adaptive Control Hypothesis (Green & Abutalebi, 2013) focuses on production, bilinguals also are exposed to language switches made by other speakers. Comprehension, depending on the language environment, might trigger less parallel co-activation and language competition than language production (see e.g., de Bruin & Xu, 2023; Coumel et al., 2024; Declerck et al., 2019). Indeed, during word comprehension, the orthographic or phonological information activates the target word, with potentially less competition from the other language than in production, where a concept can activate multiple L1 and L2 words (e.g., Declerck, Koch, et al., 2015; Green, 1998). Reduced parallel language co-activation during comprehension could in turn reduce language interference and the need for language control.

Although switching costs are observed in comprehension too (e.g., Olson et al., 2017; Litcofsky & Van Hell, 2017), they seem smaller and appear less consistently than across the production literature (see Gade et al., 2021, for a production meta-analysis). The exact switching-cost patterns also vary across the comprehension literature. Switching costs sometimes arise in one language only (e.g., Bultena et al., 2015; Gullifer & Titone, 2019; Jackson et al., 2004; Mosca & de Bot, 2017; Olson, 2017) or they are only observed for some bilinguals, for example only for bilinguals who do not switch frequently themselves (e.g., Gosselin & Sabourin, 2021; Valdés Kroff et al., 2018; see also Kaan et al., 2020). Recently, Declerck et al. (2019) reported no switching costs in French-English and French-Spanish bilinguals across a range of comprehension tasks (including e.g., parity, magnitude, and animacy judgment tasks), whereas they did find significant production switching costs. While most comprehension studies have used visual stimuli (cf. Olson, 2017; Van Hell, 2023), some studies have also examined comprehension of auditory (spoken) words. Switching costs should be expected in the auditory domain too (e.g., BIA+, Dijkstra & Van Heuven, 2002) but, similar to visual tasks, results are mixed. For instance, while Olson (2017) showed larger L1 than L2 costs, Liao and Chan (2016) showed the opposite pattern, and Declerck et al. (2019) showed no switching costs in a task including spoken and visual stimuli. However, research on switch processing of spoken words is scarce, despite bilinguals being most likely to encounter language switches in spoken form while interacting with other bilinguals (Van Hell, 2023).

Few studies have measured mixing effects in comprehension tasks, and they also report mixed results. Grainger and Beauvillain (1987) found significant mixing costs in a generalized lexical decision task, whereas Declerck et al. (2019) found mixing costs only in one of three experiments. Finally, Jylkkä et al. (2017) observed a mixing benefit for the L2 and no significant mixing effects for the L1. They interpreted the former result as reflecting opportunistic planning (Green & Abutalebi, 2013). They argue that in dual-language contexts, when doing a semantic categorisation task as often used in comprehension studies, bilinguals can take a faster lexical route from the L2 to the L1, and then to the target concept (cf. the Revised Hierarchical Modal, Kroll & Stewart, 1994). Such faster route through the L1 might be less or not possible during single-language contexts if the L1 is inhibited (Jylkkä et al et al., 2017). However, these studies all used written words, leaving open the question whether mixing costs are observed during spoken word comprehension. Studies using spoken words or sentences have typically compared dual-language contexts as a whole to single-language blocks. Gavino and Goldrick (2024) showed better transcription performance when listening to single-language sentences compared to sentences with code switches (cf. García et al., 2018, for similar findings). This suggests increased processing costs and potentially language control in dual-language blocks during spoken sentence comprehension, although the comparison with dual-language blocks as a whole (rather than comparing single-language with non-switch trials) makes it difficult to know whether differences with the single-language blocks are driven by reactive control at the moment of switching or more proactive control mechanisms.

Taken together, it is not clear how bilinguals apply language control in comprehension tasks. Studies comparing production and comprehension have furthermore shown mixed results, with some arguing similar control mechanisms are applied (e.g., Gambi & Hartsuiker, 2016; Peeters et al., 2014) while others emphasise differences (e.g., Ahn et al., 2020; Blanco-Elorrieta & Pylkkännen, 2016; Li et al., 2024; Macizo et al., 2012; Mosca & de Bot, 2017; Reynolds et al., 2016). Where production and comprehension are compared, studies often use tasks asking participants to alternate between comprehension and production (e.g., Li & Gollan, 2020, cf. H. Liu et al., 2021 for a comparison between cued and voluntary production), addressing a related but different question. A recent study by Li et al. (2024) directly compared reading in silence to reading aloud and also suggested different control mechanisms are involved when production (reading aloud) is added. However, in that study, different measures (ERPs versus reading times) were used to assess comprehension and production. Furthermore, the focus on reading adds an influence of orthographic processing that is not necessarily present (to the same extent) during typical language production in the absence of written text.

When comprehension and production switching costs are compared without asking participants to constantly alternate between the two or to do a reading task, comprehension is typically compared to *cued* production (e.g., Mosca & de Bot, 2017). These comparisons have often revealed differences in control processes. However, such differences could be due to comprehension-production differences in general or could be specific to the higher levels of language control used during *cued* production. Production and comprehension vary in multiple ways. Language-control differences could arise regardless of the type of production because production requires speakers to select a lexical word form out of a large range of options, to select the corresponding phonological form, and to articulate it. In contrast, during comprehension, the word input itself already provides strong information about the language and word form. Thus, control differences might arise between production and comprehension in general due to these processing differences. However, production-comprehension differences might also be inflated or driven by the typical comparison with the most demanding type of production: cued production. Previous research comparing both voluntary and cued production to comprehension (by manipulating the duration of response-stimulus intervals, de Bruin & Xu, 2023) suggested comprehension differs from cued production but not necessarily from voluntary production. However, comprehension and production tasks were completed by different bilingual groups, hindering a direct comparison.

Thus, theoretically, language control is argued to be lower during voluntary than cued production (e.g., Green & Abutalebi, 2013) and during comprehension than production (e.g., Declerck et al., 2019). While current empirical evidence supports these differences, it currently is not clear whether production-comprehension differences are general (in which case both cued and voluntary production should differ from comprehension) or specific to high-control types of production (in which case differences between production and comprehension should mostly or only emerge in comparison with the cued production task). No studies to date have compared comprehension to the two production types (cued and voluntary) within participants. As a first research question (RQ1), we therefore examined how the immediate language context and task (comprehension or production) influence language control. Furthermore, the literature so far has focused on written word processing during comprehension, despite bilinguals more frequently encountering spoken than written switches. When comparing comprehension and production (e.g., Mosca & de Bot, 2017; Peeters et al., 2014), this also introduces orthographic influences in the comprehension task that are not present (to the same extent) during picture naming. The current study therefore compared word production and comprehension of *spoken* words.

1.2. The impact of overall language environment on language control

In addition to language control depending on the immediate context, the overall environment bilinguals live in can also play a role. Starting with word retrieval more broadly, bilinguals' L1 and L2 use and exposure can vary, which can influence proficiency (i.e., the actual knowledge, including vocabulary) and ease of lexical access, in line with the weaker links hypothesis (Michael & Gollan, 2005). Indeed, several studies (e.g., Beatty-Martinez et al., 2020) have shown that lexical access (speed of naming) in picture-naming tasks differs when comparing bilingual groups living in different language environments. Longitudinal studies examining L1 and L2 production in *single-language contexts* also suggest language learners' L1 and L2 access can change after spending time in an L2 environment (e.g., Baus et al., 2013; Botezatu et al., 2022; Linck et al., 2009). Even short re-immersion in an L1 environment can impact lexical access (Casado et al., 2023). However, some of these studies have also suggested L2 immersion effects might only apply to certain words (e.g., low-frequency words in Baus et al., 2013, and high-frequency words in Casado et al., 2023) or might not always be observed (e.g., Casado et al., 2023, did not observe group differences between bilinguals in an L1 or L2 environment).

Beyond general L1 and L2 access, a key open question remains whether changes in overall language environments can also influence *language control.* Daily-life language environments vary in the amount of time a bilingual spends in single-language or in dual-language contexts, as well as in the frequency and type of language switching. Bilinguals moving from an L1-dominant environment to an L2-dominant/dual-language environment are likely to increase not just their L2 use, but also time spent in dual-language contexts and switching languages. On the one hand, increased daily-life switching could operate as a form of switching practice, potentially optimising the language-control system, and resulting in a decrease in switching costs. On the other hand, an environment in which a bilingual frequently has to switch to the less proficient L2 might require more L1 inhibition than an environment in which the bilingual predominantly uses their L1. In this case, increased L1 inhibition might result in an increase of L1 switching costs.

Most research examining the influence of linguistic environments on bilinguals' control systems has focused on changes in *cognitive* control rather than in *language* control (see e.g., Hartanto & Yang, 2016; Kałamała et al., 2020; Lehtonen et al., 2023; Zhang et al., 2015). Given our focus on language control, we will not review the cognitive control literature here. Rather, we will discuss three types of studies that have examined a relationship between language environment and language control: cross-sectional studies examining different types of bilinguals and/or individual differences, studies manipulating the language environment within the study's language-control task, and short-term training studies. While these are all relevant for our study, their focus on cross-sectional comparisons, immediate contexts, or short-term effects differs substantially from our study's aim to examine potential long-term changes within bilinguals after a change in real-life language environments.

Starting with comparisons between bilinguals with different language backgrounds, several studies have shown that a bilingual's language background can relate to their language control. For example, during cued production, Italian-English and Italian-Sardinian bilinguals with a higher L2 exposure have shown (marginally) smaller L1 switching and mixing costs than groups with lower L2 exposure (Bonfieni et al., 2019). Comparisons between Chinese-English bilinguals immersed in an L2 versus L1 environment have suggested those in an L2 environment apply more L1 inhibition than those in an L1 environment (Zhang et al., 2021). Moreover, bilinguals who code-switch more frequently in their everyday life have shown smaller switching costs in production tasks (e.g., Han et al., 2022; Prior & Gollan, 2011; Yim & Bialystok, 2012). During comprehension, bilinguals who switch frequently also process switches differently than bilinguals who switch less frequently (e.g., Beatty-Martinez & Dussias, 2017; Gosselin & Sabourin, 2021; Hui et al., 2022; Kaan et al., 2020; Litcofsky & Van Hell, 2017; Valdés Kroff et al. 2018; see also Adamou & Shen, 2019). These studies thus generally suggest that bilinguals who switch more often in their daily lives show smaller switching costs. However, in these between-group comparisons, groups might differ in other aspects as well (e.g., proficiency) that can make it more difficult to examine whether any differences are truly due to daily-life language environments.

These potential differences between groups can be avoided by manipulating the language environment itself. Some studies have examined the immediate impact of changing the language context in which a switching task is completed. For instance, Timmer et al (2019) showed that language control can differ depending on whether Dutch-English participants complete a switching task in an L1- or L2-dominant experimental context. Similarly, Olson (2016) showed that the pattern of switching costs varied in Spanish-English bilinguals depending on the number of L1 and L2 trials in the task. These studies highlight the flexibility of language control in response to the immediate context but do not directly speak to potentially changes in response to the overall daily-life environment.

Finally, other studies have provided participants with (short-term) training, suggesting that causal changes in language control are possible beyond the immediate language context. Wu et al. (2018) asked unbalanced Chinese-English bilinguals to complete a two-day cued production training. L1 switching costs decreased with training (Experiment 1), suggesting reactive language control can be optimised by language switching practice. Similarly, Kang et al. (2017, 2018) used an eight-day training paradigm with comprehension switching tasks in Chinese-English bilinguals and observed changes in ERP and fMRI data, which were interpreted as more efficient language competition resolution. Behavioural data showed a switching-cost reduction in terms of RTs in the 2017 study and in terms of accuracy (but not RTs) in the 2018 study. However, Zhang et al. (2021) did not show changes in switching costs after a three-day language switching training in L2-immersed bilinguals, even though some changes in overall L1 and L2 naming times were observed. Finally, examining changes after a year of classroom L2 learning, C. Liu et al. (2021) showed a reduction in language switching costs. Thus, changes in switching and language-control mechanisms appear to occur beyond the influence of the immediate language context, although in the studies conducted so far, such changes might be closely linked to the L2 or switching training provided to participants.

1.3. Research questions and present study

In summary, so far, studies have focused on comparing different groups or have trained participants on the control component of interest (language switching, with the exception of C. Liu et al., 2021). Furthermore, many of these studies focused on language comprehension (e.g., Gosselin & Sabourin, 2021; Kaan et al., 2020; Kang et al., 2017, Litcofsky & Van Hell, 2017) or on single-language contexts (e.g., Baus et al., 2013). To examine the actual adaptability of language control in response to immediate and overall language-environment changes, a within-participant study examining changes within bilinguals across a longer period of time is necessary.

The goal of the current study was therefore to examine both the influence of the immediate switching context (voluntary production, cued production, comprehension) on language control (RQ1) and potential changes in language control in adaptation to changes in the overall daily-life language environment (RQ2). We tested Mandarin-English bilinguals who had recently moved from China (an L1-dominant environment) to the UK (an L2 environment including L2-dominant and dual-language contexts; “Move group”) upon arrival to the UK (Session A) and about seven months later (Session B). As a control group, we also tested Mandarin-English bilinguals who stayed in China during the whole study (“Control group”).

RQ1 firstly examined potential effects of the immediate context on language-control. In line with the Adaptive Control Hypothesis (Green & Abutalebi, 2013) and previous research (e.g., de Bruin & Xu, 2023), we expected cued language production to pose the highest demands in terms of reactive control (applied at the moment of switching), proactive control (overall managing of language activation and inhibition), as well as overall goal monitoring and cue detection (Green & Abutalebi, 2013). Such demands are lower during voluntary switching (e.g., de Bruin et al., 2018; Green & Abutalebi, 2013). Therefore, and in line with the literature comparing cued and voluntary production (e.g., de Bruin & Xu, 2023; Gollan et al., 2014; Jevtović et al., 2020), we expected larger switching costs during cued than voluntary switching as a reflection of reactive language-control differences. Furthermore, we expected a cued mixing cost but no cost or even in a benefit in the voluntary task, reflecting lower proactive control during voluntary language mixing (e.g., de Bruin et al., 2018; de Bruin & Xu, 2023; Gollan & Ferreira, 2009; Grunden et al., 2020). In terms of production versus comprehension, if comprehension elicits less language competition and therefore requires less control (cf. Declerck et al., 2019), we expected smaller switching and mixing costs during comprehension than production. If such difference is driven by language control, we expect comprehension effects to differ from *cued* production in particular. If switching and mixing effects differ between comprehension and both production tasks, this would suggest reactive and proactive control differences are related to broader comprehension-production differences.

Our spoken word comprehension task allowed for a more direct comparison between spoken production and comprehension, removing any additional effects of orthographic processing during comprehension. We used an animacy judgement task to elicit RTs. A range of tasks has been used across the literature, but there is no consistent pattern with respect to some tasks showing (larger) switching costs (e.g., Declerck et al., 2019). An animacy judgement task was preferred over, for example, a lexical decision task to ensure some level of semantic processing was needed beyond word-form recognition. Animacy judgement tasks have also been used regularly to study comprehension switching, and have previously shown switching effects (Jylkkä et al., 2018, but cf. Declerck et al., 2019, Experiment 7), therefore also allowing us to compare our study to previous relevant literature. To avoid introducing production-related control components, we asked participants to respond through button presses rather than, for example, using a verbal reading out loud task.

Our RQ2 examined whether moving to a new language environment impacted language control in the Move group. We formulated the following sets of hypotheses in relation to this longitudinal question:

We expected the Move group to show changes in their language production and comprehension after having spent seven months in the UK (Session B), compared to at the start of their time in the new language environment (Session A). Such changes after moving language environment could affect lexical access more generally and/or language control. Previous research focusing on single-language production showed that changes in lexical access can indeed occur longitudinally (e.g., Baus et al., 2013). Short-term training studies furthermore suggest exposure to language switches can influence language control (e.g., Kang et al., 2017, but cf. Zhang et al., 2021). We expected these changes to only occur in the Move group, because the Control group stayed in the same language environment throughout the study (i.e., interaction between Group (Move/Control) and Session (A/B)).

If reduced L1 use and increased L2 use after the move influence lexical access, we expected reduced overall L1 lexical access (slower naming) and facilitated L2 lexical access (faster naming) in the Move group in Session B relative to Session A (interaction between Session and Language within the Move group).

If language control adapts, we expected switching and/or mixing effect changes in the Move group (interaction between Session and Switching/Mixing, and potentially a three-way interaction with Language). As the cued production task was expected to require the highest level of language control among the three tasks (cf. Green & Abutalebi, 2013), we expected larger language-control changes in this task than in the voluntary production task and in the comprehension task. Two directions could emerge:

a) Living in an L2/dual-language environment results in bilinguals more frequently using and switching to the L2, compared to the previous L1-dominant language environment. This might require bilinguals to apply more L1 inhibition after the move. With increased L1 inhibition linked to larger L1 switching and/or mixing effects (cf. Green, 1998), (L1) switching and/or mixing effects might be larger after 7 months in the new environment (Session B compared to A). Such hypothesis is supported by Linck et al. (2009)'s argument that their participants' decrease in L1 performance (in a single-language task) after moving to an L2 environment was due to increased L1 inhibition. It is also supported by Zhang et al's (2021) conclusion that bilinguals in an L2 environment apply more L1 inhibition than those in an L1 environment
b) In contrast, living in an L2/dual-language environment can result in more exposure to (or "practice" with) language switching, which could facilitate the development of more efficient language control and optimise language switching. In this case, we expected a reduction in switching and/or mixing effects between sessions. This switching-cost reduction would be in line with previous short-term training studies (e.g., Kang et al., 2017). This could influence L1 switching costs most strongly, if L1 inhibition is optimised the most.

Finally, previous studies comparing participants from different language backgrounds in cross-sectional designs (e.g., Bonfieni et al., 2019; Yim & Bialystok, 2012) have shown that switching and mixing costs can differ between bilinguals. Participants within our Move group were likely to differ from each other in terms of how they used their languages in the UK and how much their language environment actually changed compared to before their move. In addition to comparing changes at the group level, within the Move group we therefore also examined relationships between (potential changes in) language control and individual differences in language profiles, including age of L2 acquisition, proficiency, language use and switching habits (RQ3). This question, including the hypotheses and analyses, is fully described in the Supplementary Materials presented on the OSF page (https://osf.io/a24xv/), and not covered further within the manuscript.

2. Methodology

2.1. Transparency and openness

The preregistration for the analyses of RQ1-RQ2 (and the preregistration for the RQ3 analyses in the Supplementary Materials), the stimuli, datasets, and the analyses scripts are available on the Open Science Framework: https://osf.io/a24xv/ (Coumel et al., 2025). The study also included separate single-language tasks that addressed different research questions. These single-language production tasks were completely separate from the dual-language tasks described here. The dual-language tasks included here did also include their own single-language blocks to compute mixing effects. Throughout the manuscript, we will refer to these single-language blocks within the dual-language tasks as "single-language blocks". The separate single-language tasks are described in a separate manuscript (de Bruin et al., 2025). Below, we report how we determined our sample size, describe all data exclusions, manipulations, and measures, and provide details about the analyses.

2.2. Participants

We recruited two groups of L1-Mandarin students: one group was about to move to the UK at the beginning of the study (Move group), while the other group was staying in China (Control group). To facilitate recruitment, the participants were tested in two waves during the academic years 2021-2022 and 2022-2023. All participants were studying a degree related to linguistics and/or education, with many studying a degree related to English. We chose students with these academic backgrounds based on the assumption that they would have a higher English proficiency level than other students, allowing us to focus on potential changes in language control rather than fundamental proficiency changes. Based on a power analysis (using *simr)* using mixed effect models and 40 trials per condition, our recruitment aim was to start with 100 participants per group to obtain 95% power to detect medium-sized effects (including after 10% attrition). In the end, 158 participants completed the first online questionnaire and 135 completed Session A (15% attrition, see Table 1 for an overview of the different sections of the study). The full study (Sessions A and B) was completed by 114 Mandarin-English bilinguals (53 Move participants and 61 Control participants; 16% attrition between Session A and B). Attrition rates were relatively high (total attrition rate of 28% between the first and last study part). This was closely related to the study being completed during the COVID-19 pandemic, which resulted in a lower number of participants moving from China to the UK to complete a degree, and therefore also a smaller sample size than planned. Although our sample size was lower than planned, power calculations using *simR* showed we still had over 80% power to detect small to medium-sized changes in switching and mixing effects (i.e., a significant interaction between Trial Type and Session) within the Move group (as well as a three-way interaction with Group). Furthermore, considering the important role of the number of trials per condition (cf. Brysbaert & Stevens, 2018), we also included at least 40 trials per condition, per task (see “Procedure” for details).

On average, the Move group completed Session B 6.78 months after Session A (*SD* = 7.82), and the Control group 6.27 months after Session A (*SD* = 15.28). The time between sessions was significantly shorter in the Control than in the Move group (p <.001) but between six and seven months for both groups. All participants had normal or corrected-to-normal vision and no known neurological, reading, or hearing impairments. The study was approved by the Ethics Committee of the Department of Psychology at the University of York, with additional approval for testing of the control group provided by the Department of Psychology at Qingdao University. Participants provided written informed consent at the start of the study and for each individual session.

Table 1. Overview of the relevant dual-language tasks and questionnaires completed in each session. Descriptions of and full references for each task/questionnaire are given in the text of the Methods section and/or Supplementary Materials. LexTALE stands for Lexical Test for Advanced Learners of English (Lemhöfer & Broersma, 2012); BEST picture naming test is based on the Basque-English-Spanish Test (de Bruin et al., 2017); LSBQ refers to the Language and Social Background Questionnaire (Anderson, et al., 2018); the BSWQ to the Bilingual Switching Questionnaire (Rodriguez-Fornells et al., 2012); and the BICQ to the Bilingual Interactional Context Questionnaire (Hartanto & Yang, 2020).

About 2 weeks before Session A (i.e., before move to the UK)	**Session A**	**Session B**
<u>**Online questionnaire:**</u> • **Age of Acquisition (AoA)** • **Language proficiency**: • *Self-rated* • *Objective*: o LexTALE o BEST picture naming test • **Language Use:** • *Frequency*: o LSBQ (also used to compute entropy) o General language use and exposure	**Experimental dual-language tasks**: o Cued production task o Voluntary production task o Comprehension task • **Language proficiency**: • *Self-rated* • *Objective*: o LexTALE o BEST picture naming test o Interview o Vocabulary size • **Language Use:**	**Experimental dual-language tasks**: o Cued production task o Voluntary production task o Comprehension task • **Language proficiency**: • *Self-rated* • *Objective*: o LexTALE o BEST picture naming test o Interview o Vocabulary size • **Language Use:**

• **Language switching**: • *Frequency*: o BSWQ o Switching questions • *Type (Language environment):* BICQ	• *Frequency* o LSBQ o General language use and exposure • **Language switching**: • *Frequency:* o BSWQ o Code-switching frequency judgment task o Switching questions o Story telling task (second wave only) • *Type (Language environment):* BICQ	• *Frequency* o LSBQ o General language use and exposure • **Language switching**: • *Frequency:* o BSWQ o Code-switching frequency judgment task o Switching questions o Story telling task (second wave only) • *Type (Language environment):* BICQ

Note: During Sessions A and B, the experimental tasks, the picture naming tests, and the interviews were completed in the lab. All other tasks and questionnaires were completed online using Gorilla (Anwyl-Irvine et al., 2020). The online part of Session A was completed on average 8.96 days after the in-person part. The online part of Session B was completed on average 14.44 days after the in-person part. The study included additional tasks (e.g., completely separate single-language production tasks, a grammaticality judgement task) but here we only include the tasks relevant for this manuscript. Finally, Move-group participants also completed the language-experience questionnaires at an additional time point approximately three months after their arrival in the UK. This is not included here as it was only included to better understand changes in daily-life language experiences throughout the first months.

At the beginning of the study, before any participants moved to the UK, an online questionnaire asked the participants about their (language) background (see Supplementary Materials on the OSF page for details). Participants' general background was comparable between the groups (all *ps* >.2), in terms of age (Move group: *Mage* = 24, *SD* = 3; Control group: *M* = 23, *SD* = 2), gender (with 43 female participants in the Move group and 53 in the Control group), completed years of education (Move group: *M* = 16, *SD* = 1; Control group: *M* = 17, *SD* = 1), and non-verbal reasoning (Move group: *M* = 69%, *SD* = 20; Control group: *M* = 65%, *SD* = 26). In the latter task participants performed a progressive matrices test, in which they were presented with an array of geometrical shapes and asked to select from a set of options the missing shape that completed the figure (ICAR, 2014, Condon & Revelle, 2014). In the Move group (for those who answered the question), parental education was higher (63% reported one or both parents having completed higher education) than in the Control group (24% of parents completed higher education).

In Table 2, we report data regarding participants' background at baseline in terms of L2-English age of acquisition and L1-Mandarin and L2-English proficiency levels. These did not differ between the Move and Control groups. Participants acquired L1-Mandarin from birth and typically started acquiring L2-English during childhood. They were more proficient in their L1 than in their L2 (all tests showed a significant language difference) but had typically reached at least an intermediate L2-proficiency level. Proficiency was measured at baseline through self-ratings as well as several objective measurements (further described in the Supplementary Materials). This included the LexTALE (Lexical Test for Advanced Learners of English) in English (Lemhöfer & Broersma, 2012) and in Mandarin (Wen et al., 2023, only for participants in the second wave), two short lexical decision tasks in which participants were asked to indicate whether presented letter strings corresponded to real words in the tested language or not. Productive vocabulary was examined through the BEST (based on the Basque-English-Spanish test) picture naming tests in L1-Mandarin and L2-English, in which participants had to name pictures corresponding to non-cognate words in English and Mandarin (de Bruin et al., 2017).

Table 2. Summary of objective and subjective measurements of Mandarin and English proficiency in the baseline (pre-move) assessment for the Move group (left) and the Control group (right)[1]. Groups did not significantly differ on any of the measurements reported in this table (see "Comparison" for *t* and *p* values). Participants performed better in L1-Mandarin than in L2-English in all measures.

	Move group (N = 53)			Control group (N = 61)			Comparison
	Mean	**SD**	**Min-Max**	**Mean**	**SD**	**Min-Max**	
L2 Age of Acquisition							
AoA: onset	8.2	2.4	3-14	8.4	2.0	3-13	t = 0.569, p = .570
First conversation	10.6	3.5	5-20	11.2	3.0	5-20	t = 0.899, p = .371
Outside classroom use	15.9	3.9	6-25	14.9	5.0	5-26	t = -0.925, p = .358
LexTALE (0-100%)[2]							
L1-Mandarin	94.1	9.5	57.5-100	96.8	4.4	85-100	t = 1.388, p = .171
L2-English	67.3	12.5	47.5-95	64.6	11.5	41.3-100	t = -1.171, p = .244
Picture Naming Test (0-100%)							

L1-Mandarin	96.8	3.6	83.1-100	94.6	9.7	50.8-100	*t* = -1.442, *p* = .152
L2-English	75.7	12.2	47.7-96.9	76.7	20.3	35.4-98.5	*t* = 0.288, *p* = .774
Mandarin Self-Rated Proficiency (0-10)							
Speaking	9.4	0.9	7-10	9.2	1.1	6-10	*t* = -1.043, *p* = .299
Understanding	9.5	1.0	5-10	9.2	0.9	7-10	*t* = -1.360, *p* = .177
Writing	8.5	1.5	4-10	8.0	1.6	4-10	*t* = -1.613, *p* = .110
Reading	9.2	1.0	6-10	8.7	1.2	6-10	*t* = -1.978, *p* = .050
English Self-Rated Proficiency (0-10)							
Speaking	5.7	1.2	3-8	5.6	1.3	3-9	*t* = -0.170, *p* = .865
Understanding	6.5	1.4	2-9	6.5	1.3	3-9	*t* = -0.081, *p* = .935
Writing	5.6	1.1	3-8	5.8	1.4	3-9	*t* = 0.874, *p* = .384
Reading	6.8	1.2	4-9	6.7	1.4	3-9	*t* = -0.403, *p* = .687

[1]Not all participants actually used English outside the classroom and therefore 14 Move and 26 Control participants left the third AoA question unanswered. AoA onset data are missing for 2 participants and AoA first conversation data are missing for 4 participants. Only the participants of the second wave of testing completed the Mandarin LexTALE test (N = 57), as this was published after we started data collection. Moreover, due to technical issues or because participants did not fill in the entire questionnaire, some data are missing for some tasks (L2-English LexTALE N=3; L1-Mandarin picture naming N = 14; L2-English picture naming N = 16; proficiency self-reports N = 3).

[2]These mean L2-Lextale scores correspond to a B2 level in the Common European Framework of Reference (Council of Europe, 2001) at baseline, with some participants at B1 level and some participants at C1 level.

In Table 3, we report data regarding participants' L1-Mandarin and L2-English use and their language switching behaviour before the move. Language use was assessed through the Language and Social Background Questionnaire (LSBQ; Anderson, et al., 2018; with some small modifications based on Mann & de Bruin, 2022) as well as a general language use/exposure score (see Supplementary Materials). Participants, across groups, were exposed to, used, and had been using L1-Mandarin more than L2-English since their childhood. Although this was measured before their move to the UK, participants of the Move group reported that they were currently using L2-English more than participants of the Control group. This may be because Move group participants increased their use of English in their everyday life in preparation for their move to the UK, for example to deal with administrative procedures. The two groups did not significantly differ in the other language use measures.

Regarding language switching behaviours, across groups, participants reported they sometimes switched languages and spent some time in dual-language environments before the move. Switching frequency (see Supplementary Materials for details) was assessed through the Bilingual Switching Questionnaire, BSWQ (Rodriguez-Fornells et al., 2012) and three additional switching questions asking about language switching on a daily basis, within a conversation, and in sentences (de Bruin et al., 2018). Type of switching was further assessed through the Bilingual Interactional Context Questionnaire, BICQ (Hartanto & Yang, 2020), which asked bilinguals about time spent in dual-language environments switching languages between different people in a context (but not mixing languages within a sentence) and time spent mixing languages freely within sentences (dense code switching). There were no significant between-group differences on these measures, with the exception of the Move group reporting slightly higher within-sentence switching on one of the questions (see Table 3). Further switching measures were included in Sessions A and B but not in the baseline measures reported in Table 3.

Table 3. Summary of Mandarin and English language use in the baseline (pre-move) assessment for the Move group (left) and the Control group (right)[1]. The two groups showed small significant differences in terms of their LSBQ current language use score and the within-sentence switching question but did not differ on the other measures at baseline (see "Comparison" for *t* and *p* values).

	Move group (N = 53)			**Control group (N = 61)**			**Comparison**
	Mean	**SD**	**Min-Max**	**Mean**	**SD**	**Min-Max**	
LSBQ Language Use (1 = all English; 5 = all Mandarin)							
Childhood use at home and school	4.4	0.3	3.7-5	4.3	0.3	3.4-5	*t* = -0.361, *p* = .718
*Current use across people and contexts**	3.7	0.7	1.8-4.8	4.1	0.5	2.7-4.8	*t* = 3.306, *p* = .001
Language Use (0-100% of the time)							
L1-Mandarin	81.5	15.3	40-100	84.3	10.8	50-99	*t* = 1.141, *p* = .256

L2-English	18.5	15.3	0-60	15.7	10.8	1-50	*t* = -1.141, *p* = .256
Language Exposure (0-100% of the time)							
L1-Mandarin	82.3	17.3	30-100	81.6	15.6	50-100	*t* = -0.248, *p* = .805
L2-English	17.7	17.3	0-70	18.4	15.6	0-88	*t* = 0.248, *p* = .805
BSWQ Score (1 = never – 5 = always)	2.8	0.4	2-3.7	2.8	0.4	1.9-3.7	*t* = -0.100, *p* = .921
Switching Frequency (1 = never, 5 = always)							
Daily switching	2.9	1.0	1-5	2.6	0.8	1-4	*t* = -1.365, *p* = .175
In a conversation	2.6	0.8	1-5	2.4	0.8	1-4	*t* = -1.339, *p* = .183
*Within sentences**	2.3	0.8	1-5	2.0	0.7	1-4	*t* = -2.016, *p* = .046
Time Spent in Each Environment (0-100% of the time, BICQ)							
Dual-language environment	20.1	20.3	0-74.5	21.4	19.6	0-90	*t* = 0.351, *p* = .727
Dense code-switching environment	12.1	17.4	0-96.6	15.1	17.5	0-86.4	*t* = 0.904, *p* = .368

[1] *Because some participants did not fill in the entire questionnaire or due to uninterpretable answers on the BICQ questionnaire, 3 participants had missing data for all questionnaires and an additional 2 had missing data for the BICQ. Data for Mandarin/English exposure were removed for one participant who reported 0 for both languages by mistake.*

We also assessed the participants' language proficiency, daily-life use, and switching throughout the study. Changes in language proficiency and daily-life use and switching are reported in Table 4 (see Supplementary Materials for how these composite scores were created for each language-profile component). Participants in the Move group significantly increased in terms of their self-rated and objective L2 proficiency levels. Their self-rated L1 proficiency scores were lower in the final session than at baseline. The proficiency changes, however, were small, as intended through our recruitment of participants who already had an intermediate L2 proficiency level. Larger changes were observed in terms of use of the L1 and L2, with an increase in L2 use and entropy scores showing more time spent in dual- than single-language contexts. This was also captured in the language switching scores, which showed a small increase in switching frequency, as well as participants spending more time in dual-language contexts in which they would switch languages within a given environment. It is this type of dual-language switching (e.g., switching languages between different conversation partners within the same language environment) that is argued to rely most strongly on language control (Adaptive Control Hypothesis, 2013). In contrast, dense code-switching (mixing languages within an utterance) is argued to rely more on opportunistic use (using the words that come to mind fastest) and is furthermore only possible when interacting with other Mandarin-English bilinguals. Indeed, these dense code switching experiences did not change after participants' move at the group level, while (crucially) their time spent switching in dual-language contexts did. The changes observed in the Move group were (as expected) not present in the Control group on any of the measures reported in Table 4 for the Move group. The absence of such changes in the Control group suggests the Move group indeed experienced these changes in their daily-life language experiences as a consequence of their move.

Table 4. Summary of the Move group participants' language profile in Session B and the difference relative to baseline assessment. * indicates significant changes in language profile (p < .05). This table only includes the 43 participants in the Move group who completed the language profile questionnaires needed to compute all scores reported below at baseline and in Session B.

Language profile	**Baseline**		**Session B**		**Difference score to baseline**	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Self-Rated Proficiency						
*L2-English**	62 (9)	40-85	66 (10)	40-85	4 (8)	-10-+25
*L1-Mandarin**	92 (9)	75-100	86 (12)	63-100	-6 (13)	-38-+10
Objective L2-English Proficiency*	67 (9)	50-81	70 (7)	53-82	3 (6)	-9-+23
Language Use						

*Frequency**	75 (13)	49-95	55 (12)	28-74	-20 (16)	-56-+17
*Type (Entropy)**	53 (18)	20-94	65 (13)	30-87	12 (23)	-29-+59
Language Switching						
<u>*Frequency*</u>						
*Language Switching Frequency**	43 (11)	23-70	48 (12)	25-71	5 (12)	-17-+31
Code-Switching Frequency Judgment	33 (13)	5-54	31 (12)	7-52	-2 (16)	-41-+48
<u>*Type*</u>						
*Dual-Language**	18 (18)	0-75	31 (19)	5-82	14 (26)	-39-+82
Dense Code-Switching	13 (18)	0-97	11 (12)	0-44	-3 (21)	-97-+38

Notes: All scores are presented on a 0-100% scale. A higher frequency of language use means more L1-Mandarin use. For the entropy (type of language use) score, a score of 0 means that participants spend all their time in a single-language environment, a score of 100% means that participants spend all their time in a dual-language environment. The difference scores are between Session B scores and baseline scores. A few measures were not completed at baseline; in those cases we used the Session A data (for the interview and vocabulary size test of the Objective Proficiency score and the code-switching frequency judgement task). Further information about how these composite scores were computed is provided in the Supplementary Materials.

2.3. Design

In all tasks (cued production, voluntary production, comprehension), there was one between-participant variable, Group (Move group vs. Control group), and three within-participant variables, Session (Session A vs. Session B), Language (L2-English vs. L1-Mandarin), and Switching or Mixing. Switching was a comparison between non-switch trials and switch trials in the dual-language blocks. Mixing was a comparison between the non-switch trials in the dual-language blocks and the trials in the single-language blocks (see Figure 1). The main dependent variable was participants' reaction times, which was measured as naming onset relative to onset of picture (production) or word (comprehension) presentation.

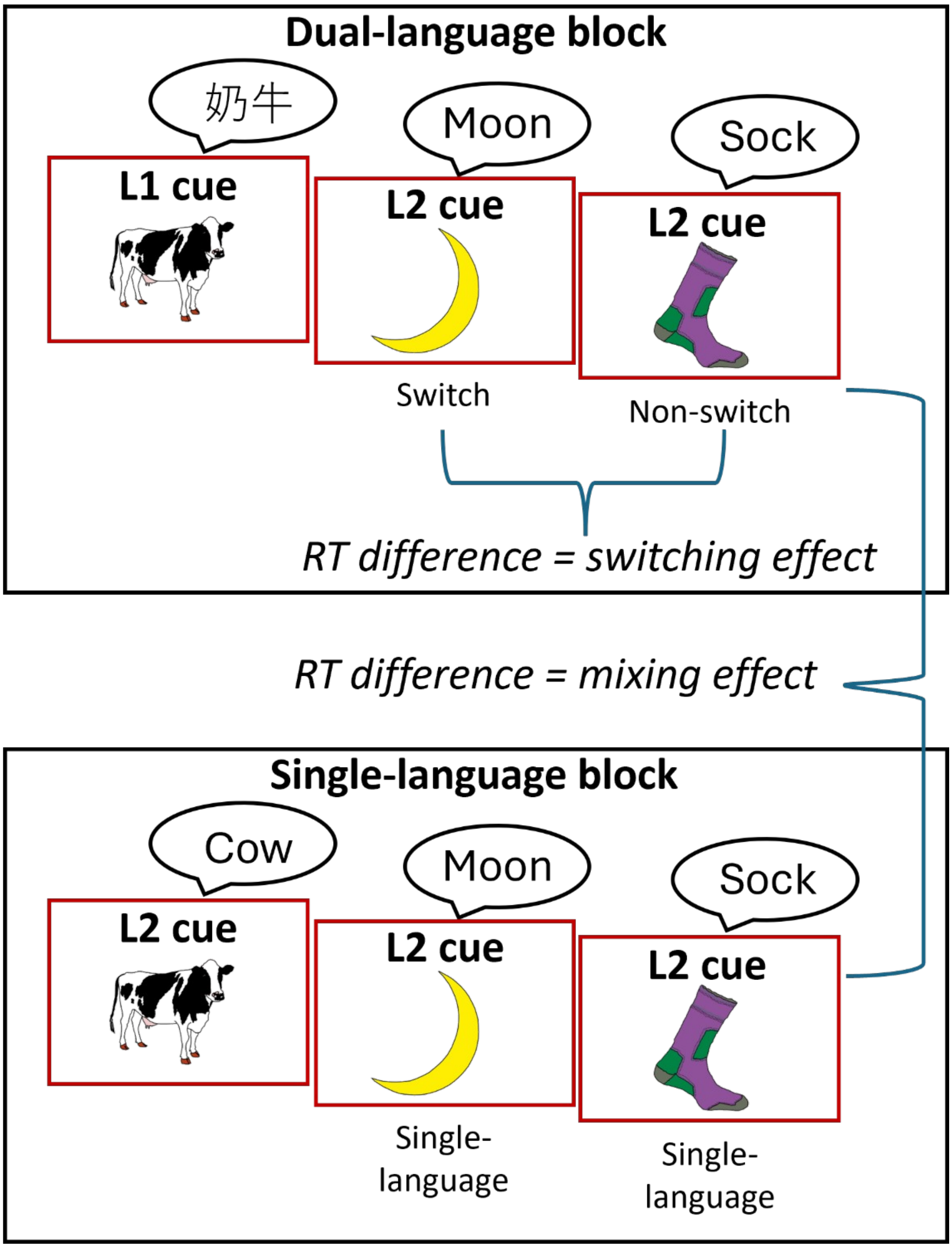

Figure 1. Overview of the dual-language and single-language blocks included in the cued picture-naming task. The switching effect is the RT (reaction time) difference between switch and non-switch trials. The mixing effect is the RT difference between single-language trials and non-switch trials from the dual-language block. Please note that, due to restrictions on the dissemination of the face cues used, we here use "L1 cue" to represent the cue given to participants; in the experiment, participants saw a face of typically East Asian or White European appearance as their language cue. In the voluntary task, trials looked similar but no longer included these face cues. Instead, switch or non-switch trials depended on the language choice of the participant for each picture. In the comprehension task, participants heard a word and pressed a button to indicate if the word referred

to an animate or inanimate entity. The images shown here and used in the study are from the MultiPic database (Duñabeitia et al., 2018; 2022).

2.4. Materials

We selected 120 simple, high-frequency pictures/words that were non-cognates between L1-Mandarin and L2-English (MultiPic, Duñabeitia et al., 2018; 2022). Items were matched in frequency across languages (see OSF page for the stimuli per list). The Mandarin words included no more than three characters and the English words no more than three syllables. 42 words referred to living entities and 78 to non-living entities, for the purpose of the animacy judgments in the comprehension task. The three experimental tasks used the same materials, but divided into lists and counterbalanced across participants so that individual participants never saw the same materials across tasks or sessions. Each task (cued, voluntary, comprehension) used a unique set of 20 individual pictures (corresponding to 20 Mandarin and 20 English words) for a given individual per session. This ensured that each participant saw all 120 items across the three tasks in the two sessions. Pictures were never repeated across tasks or sessions. That is, a given participant would see one set of pictures for the cued task, another one for the voluntary task, and a third one for the comprehension task in session A, with three other sets in session B. Within each task, each picture was presented twelve times in the dual-language block (three times per combination of language and trial type, resulting in 240 trials) and four times in the single-language blocks (resulting in 80 trials).

For the comprehension task, one female Mandarin native speaker recorded the Mandarin and the English words. The speaker had a neutral Mandarin accent and an L2-English accent that was representative of L1-Mandarin speakers of L2-English. We opted for a Mandarin speaker to record both languages to avoid differences between the Move and Control groups in exposure to British English. The mean duration of the recordings (i.e., the duration of each spoken word) was 828ms (SD = 151) for L1-Mandarin words and 795ms (SD = 146) for L2-English words ($t(238) = 1.746$, $p = .082$). Although the use of one speaker in the current study could not avoid talker-specific effects, other work (Coumel et al., 2024) compared the same task when words were spoken by one versus four speakers and showed no significant impact of that manipulation on either overall RTs or on switching costs.

A detailed description of all language-profile measures (age of acquisition, proficiency, language use, language switching) is provided in the Supplementary Materials.

2.5. Procedure

The order of the tasks in the lab for Session A and B was always: single language tasks (see de Bruin et al., 2025) – dual-language tasks – proficiency measures. The order of the three dual-language tasks was counterbalanced across participants. Approximately half of the participants (27/53 Move participants and 32/61 Control participants) completed the three dual-language tasks in the order comprehension – cued production – voluntary production. The other half completed the order voluntary production – cued production – comprehension. We only included these order options to avoid participants switching more frequently between comprehension and production tasks. This way, approximately half of the participants completed the comprehension task first (the other half the production tasks) and approximately half of the participants completed the cued before the voluntary production task (the other half completed the voluntary production task first)[1]. Each task took about 15 minutes. The tasks were presented in PsychoPy (Peirce et al., 2019) and the sessions were led in person by a Mandarin-English bilingual for both groups. Participants always saw the instructions on the screen in both languages.

In the cued production task, participants were instructed to name pictures either in L1-Mandarin or in L2-English, responding to a face cue that indicated which language to use (i.e., a face of East Asian appearance cued responses in L1-Mandarin and a face of White European appearance cued responses in L2-English). There were two different face cues (Ma et al., 2020) per language. This ensured there was no confound between language switching and cue switching, as cues always switched on both language switch and non-switch trials (Heikoop et al., 2016). In the voluntary production task, participants also named pictures but were free to name each picture in their language of choice in the dual-language parts. They were told, in both languages, that they could name each picture in Mandarin or in English and that they were free to switch between languages whenever they wanted. We asked them to try and use both languages in the task, but gave no further instructions on how often or when to use each language. Instead, we told them to use the first word that comes to mind for each picture, regardless of the language. In the comprehension

[1] We examined whether switching and mixing effects were affected by task order. In the comprehension and cued production tasks, neither mixing nor switching effects significantly differed between the two task orders. In the voluntary task, switching effects did not differ significantly either. Voluntary mixing effects were slightly larger in task order 1 (larger mixing benefit when the voluntary task was completed after the cued task). Across the tasks, and considering task order was distributed equally across groups and remained constant for a participant across testing sessions, task order effects did not appear to influence any (null) effects of Group or Session.

task, participants heard L1 or L2 words and indicated with a button press whether the word referred to a living/animate entity (press "L") or an inanimate object (press "A").

The three tasks had the same structure: familiarisation phase, practice phase in language A, single-language block in language A, practice phase in language B, single-language block in language B, dual-language practice, dual-language block, single-language block in language B, single-language block in language A. The order of languages in the single-language blocks was counterbalanced across participants.

In the familiarisation phase, participants were shown the target pictures with corresponding words (production tasks) or the spoken and written words (comprehension task) in L1-Mandarin and in L2-English. The participants were instructed to just listen to the words and/or look at the written form. We included this familiarisation phase to make sure participants could recognise each picture and spoken word easily.

Each single-language block included twenty trials (80 trials in total). The dual-language block included 240 trials (with a short break after every 80 trials). In the cued production and comprehension tasks, half of the trials were in English and the other half in Mandarin; 33% of the trials were language switch trials and these trials were distributed equally across languages. This resulted in 40 L1 switch, 40 L2 switch, 80 L1 non-switch, and 80 L2 non-switch trials in the cued production and comprehension tasks. In the comprehension task, category switches between living and non-living responses were also distributed equally across trial type and language combinations. In the voluntary task, the exact number of trials per condition depended on the participants' responses. The switching frequency in the cued and comprehension tasks was therefore chosen to match the switching frequency we expected to observe in the voluntary production task, based on previous voluntary production tasks conducted with Mandarin-English bilinguals with a similar background (de Bruin & Xu, 2023). The mean voluntary switching rate in the current study was 36% for the Move group.

In the cued production and comprehension tasks, each picture or word was presented an equal number of times in each condition. For the voluntary task, trial type and language depended on the participants' responses. The pictures were presented pseudo-randomly so that the participants would not see the same picture twice in a row. Pictures were repeated within each task but never across tasks or sessions within the same participant. The trials were pseudo-randomised with no more than four trials of the same type (language switch or non-switch), language, or animacy category (for the comprehension task) in a row. In the production tasks, a trial consisted of a 500ms fixation cross followed by the presentation of the target picture (and cue, in the cued task) for 2500ms, regardless of when the response was given. In the comprehension task, a trial consisted of a 500ms fixation cross followed by the presentation of the target word until a response was given or for 2500ms if no response was given.

2.6. Data analysis

For all three tasks, we analysed the data with mixed-effect models in R (4.4.1) using the packages lme4 version 1.1.35.4 (Bates et al., 2015) and lmerTest version 3.1.3 (Kuznetsova et al., 2017). We started the analysis with a maximal model (Barr et al., 2013) that included random intercepts for participants and items, and all within-participant/-item predictors. In the case of non-convergence, we removed the correlations between random slopes and random intercepts. We then reduced the random effect structure by removing the random slopes (across item and participant slopes) explaining the least variance until reaching convergence (see the "ModelOutput" file on the OSF page for a description of the steps taken per model, as well as the information for the random effects). The final random-effects structure of the models can be found in the tables. All independent variables were two-level categorical fixed effects and coded as follows: Language (Mandarin (-0.5), English (+0.5)), Session (Session A (-0.5), Session B (+0.5)), and Group (Control (-0.5), Move (+0.5)). For the switching effects, Switching was coded as non-switch (-0.5) vs. switch (+0.5) trials in the dual-language blocks; for the mixing effects, Mixing was coded as single-language (blocked) trials (-0.5) vs. non-switch trials in the dual-language block (+0.5; see Figure 1).

For all three tasks, we only included participants who scored at least 70% accuracy. In the cued production task, we therefore had to exclude three participants, resulting in 111 participants in the analysis. In the voluntary production task, no participants were excluded due to not reaching 70% accuracy. However, we excluded one participant because they named each picture in *both* languages in Session A, three participants because their responses did not include trials in one of the conditions (e.g., no L1 non-switch trials), and two participants because one of the conditions did not have any trials left after removal of inaccurate responses and RT outlier removal. One additional participant could not complete the voluntary task in Session B due to technical difficulties. After removal of these participants, the sample size was 107. In the comprehension task, we excluded one participant for whom the data had not been recorded in Session A, leaving 113 participants.

Per participant and combination of conditions, we removed RT outliers (2.5*SD* above/below mean per participant and condition combination, plus trials faster than 300ms; Grange, 2015). This removed 1.7% of correct trials from the cued analysis, 1.9% of voluntary trials, and 2.3% of comprehension trials. In all tasks, we also excluded inaccurate trials and trials occurring after a break, as these trials were neither switch nor non-switch trials. For the production tasks, we also removed trials preceded by no response (cued and voluntary) or an incorrect language (cued), as trial type could not be determined.

The main analyses focused on RTs as the dependent variable (DV). RTs in the production tasks were scored using CheckVocal (Protopapas, 2007, using CheckFile). RTs were not normally distributed and therefore log transformed for the analyses but we report the means of the untransformed RTs in the text and tables.

In addition to the pre-registered main analyses reported here, we also report the following additional analyses in the Supplementary Materials (also available on https://osf.io/a24xv/). The first analysis is accuracy analyses in the cued task (pre-registered). We did not examine accuracy in the voluntary task, where accuracy was close to ceiling, or the comprehension task, where accuracy reflected incorrect animacy judgements. The second section of the Supplementary Materials presents further exploratory analyses on switching frequency and language choice in the voluntary task. Third, we present exploratory Bayesian analyses to quantify evidence for/against an effect of language-environment move on language control (switching and mixing effects in each task). Fourth, we report exploratory analyses comparing switching and mixing effects across the three tasks (cued production, voluntary production, comprehension). Finally, further pre-registered analyses examining individual differences in language profile are reported in section 5 of the Supplementary Materials.

3. Results

3.1. Cued production task

Overall accuracy in the cued task was high in both groups and in both sessions (Control Group: Session A: *M* = 93.2%, *SD* = 6.2; Session B: *M* = 92.9%, *SD* = 6.0; Move Group: Session A: *M* = 95.0%, *SD* = 5.0; Session B: *M* = 95.2%, *SD* = 4.7). Further analyses are provided in the Supplementary Materials.

3.1.1. RT: Switching effects

Table 5 shows the full statistics from the RT analysis, Table 6 the means by condition, and Figure 2 the switching costs in the Move group. The RT analysis showed a main effect of Language, with slower responses in L1-Mandarin (*M* = 1234ms, *SD* = 133) than in L2-English (*M* = 1109ms, *SD* = 118). It also showed a main effect of Switching, with shorter RTs on non-switch (*M* = 1141ms, *SD* = 118) than on switch (*M* = 1237ms, *SD* = 132) trials, indicating a switching cost. However, neither overall RTs nor switching costs changed across sessions. Overall RTs remained stable across sessions (*M* Session A = 1179ms, *SD* = 139; *M* Session B = 1168ms, *SD* = 125). Similarly, the switching costs were comparable across sessions (*M* cost Session A = 99ms, *SD* = 61; *M* Session B = 94ms, *SD* = 67). Switching costs also did not differ significantly between groups (*M* cost Move group = 102ms, *SD* = 57; *M* cost Control group= 91ms, *SD* = 53). Moreover, there was no significant interaction between the Switching cost, Session, and Group (see Table 5). Thus, the cued RT data showed no significant changes in terms of the switching cost across sessions, or differences between groups across sessions.

Switching did interact with Language. The switching cost was larger when switching to L2-English (*M* cost = 133ms, *SD* = 63) than to L1-Mandarin (*M* cost = 62ms, *SD* = 68). Finally, Language interacted with Group. The Move group was significantly faster than the Control group in L2-English (*M* Move group = 1081ms, *SD* = 113; *M* Control group = 1133ms, *SD* = 118; β = -0.048, *SE* = 0.019, *t* = -2.511, *p* = .014) while there was no significant difference in L1-Mandarin (*M* Move group = 1228ms, *SD* = 135; *M* Control group = 1239ms, *SD* = 131; β = -0.012, *SE* = 0.020, *t* = -0.589, *p* = .557). Importantly, this effect of Group did not interact with Session nor Switching, suggesting this reflected a baseline difference between groups rather than an effect related to language control or environment changes.

Table 5. Outcome of the linear mixed effect model for the analysis of switching effects in the cued production task. The final model converged with by-participant and by-item intercepts and all slopes apart from item slopes for Switching x Session, Switching x Language x Group, Switching x Session x Language, and Switching x Session x Group. For significant effects, the *p* value is stated in bold.

Fixed effects	**Estimate**	**Standard Error**	**t-value**	***p*-value**
Intercept	7.052	0.010	693.563	**<.001**
Switching	0.086	0.005	19.001	**<.001**
Session	-0.012	0.008	-1.479	.142
Language	-0.098	0.007	-13.593	**<.001**
Group	-0.030	0.019	-1.585	.116
Switching x Session	-0.003	0.005	-0.643	.522
Switching x Language	0.063	0.006	9.795	**<.001**
Session x Language	0.003	0.007	0.383	.702
Switching x Group	0.008	0.008	1.027	.307
Session x Group	-0.0001	0.016	-0.004	.996
Language x Group	-0.036	0.011	-3.303	**.001**
Switching x Session x	0.008	0.009	0.884	.378

Language				
Switching x Session x Group	0.012	0.009	1.377	.172
Switching x Language x Group	0.012	0.011	1.158	.249
Session x Language x Group	-0.025	0.014	-1.737	.085
Switching x Session x Language x Group	-0.010	0.019	-0.524	.601

Table 6. Reaction times, switching, and mixing effects (reported in ms) in the cued production task.

	Session A		Session B	
		Move group		
	Mandarin	**English**	**Mandarin**	**English**
Single-language	1011 (128)	928 (114)	1011 (109)	926 (114)
Non-switch	1212 (160)	1048 (134)	1204 (127)	1026 (119)
Switch	1276 (176)	1185 (166)	1270 (149)	1167 (132)
Switching effect	64 (80)	138 (75)	66 (87)	141 (79)
Mixing effect	201 (107)	119 (95)	193 (101)	100 (100)
		Control group		
	Mandarin	**English**	**Mandarin**	**English**
Single-language	1037 (142)	1014 (119)	1003 (140)	992 (106)
Non-switch	1228 (147)	1091 (132)	1215 (155)	1093 (133)
Switch	1297 (148)	1220 (152)	1263 (137)	1218 (153)
Switching effect	69 (79)	129 (79)	48 (89)	124 (74)
Mixing effect	191 (101)	78 (105)	212 (105)	101 (101)

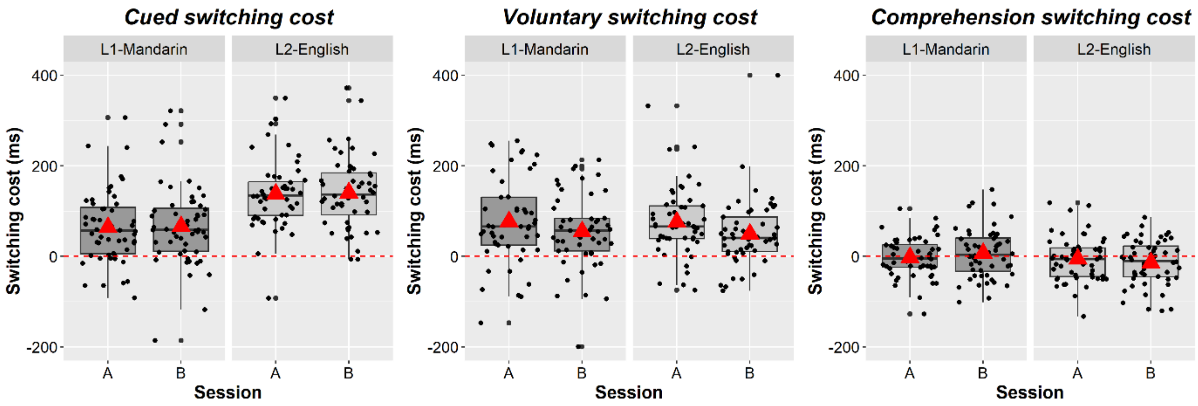

Figure 2. Switching effects (RT difference between switch and non-switch trials) per task (cued production, voluntary production, comprehension) for the Move group. Within each task plot, the left panel shows the L1 and the right panel the L2. Within each panel, the left boxplot shows Session A and the right boxplot Session B. Red triangles represent mean switching costs; the black horizontal lines in the box plots represent the medians; the whiskers represent the lower and the upper quartiles; dots represent individual means.

3.1.2. RT: Mixing effects

The RT analysis (see Tables 6 and 7, and Figure 3) showed a main effect of Language, reflecting (in line with the previous analysis) slower responses in L1-Mandarin than in L2-English. It also showed a main effect of Mixing, with slower naming on non-switch trials in the dual-language block (*M* = 1141ms, *SD* = 118) than on single-language trials (*M* = 991ms, *SD* = 93), indicating a mixing cost. There was no significant main effect of Session on overall RTs, similar to the switching-cost analysis, reflecting that RTs remained stable across sessions (*M* Session A = 1095ms, *SD* = 121; *M* Session B = 1083ms, *SD* = 111).

Mixing interacted with Session and Group. In the Control Group, the mixing cost was somewhat higher in Session B (*M* cost = 160ms, *SD* = 85) than in Session A (*M* cost = 138ms, *SD* = 79). In the Move group the opposite pattern was present, with a slightly higher mixing cost in Session A (*M* cost = 160ms, *SD* = 83) than in Session B (*M* cost = 145ms, *SD* = 82). However, neither the Move group (β = -0.011, *SE* = 0.010, *t* = -1.115, *p* = .270) nor the Control Group (β = 0.016, *SE* = 0.008 *t* = 1.888, *p* = .064) showed a significant interaction between Mixing and Session. This suggests the three-way interaction reflects patterns going in the opposite direction for the Move and Control group (largely driven by the Control group), rather than significant mixing-cost changes in the Move group.

The mixing cost also interacted with language. The mixing cost was significantly higher in L1-Mandarin (*M* cost = 199ms, *SD* = 87) than in L2-English (*M* cost = 98ms, *SD* = 86). Similar to the switching-cost analysis, Language also interacted with Group. This again reflected that the Move

Group showed a larger difference between the L1 and L2 than the Control Group. This, together with the main effect of Group, appeared most strongly driven by relatively long L2 single-language RTs in the Control group.

Thus, in summary, the cued RT analyses showed significant switching and mixing costs. Switching costs were larger when switching to L2-English while mixing costs were larger when using L1-Mandarin. Neither overall RTs nor switching or mixing costs showed significant changes between sessions.

Table 7. Outcome of the linear mixed effect model for the analysis of mixing effects for the cued production task. The final model included by-participant and by-item random intercepts, all participant slopes, and item slopes for all main effects, plus the interactions for Mixing x Language, Session x Language, Session x Group, Language x Group, and Session x Language x Group.

Fixed effects	**Estimate**	**Standard Error**	**t-value**	***p*-value**
Intercept	6.940	0.010	723.786	**<.001**
Mixing	0.140	0.006	23.398	**<.001**
Session	-0.011	0.008	-1.377	.172
Language	-0.087	0.009	-10.049	**<.001**
Group	-0.039	0.017	-2.300	**.023**
Mixing x Session	0.002	0.006	0.291	.772
Mixing x Language	-0.083	0.008	-10.382	**<.001**
Session x Language	0.002	0.008	0.281	.779
Mixing x Group	0.010	0.011	0.868	.387
Session x Group	0.008	0.016	0.474	.637
Language x Group	-0.058	0.014	-4.214	**<.001**
Mixing x Session x Language	-0.007	0.011	-0.632	.529
Mixing x Session x Group	-0.027	0.013	-2.082	**.040**
Mixing x Language x Group	0.030	0.016	1.927	.057
Session x Language x Group	-0.017	0.015	-1.108	.270
Mixing x Session x Language x Group	-0.006	0.022	-0.270	.788

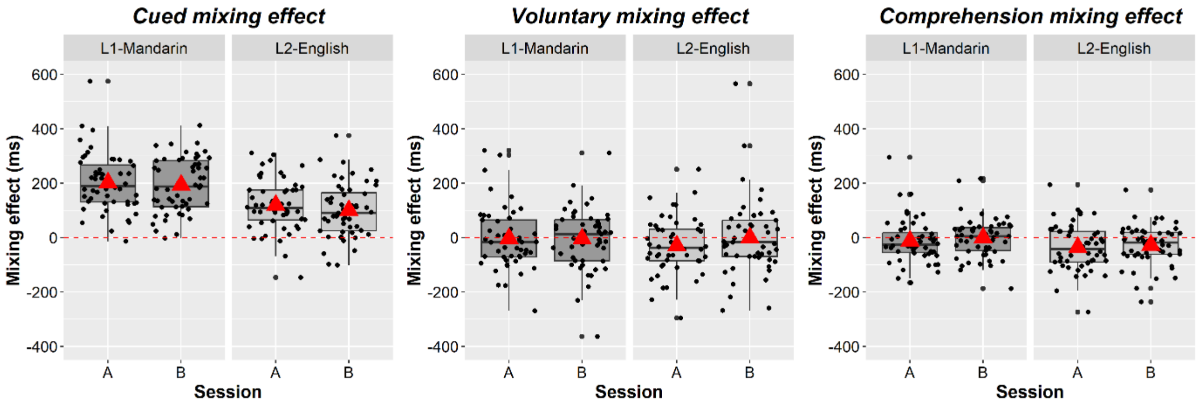

Figure 3. Mixing effects (RT difference between non-switch and single-language trials) per task (cued production, voluntary production, comprehension) for the Move group. Within each task plot, the left panel shows the L1 and the right panel the L2. Within each panel, the left boxplot shows Session A and the right boxplot Session B. Red triangles represent mean mixing effects; the black horizontal lines in the box plots represent the medians; the whiskers represent the lower and the upper quartiles; dots represent individual data points.

3.2. Voluntary production task

3.2.1. Accuracy, switching frequency, and language choice

Accuracy was high in both groups and in both sessions (Move group: *M* Session A = 97.4%, *SD* = 3.4; *M* Session B = 97.3%, *SD* = 4.5; Control group: *M* Session A = 96.3%, *SD* = 3.9; *M* Session B = 96.8%, *SD* = 3.8). Most mistakes concerned using a different picture name than the target. As accuracy was well above 95%, it was not analysed further. The Supplementary Materials (Section 2) provide further exploratory analyses and information on participants' switching frequency and language choice in the task. Mean switching frequency in the Move group was 35.7% switches (*SD* = 14.6) in Session A and 36.2% switches (*SD* = 13.1) in Session B. In the Control group, mean switching frequency was 39.7% (*SD* = 13.5) in Session A and 39.0% (*SD* = 11.8) in Session B. Switching frequency did not change significantly between sessions and did not differ between the two groups. In terms of language choice, in both groups both languages were used frequently, with slightly more frequent use of English than Mandarin in both groups.

3.2.2. RT: Switching effects

The RT analysis (see Table 8 for the statistics, Table 9 for the means, and Figure 2 for the Move group switching costs) showed a main effect of Language, with slower responses in L1-Mandarin (*M* =

1042ms, *SD* = 134) than in L2-English (*M* = 944ms, *SD* = 111). There was a main effect of Group, whereby the Move group (*M* = 948ms, *SD* = 105) was overall faster than the Control group (*M* = 1010ms, *SD* = 117). The analysis also showed a main effect of Switching, with longer naming times on switch (*M* = 1029ms, *SD* = 123) than on non-switch (*M* = 955ms, *SD* = 114) trials, indicating switching costs. Overall RTs did not change between sessions (*M* Session A = 983ms, *SD* = 124; *M* Session B = 979ms, *SD* = 133).

Of main interest for the current study, the switching cost did differ significantly between the two sessions and was smaller in Session B (*M* cost = 64ms, *SD* = 62) than in Session A (*M* cost = 83ms, *SD* = 72, see Figure 2). Importantly, the switching cost did not differ significantly between the two groups, and the switching-cost decrease between Session A and B was also comparable for the Move group and Control Groups (see Table 8).

Switching interacted with Language, reflecting that the switching cost was larger when switching to L2-English (*M* cost = 67ms, *SD* = 56) than to L1-Mandarin (*M* cost = 50ms, *SD* = 71). This asymmetry in switching cost did not differ between the two sessions (see Table 9).

Finally, Language interacted with Group. This reflected that, similar to the cued task, the RT difference between L1-Mandarin and L2-English responses was larger in the Move group (see Table 8). The difference was mostly driven by the L2-English trials, where the Move group was significantly faster than the Control group (β = -0.074, *SE* = 0.020, *t* = -3.655, $p < .001$), while the group difference did not reach significance in L1-Mandarin (β = -0.042, *SE* = 0.023, *t* = -1.794, *p* = .076). The three-way interaction between Session x Language x Group reached a *p* value of .050, but neither by-language analysis showed a significant Session x Group interaction, suggesting these language differences between Groups did not change with time.

Table 8. Outcome of the linear mixed effect model for the analysis of switching effects for the voluntary production task. The final model included by-participant and by-item intercepts, all by-participant random slopes, and all by-item slopes apart from Switching x Session, Switching x Session x Group, and Switching x Language x Session.

Fixed effects	**Estimate**	**Standard Error**	**t-value**	***p*-value**
Intercept	6.877	0.011	609.958	**<.001**
Switching	0.057	0.004	14.056	**<.001**
Session	-0.010	0.009	-1.019	.311
Language	-0.077	0.008	-9.918	**<.001**
Group	-0.058	0.021	-2.745	**.007**

Switching x Session	-0.013	0.006	-2.280	**.025**
Switching x Language	0.016	0.006	2.635	**.010**
Session x Language	0.005	0.008	0.699	.486
Switching x Group	0.014	0.008	1.692	.094
Session x Group	0.011	0.019	0.604	.547
Language x Group	-0.031	0.012	-2.636	**.010**
Switching x Session x Language	-0.010	0.010	-0.977	.331
Switching x Session x Group	-0.016	0.012	-1.381	.170
Switching x Language x Group	-0.019	0.012	-1.566	.120
Session x Language x Group	-0.030	0.015	-1.983	.050
Switching x Session x Language x Group	0.010	0.021	0.472	.638

Table 9. Reaction times, switching costs, and mixing effects (in ms) in the voluntary production task.

	Move group			
	Session A		**Session B**	
	Mandarin	**English**	**Mandarin**	**English**
Single-language	989 (127)	918 (111)	1005 (123)	893 (111)
Non-switch	985 (139)	887 (102)	1000 (140)	892 (125)
Switch	1062 (150)	963 (132)	1055 (157)	942 (138)
Switching effect	77 (96)	76 (77)	55 (83)	50 (77)
Mixing effect	-4 (116)	-31 (104)	-5 (113)	-1 (141)
	Control group			
	Session A		**Session B**	
	Mandarin	**English**	**Mandarin**	**English**
Single-language	1032 (137)	1004 (123)	1026 (151)	1015 (136)
Non-switch	1054 (141)	949 (114)	1031 (166)	959 (129)
Switch	1093 (150)	1025 (140)	1071 (146)	1013 (140)
Switching effect	39 (77)	75 (62)	40 (82)	54 (72)
Mixing effect	22 (104)	-55 (114)	5 (104)	-56 (117)

Though overall switching costs decreased from Session A to Session B, these decreases, as well overall switching costs, were similar across groups, suggesting they were not specifically related to changes in language environment. We therefore conducted a further exploratory analysis to better understand this change. We examined whether changes in switching costs related to changes in switching frequency, as previous research has suggested a larger switching frequency relates to a lower switching cost within the task (de Bruin et al., 2018). While overall switching frequency in the task did not change between sessions, at an individual level, some people switched more frequently in Session B, while others switched less often. A participant's change in switching frequency indeed correlated with their decrease in switching costs ($r(105) = -0.274$, $p = .004$). Participants who increased their switching frequency more between sessions showed a larger decrease in switching costs between sessions (see Figure 4). Further information about the relationship between a participants' language profile and their changes in switching costs and frequency is provided in the Supplementary Materials.

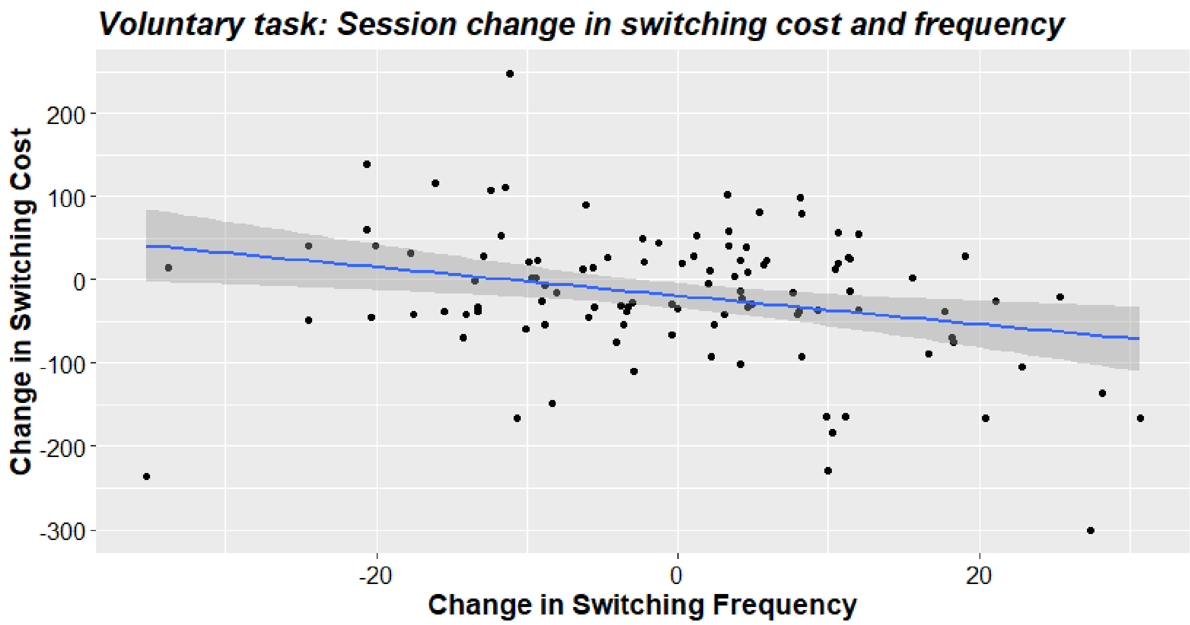

Figure 4. The relationship between a participant's change in switching frequency in the voluntary task (Session B versus Session A) and their change in Switching Costs (difference between RTs on switch – non-switch trials in Session B versus Session A). Positive numbers on the x-axis reflect an increase in switching frequency between sessions. Negative numbers on the y-axis reflect a decrease in switching costs between sessions. The figure includes both Move and Control participants.

3.2.3. RT: Mixing effects

The RT analysis for the mixing effect is reported in Table 10, with the means reported in Table 9 and Figure 3. Similar to the switching cost analysis, responses were slower in L1-Mandarin than in L2-English and the Control group responded more slowly than the Move group overall. There was no significant difference in RTs between non-switch (*M* = 955ms, *SD* = 114) and single-language trials (*M* = 987ms, *SD* = 101), reflecting no significant mixing effect. Of main interest for the study, this (absence of a) mixing cost did not change between Session A (*M* effect = -32ms, *SD* = 94) and B (*M* effect = -29ms, *SD* = 103). Similar to the switching analysis, overall RTs did not differ between sessions either.

The mixing effect did interact with language. While L1-Mandarin showed no significant mixing effect (*M* effect = 5ms, *SD* = 92, β = 0.004, *SE* = 0.008, *t* = 0.548, *p* = .585), there was a significant mixing benefit in L2-English (*M* effect = -40ms, *SD* = 107, β = -0.027, *SE* = 0.010, *t* = -2.866, *p* = .005) with faster responses on English non-switch than single-language trials.

Similar to the switching analysis, there was an interaction between Language and Group. The Move group was significantly faster on L2-English trials than the Control group (β = -0.093, *SE* = 0.018, *t* = -5.243, $p < .001$). In L1-Mandarin, a similar direction was observed but this difference did not reach significance (β = -0.043, *SE* = 0.022, *t* = -1.978, *p* = .051). This also interacted with Session, reflecting that this pattern was most pronounced in Session B. The Move group showed a significant interaction between Session and Language (β = -0.026, *SE* = 0.012, *t* = -2.164, *p* = .035), reflecting that the RT difference between languages increased between sessions (see Table 9). In contrast, the Control Group showed the opposite effect (β = 0.022, *SE* = 0.010, *t* = 2.318, *p* = .024), with the RT difference between languages decreasing between sessions.

Finally, there was a three-way interaction between Mixing, Language, and Group. The asymmetry in mixing effect was present in the Control Group only (β = -0.057, *SE* = 0.012, *t* = -4.613, $p < .001$). In that group, English showed a significant mixing benefit ($p < .001$) while Mandarin did not show a significant mixing effect (*p* = .136). In contrast, in the Move group there was no significant interaction (β = -0.004, *SE* = 0.014, *t* = -0.263, *p* =0.794), with neither language showing a significant mixing effect (see Table 9). However, this did not interact with Session, suggesting a baseline difference between groups.

To summarise, the voluntary task showed a switching cost but no mixing cost. The switching cost decreased between sessions, but this was related to how often participants were switching within the task itself and the switching-cost decrease was also found for the control group. This suggests this decrease was not due to a change in language environment. The mixing cost did not change between sessions.

Table 10. Outcome of the linear mixed effect model for the analysis of mixing effects for the voluntary production task. The final model included by-participant and by-item random intercepts, all by-participant slopes, and all by-item slopes apart from Mixing x Session x Group and Mixing x Session.

Fixed effects	Estimate	Standard Error	t-value	*p*-value
Intercept	6.856	0.010	683.622	**<.001**
Mixing	-0.012	0.008	-1.549	.124
Session	-0.004	0.008	-0.430	.668
Language	-0.068	0.010	-6.899	**<.001**
Group	-0.068	0.018	-3.759	**<.001**
Mixing x Session	0.003	0.007	0.367	.714
Mixing x Language	-0.030	0.009	-3.221	**.002**
Session x Language	-0.002	0.008	-0.252	.801
Mixing x Group	0.006	0.015	0.389	.698
Session x Group	0.007	0.017	0.385	.701
Language x Group	-0.050	0.016	-3.154	**.002**
Mixing x Session x Language	0.020	0.012	1.671	.098
Mixing x Session x Group	0.025	0.015	1.661	.100
Mixing x Language x Group	0.052	0.018	2.897	**.005**
Session x Language x Group	-0.048	0.016	-3.082	**.003**
Mixing x Session x Language x Group	0.014	0.023	0.584	.561

3.1.3. Comprehension task

Overall accuracy was high in the Comprehension task (Move group: *M* Session A = 93.8%, *SD* = 4.6; *M* Session B = 94.4%, *SD* = 4.1; Control Group: *M* Session A = 93.5%, *SD* = 5.3; *M* Session B = 93.4%, *SD* = 5.3). Given that accuracy was high and errors reflected semantic categorisation mistakes (animate/inanimate judgments) rather than language mistakes, accuracy was not analysed further.

3.3.1. RT: Switching effects

The RT analysis (see Table 11 for the full statistics and Table 12 for the means, as well as Figure 2) revealed a main effect of Language. Participants responded faster to L1-Mandarin (*M* = 1033ms, *SD* =

131) than to L2-English words (*M* = 1131ms, *SD* = 136). There was no significant main effect of Switching, which meant that there was no overall switching cost. However, Switching did interact with Language. Only L1-Mandarin showed a small but significant switching cost (*M* cost = 9ms, *SD* = 38; β = 0.008, *SE* = 0.004, *t* = 2.317, *p* = .022). For L2-English, the reverse pattern was observed (*M* cost = -8ms, *SD* = 38), although this did not reach significance (β = -0.008, *SE* = 0.004, *t* = -1.840, *p* = .068).

Of main interest for the current study, Session did not interact with the switching cost, nor did the three-way interaction between Switching, Session, and Language reach significance. This suggests the switching costs did not change with time (*M* cost Session A = 2ms, *SD* = 38; *M* cost Session B = -1ms, *SD* = 36; see Table 12 for the switching costs by Language and Group per Session). Overall RTs did show a main effect of Session, whereby participants were faster in Session B (*M* = 1055ms, *SD* = 125) than in Session A (*M* = 1106ms, *SD* = 148; see Tables 11-12). However, this did not interact with Group, suggesting this decrease was comparable for the Move and Control groups (see Table 12). This decrease in RTs is likely to reflect a practice effect in relation to the animacy judgment task. This Session effect did interact with Language, suggesting this RT decrease was larger for L2-English (*M* change = -59ms, *SD* = 116) than for L1-Mandarin (*M* change = -44ms, *SD* = 104).

Finally, there was a small overall difference between groups in terms of overall RTs, which interacted with Language. The two groups did not differ in L2-English RTs (*M* Move group = 1139ms, *SD* = 134; *M* Control group = 1124ms, *SD* = 138; β = 0.015, *SE* = 0.022, *t* = 0.682, *p* = .497). However, the Control group was significantly faster in L1-Mandarin (*M* = 1000ms, *SD* = 124) than the Move group (*M* = 1071ms, *SD* = 129; β = 0.066, *SE* = 0.021, *t* = 3.156, *p* = .002). Importantly these differences between groups did not interact with Session, suggesting they reflected baseline differences.

Table 11. Outcome of the linear mixed effect model for the analysis of switching effects for the Comprehension task. The final model included by-participant and by-item intercepts, all by-participant slopes, and all by-item slopes apart from Switching x Session, Switching x Language x Session, Switching x Language x Group, and Switching x Language x Group x Session.

Fixed effects	**Estimate**	**Standard Error**	**t-value**	***p*-value**
Intercept	6.962	0.011	643.715	**<.001**
Switching	0.0001	0.003	0.037	.971
Session	-0.043	0.008	-5.448	**<.001**
Language	0.089	0.008	10.777	**<.001**

Group	0.041	0.020	1.992	**.049**
Switching x Session	-0.002	0.004	-0.477	.634
Switching x Language	-0.016	0.006	-2.683	**.008**
Session x Language	-0.011	0.005	-2.221	**.029**
Switching x Group	-0.007	0.004	-1.771	.081
Session x Group	-0.013	0.016	-0.807	.421
Language x Group	-0.052	0.013	-4.096	**<.001**
Switching x Session x Language	-0.007	0.007	-0.915	.362
Switching x Session x Group	0.006	0.008	0.715	.476
Switching x Language x Group	0.012	0.008	1.458	.148
Session x Language x Group	-0.015	0.010	-1.530	.129
Switching x Session x Language x Group	-0.013	0.014	-0.887	.377

Table 12. Reaction times, switching, and mixing effects (in ms) in the Comprehension task.

	Move group			
	Session A		**Session B**	
	Mandarin	**English**	**Mandarin**	**English**
Blocked trials	1109 (159)	1213 (174)	1045 (125)	1136 (153)
Non-switch trials	1096 (145)	1178 (158)	1044 (137)	1108 (137)
Switch trials	1094 (151)	1172 (163)	1051 (129)	1094 (133)
Switching effect	-3 (44)	-6 (50)	7 (55)	-14 (50)
Mixing effect	-13 (81)	-36 (85)	-1 (73)	-28 (72)
	Control group			
	Session A		**Session B**	
	Mandarin	**English**	**Mandarin**	**English**
Blocked trials	1016 (136)	1181 (160)	973 (115)	1126 (133)
Non-switch trials	1014 (144)	1151 (162)	976 (126)	1105 (141)
Switch trials	1033 (157)	1145 (164)	988 (122)	1095 (143)
Switching effect	19 (52)	-6 (55)	12 (53)	-10 (57)
Mixing effect	-2 (77)	-30 (76)	3 (76)	-21 (71)

3.3.2. Mixing effects

Similar to the switching-cost analysis, there was a main effect of Language reflecting faster L1 than L2 responses (See Tables 12 and 13). There was also a significant effect of Mixing (see Figure 3). Participants were slower on single-language trials (*M* = 1095ms, *SD* = 123) than on non-switch trials in the dual language block (*M* = 1080ms, *SD* = 127). This interacted with Language. L1-Mandarin showed no significant mixing effect (*M* difference = -3ms, *SD* = 53; β = -0.004, *SE* = 0.005, *t* = -0.749, *p* = .455), while L2-English did (*M* difference = -29ms, *SD* = 55; β = -0.024, *SE* = 0.005, *t* = -5.056, *p* = <.001).

Of main interest for the current study, Mixing did not interact with Session. The mixing effect did not change between Session A (*M* effect = -19, *SD* = 66) and B (*M* effect = -10, *SD* = 57), and this did not interact with Group either (see Table 12 for the mixing effects by Language and by Group per Session). Similar to the switching analysis, there was a main effect of Session, reflecting overall shorter RTs in Session B than in Session A. Again, this did not interact with Group.

Finally, similar to the switching analysis, the two groups differed in L1-Mandarin production, which was faster in the Control than Move group (β = 0.076, *SE* = 0.021, *t* = 3.713, *p* <.001), while no group difference was found in L2-English (β = 0.017, *SE* = 0.021, *t* = 0.805, *p* = .423). Again, this did not interact with Session or Mixing.

Table 13. Outcome of the linear mixed effect model for the analysis of mixing effects for the Comprehension task. The final model included by-participant and by-item random intercepts, all by-participant slopes, and by-item slopes for the four main effects, Mixing x Language, Session x Language, Session x Group, Language x Group, and Session x Language x Group

Fixed effects	Estimate	Standard Error	t-value	*p*-value
Intercept	6.969	0.011	659.127	**<.001**
Mixing	-0.014	0.004	-3.435	**<.001**
Session	-0.046	0.007	-6.223	**<.001**
Language	0.107	0.009	12.034	**<.001**
Group	0.047	0.020	2.356	**.020**
Mixing x Session	0.008	0.007	1.163	.247
Mixing x Language	-0.020	0.005	-3.831	**<.001**
Session x Language	-0.005	0.006	-0.880	.381
Mixing x Group	-0.005	0.006	-0.832	.407
Session x Group	-0.016	0.015	-1.128	.262

Language x Group	-0.059	0.013	-4.661	**<.001**
Mixing x Session x Language	-0.004	0.009	-0.397	.692
Mixing x Session x Group	0.002	0.013	0.177	.860
Mixing x Language x Group	0.004	0.009	0.407	.685
Session x Language x Group	-0.009	0.012	-0.698	.487
Mixing x Session x Language x Group	-0.002	0.018	-0.094	.925

To summarise, the comprehension task showed a switching cost in the L1 but not in the L2. No mixing cost was observed either. Neither switching nor mixing effects changed between sessions, but overall RTs decreased. However, this decrease in RTs was present in both groups, suggesting this was related to task practice and not due to a change in language environment.

3.4. Task comparisons

Across the tasks, no effects of Session were observed that were specific to the Move group, suggesting that neither overall RTs nor language control changed in response to a change in language environment. However, differences were observed between the tasks in terms of the language-control measures themselves (i.e., the mixing and switching costs; see Table 14 for an overview). Below, we summarise these differences between the tasks. We also ran further exploratory analyses to compare the switching and mixing effects across the three tasks (cued production, voluntary production, and comprehension). The full results are reported in Section 4 of the Supplementary Materials ("task comparisons").

Participants showed larger switching costs in each of the two production tasks compared to the comprehension task. The switching cost was also significantly larger in the cued than voluntary production task. In the two production tasks, participants showed larger L2 than L1 switching costs (and this asymmetry was larger in the cued than voluntary task). The comprehension task showed a larger L1 than L2 switching cost.

Only the cued task showed a mixing cost, with the mixing effect not differing between the voluntary production and comprehension tasks. The cued task showed a larger L1 than L2 mixing cost. Both the voluntary and the comprehension task showed no L1 mixing cost and an L2 mixing benefit.

In terms of overall RTs, only small differences were observed. When looking at the dual-language trials, cued naming times were significantly slower than comprehension or voluntary RTs. Voluntary RTs were faster than responses in the other two tasks. In the mixing-effect analysis, comprehension RTs were significantly slower than the two production tasks. In terms of RTs per language, both production tasks showed faster L2 than L1 responses (a difference, similar to the switching cost asymmetry, that was more pronounced in the cued than voluntary task). The comprehension task showed faster L1 than L2 responses.

Table 14. Summary of the findings per task, listing the presence of switching and mixing costs; whether these effects and overall RTs changed with Session (indicating an effect of change in language environment, if only present in the Move group and not in the Control group); and whether any language differences were observed in terms of the switching/mixing effect or overall RTs.

RT Measure per task	Cued production	Voluntary production	Comprehension
Switching cost			
Present?	Yes	Yes	L1: Yes; L2: No
Change with Session?	No	Yes (Move & Control)	No
Language difference?	L2 cost > L1 cost	L2 cost > L1 cost	L1 cost > L2 cost
Mixing cost			
Present?	Yes	L1: No; L2: benefit	L1: No; L2: benefit
Change with Session?	No	No	No
Language difference?	L1 cost > L2 cost	Benefit in L2, not L1	Benefit in L2, not L1
Overall RTs			
Change with Session?	No	No	Yes (Move & Control)
Language difference?	L2 RTs < L1 RTs	L2 RTs < L1 RTs	L1 RTs < L2 RTs

4. General Discussion

This study aimed to assess how and when language control adapts to the context a bilingual is communicating in. We examined this at two levels: language control in response to the immediate (switching) environment (RQ1) and potential changes in language control after a change in a bilingual's overall language environment (RQ2). We examined Mandarin-English bilinguals' language-control mechanisms during a cued production task, a voluntary production task, and a comprehension task. We tested a group of bilinguals who moved from China, an L1-dominated environment, to the UK, a dual-language environment potentially dominated by the L2 (Move group). This group was compared to a group of bilinguals who stayed in China throughout the study (Control group). Both groups completed a series of language-switching tasks at multiple time points, which for the Move group compared their language control immediately after arrival in the UK and after seven months.

Overall, regarding RQ1, we found that the patterns of switching and mixing effects varied across tasks, with bilinguals showing the largest mixing and switching costs in the cued production task. Switching costs were smaller in the voluntary production task and only present in one language in the comprehension task. Mixing costs were absent in the voluntary production and comprehension tasks. Regarding RQ2, however, even though bilinguals in the Move group significantly increased their L2 use and time in dual-language contexts, no significant language-control changes were observed after a change in language environment. This suggests that bilinguals' language control mechanisms might adapt more to the immediate context than to the overall daily-life language environment, at least in the current bilinguals and circumstances tested here (as discussed further below).

4.1. The influence of immediate interactional contexts and task (RQ1)

4.1.1. Cued production task

As expected, and in line with previous literature, we observed significant switching costs in the cued production task (e.g., Costa & Santesteban, 2004; Meuter & Allport, 1999; see Gade et al., 2021 for a meta-analysis). This finding indicates that switching can be effortful and recruit reactive language control. This was observed with naturalistic cues in the form of faces (see Blanco-Elorrieta & Pylkkänen, 2017). Participants also showed significant mixing costs, reflecting proactive control (e.g., Christoffels et al., 2007; de Bruin et al., 2018). This suggests that participants used both reactive and proactive control, through inhibition of the non-target language (Green, 1998) and/or over-activation of the target language (Philipp et al., 2007). The switching costs were largest when switching to L2-English, whereas the mixing costs were largest when using L1-Mandarin.

Participants' production was faster in the L2 than in the L1 in all production contexts, including the single-language blocks. This suggests that the L2 – although not the dominant language - was the most active language during testing. The Mandarin-English bilinguals we tested might have generally over-activated their L2 because they were living in the UK (in the Move group) or because, in both groups, they were studying an English degree and tested in a bilingual environment (university setting) that required higher levels of L2 use. Other studies using similar populations (e.g., de Bruin & Xu, 2023) or with other L1-dominant bilinguals living in an L2-environment (e.g., Italian-English bilinguals in Bonfieni et al., 2019) have found similar results. In these types of settings and bilingual groups, bilinguals might over-activate their L2 to allow for fast access, especially in dual-language tasks where both languages have to be used flexibly.

The same participants also showed larger L2 *switching costs*, contrary to some previous studies reporting larger L1 costs in cued production tasks (Jin et al., 2014; Li & Gollan, 2021; Macizo et al., 2012; Meuter & Allport, 1999; Peeters et al., 2014; Philipp et al., 2007, but cf. Gade et al., 2021 for a review). However, larger L2 than L1 switching costs have previously been observed if the L2 is the participants' most proficient, used, or active language (cf. e.g., Bonfieni et al., 2019; de Bruin & Xu, 2023; Declerck, Stephan, et al., 2015; Zheng et al., 2020 for similar results). Similarly, larger L2 switching costs have been observed in task contexts where the L2 was manipulated to be the more active language (Timmer et al., 2019). When switching-cost asymmetries are observed, across the literature and as observed here, the arising pattern is that the largest costs are often observed for the most active language (in the task).

Both activation and inhibition are key mechanisms underlying bilingual language control (e.g., Green, 1998). Both overall faster L2 than L1 responses and asymmetries in switching costs are often interpreted through (over-)inhibition of the L1 (e.g., Goldrick & Gollan, 2023). However, the complex combination of faster L2 RTs, smaller L2 mixing costs, and larger L2 switching costs in the current study seems most easily explained through a combination of proactive L2 over-activation and a need to reactively inhibit the L2 when switching to the L1. These at least appear to be the mechanisms involved when participants are asked to produce easy-to-retrieve words (see the next paragraph for further discussion). Bilinguals might over-activate their L2, leading to faster L2 responses throughout and smaller L2 mixing costs if this over-activation is applied most strongly in the dual-language context. When having to switch back to the L1, this then might have required more *reactive* L2 inhibition. This in turn would have required relatively more time to lift this reactive inhibition at the moment of switching back to the L2 (larger L2 switching costs). An explanation purely involving inhibition seems less plausible as this would require the bilinguals to constantly inhibit the L1 (larger L1 mixing costs) *and* reactively over-inhibit the L2 (larger L2 switching costs), with the latter appearing redundant if the L1 is already inhibited strongly. Similarly, an account purely involving over-activation seems less likely as it would require a bilingual to constantly over-activate the L2 (smaller L2 mixing costs) *and* to reactively over-activate the L1 (larger L2 switching costs), which again would appear to counteract the (likely demanding) process of ensuring the L2 is active enough for use in a dual-language context.

Such combination of L2 proactive over-activation and reactive inhibition resulting in faster L2 than L1 responses and larger L2 switching costs might specifically apply to bilinguals who acquired one language from birth, have a higher proficiency in that L1, but are living in L2-dominant contexts or bilingual societies (see e.g., Bonfieni et al., 2019; de Bruin & Xu, 2023; Jevtović et al., 2020). These patterns have not been observed in studies using similar paradigms with unbalanced bilinguals living in an L1-dominant environment who rarely use their L2 (e.g., de Bruin et al., 2023). This suggests that specifically when immersed in a context where the non-dominant language is over-represented, such as when studying in a foreign country or in a setting associated with the L2 (like a university), bilinguals might over-activate the non-dominant language to facilitate its retrieval for production. Furthermore, faster L2 than L1 production might only be achievable when words can easily be retrieved in both languages. In our tasks, all pictures represented high-frequency, easy-to-name items that participants were familiarised with before the study. L2 over-activation might not result in faster L2 than L1 production (or be applied) when the words are more difficult to retrieve, especially in the L2 (e.g., when lower frequency items are used or when items are not repeated within the task). Indeed, in our completely separate single-language tasks (de Bruin et al., 2025) that used picture naming with lower-frequency unrepeated items (as well as verbal fluency tasks requiring free word generation), the same participants' production was faster in their L1 than L2.

4.1.2. Voluntary production task

The voluntary task was included to examine language-switching practices more similar to the dense code-switching context described by Green & Abutalebi (2013). Similar to previous studies, we found significant switching costs in the voluntary production task (e.g., de Bruin et al., 2018, 2020; Fricke et al., 2016; Gollan & Ferreira, 2009; Gross & Kaushanskaya, 2015; Jevtović et al., 2020; but see e.g., Blanco-Elorrieta & Pylkkänen, 2017; Zhu et al., 2022), suggesting there was still some language competition, and potential control, at the moment of switching. However, as discussed below in the task comparison section, these voluntary switching costs were smaller than the cued costs.

Contrary to the cued task, no mixing cost was observed in either group (cf. de Bruin et al., 2018; de Bruin & Xu, 2023; Gollan & Ferreira, 2009; Grunden et al., 2020). The patterns of mixing effects in this task align with the predictions of the Adaptive Control Hypothesis (Green & Abutalebi, 2013). Relative to the single-language blocks, where proactive control is also necessary to manage language competition, no more (or sometimes even less) proactive control might be needed when freely using two languages. Bilinguals can pick which language to use and rely on opportunistic language use (Green & Wei, 2014). Neither group showed a mixing cost, but the Control group also showed an L2 mixing benefit. For this Control group, L2 word retrieval might have been more difficult, as suggested by the longer RTs and less use of English. Being able to use L2-English opportunistically in the dual-language part might therefore have been more beneficial to the Control group (see e.g., Gollan & Ferreira, 2009). That is, these participants might have benefited more from only using their L2-English when words were fast to retrieve in that language, reserving their L1-Mandarin for more difficult words (cf. also de Bruin et al., 2018, showing a larger mixing benefit for highly proficient bilinguals who relied more strongly on voluntarily using the language in which they were fastest to retrieve a specific word).

Daily-life language switching can happen in different ways, including in contexts where people switch languages in response to cues (such as the face of a conversation partner) as well as in contexts where there are no external demands or cues indicating language choice or moment of switching. Our voluntary task was included to capture the latter type of context. Nevertheless, the switching tasks used here are still far from mimicking daily-life interactions, including in terms of the absence of a sentence context (cf. e.g., de Bruin & Shiron, 2024; Sánchez et al., 2022); the potential influence of a conversation partner (cf. e.g., de Bruin & Shiron, 2024; Kootstra et al., 2020); as well as other linguistic and non-linguistic cues present in one's daily-life language environment (cf. e.g., de Bruin & Martin, 2022; Vaughn-Evans, 2023). At the same time, lab-based tasks may increase participants' awareness of their switching behaviour and might increase strategic choices (e.g., using both languages regularly, even without instructions asking participants to do so). Indeed, participants switch more frequently in these tasks than in real-life conversations (e.g., Fricke & Kootstra, 2016), even if switching frequencies between these lab-based tasks and real-life conversations correlate (de Bruin, in press). These different factors can increase the need for top-down control and reduce the role of opportunistic language use, thereby potentially inflating voluntary switching costs (although such costs have been observed in corpus data too, Fricke et al., 2016).

4.1.3. Comprehension task

The comprehension task showed a significant cost when switching to L1-Mandarin, but not to L2-English. While the significant comprehension switching costs stand in contradiction with previous research not finding such costs in comprehension tasks (see e.g., Declerck et al., 2019), it does align with previous studies finding significant comprehension switching costs in one language only (e.g., Jackson et al., 2004; Mosca & de Bot, 2017), or a larger cost in one language than in the other (e.g., Declerck & Grainger, 2017; Litcofsky & Van Hell, 2017; Olson, 2017; Philipp & Huestegge, 2015). In some previous studies, contrary to our findings, bilinguals showed a larger L2 switching cost in comprehension (Aparicio & Lavaur, 2014; Bultena et al., 2015; Grainger & Beauvillain, 1987; Liu et al., 2020; Proverbio et al., 2004; Struck & Jiang, 2022). However, our findings do align with other studies showing larger L1 than L2 switching costs during comprehension (e.g., Litcofsky & Van Hell, 2017). Furthermore, further studies using stimuli produced by speakers with L1 rather than L2 accents (Coumel et al., in 2024) replicated these findings, suggesting the observed L1 switching cost was not due to the specific type of speakers used in the current study.

These patterns are typically explained in a similar fashion as asymmetrical switching costs during production, using an inhibition-based account (Green, 1998). Specifically, during the task, the bilinguals might have applied more inhibition towards their dominant L1 (which, in the comprehension task, was indeed the faster language). This, in turn, may have led to longer times to release inhibition when switching to the L1 than to the L2, and thus resulted in significant L1 switching costs. Alternatively, this pattern could be explained through over-activation of the L2, resulting in more L2 interference (and larger costs) when switching back to the L1. The asymmetry pattern, interestingly, differed between production and comprehension studies, as discussed further below.

In terms of the mixing effect, the L1-Mandarin showed no mixing effect while the L2-English showed a mixing benefit. Previous comprehension studies have often focused on switching effects without assessing mixing effects. Some studies have observed mixing costs in some experiments (see Grainger & Beauvillain, 1987), but the current study showed no mixing cost in either language. This suggests that responding to two languages interchangeably was not more demanding than responding to words in one language. Within comprehension (as opposed to production), there might have been less need for conflict monitoring and cue detection, as responses (button presses indicating animacy) could be made regardless of the language the word was in. Relatedly, proactively balancing the activation and/or inhibition level of each language might be less relevant in these types of comprehension tasks.

Interestingly, a language difference was found, with no mixing effect for the L1 and a mixing benefit in the L2, similar to Jylkkä et al. (2017)'s findings. This is despite the current study using spoken stimuli while Jylkkä et al. used written words. Jylkkä and colleagues interpreted their L2 mixing benefit through L2 animacy responses benefiting from going through L1 representations in the dual-language context. This could allow bilinguals to use a faster route to reach a semantic judgement, namely by reaching the semantic concept through the corresponding L1 representation rather than through the L2 representation itself. This could benefit the L2 in particular as semantic representations are argued to be weaker for the L2 than the L1 (e.g., Kroll & Stewart, 1994). This benefit might only be possible in the dual-language but not in the single-language blocks, if the other language is suppressed in the single-language block. However, such explanation does assume bilinguals do *not* apply language control at all in the comprehension dual-language context, which is at odds with the asymmetry in language switching costs observed in the current study.

Another potential explanation, which does not relate to the role of language control, is that this mixing benefit might relate to task practice effects. While producing object names is a relatively natural task, pressing a button to indicate a word's animacy is less natural and the first few times a participant makes the judgement, RTs can be longer, perhaps even more so in the L2. If this underlies the L2 mixing benefit, we would expect the L2-English in particular to benefit more from practicing this task than the L1. Indeed, comparing the first (completed before the dual-language part) and final single-language blocks (completed after the dual-language part) showed a larger decrease in RTs for L2-English (*M* first block = 1209ms versus *M* final block= 1119ms; *M* change = -90ms) than L1-Mandarin (*M* first block = 1049ms versus *M* final block= 1016ms; *M* change = -33ms). In contrast, for the two production tasks, the L1 and L2 showed more similar RT decreases between the first and last single-language blocks (Cued production task: *M* L1 change = -115ms, *M* L2 change = -106ms; Voluntary production task: *M* L1 change = -91ms, *M* L2 change = -104ms). Thus, it is possible that the L2 mixing benefit in the comprehension task reflects the L2 being particularly slow in the very first single-language block, with the dual-language part and final single-language block benefiting from more practice.

4.1.4. Switching and mixing effects across tasks

Overall, the cued production task was the one where, based on the Adaptive Control Hypothesis (Green & Abutalebi, 2013), we expected the highest levels of proactive and reactive control and therefore the highest switching and mixing costs. The results align with these predictions. Switching costs were significantly larger in the cued production task than in the voluntary production task and in the production than comprehension tasks (see e.g., de Bruin & Xu, 2023; Gollan et al., 2014; Jevtović et al., 2020). Mixing costs were only found in the cued task and not in the other tasks. These findings support the hypothesis that voluntary switching contexts, relative to cued switching contexts, might enable opportunistic language planning, have lower attentional and control demands as there is less need for cue monitoring, and require less reactive and proactive language control. Moreover, it supports the hypothesis that comprehension tasks trigger less language competition under certain conditions (see e.g., Declerck et al., 2019), and therefore recruit less reactive and proactive control too.

We also aimed to better understand whether any production-comprehension differences were related to general process differences or specifically related to comparisons between *cued* (high-control) production and comprehension. Given that both voluntary and cued production showed significantly higher costs than the comprehension task (although this cost was smaller in voluntary than cued production), language competition at the moment of switching, and the reactive control used to manage it, might generally be higher during production than during comprehension. This is possibly related to the competition between the many word forms available for selection during production, without the strong exogenous bottom-up language activation that the word forms presented during comprehension provide. However, mixing effects only differed between the cued production and comprehension task and not between voluntary production and comprehension. This suggests *proactive* control is especially involved during *cued* production, while neither voluntary production nor comprehension appears to require high levels of proactive control or goal monitoring to select a specific language in response to the environmental demands.

Production tasks showed faster L2 than L1 responses across all contexts while the comprehension task showed faster L1 responses. Thus, if bilinguals over-activated their L2, this was specific to production. During comprehension, over-activating (or proactively over-inhibiting) one of the languages might be less beneficial if language competition is lower and when the words are provided to you by the other speaker, as opposed to having to retrieve word forms yourself for production. Intriguingly, in all tasks, it was the fastest language that showed the largest switching costs (L2 during production and L1 during comprehension). This suggests that across production and comprehension tasks, bilinguals applied most reactive control over the fastest language, potentially because that language was most active and required more control to minimise interference at the moment of switching. Thus, *why* reactive control is used might be similar between production and comprehension but *when* and *over which* language that control is used differs. This suggests a complex interplay between partly shared and partly distinctly used control mechanisms (e.g., Ahn et al., 2020; Blanco-Elorrieta & Pylkkännen, 2016; Macizo et al., 2012; Mosca & de Bot, 2017; Reynolds et al., 2016; but see Gambi & Harstuiker, 2016; Peeters et al., 2014).

Across the literature, as well as in our studies, production and comprehension paradigms typically vary in several ways, including the additional task demands introduced in our comprehension paradigm (through the animacy judgement paradigm). We opted for this task given the demands it placed on semantic and lexical processing (beyond what could be achieved through e.g., a lexical decision task) and given its previous use in the literature (e.g., Jylkkä et al., 2017). However, future research might opt for paradigms more similar to production tasks (e.g., a picture-name verification tasks). In the current study, however, RT differences between tasks were small and mixing and switching cost task differences survived when RT differences were taken into account through analysis of proportional effects. Furthermore, our previous research using this task (Coumel et al., 2024) showed across multiple experiments that language switching costs in the comprehension task were not modulated by task demands such as category response switching.

4.2. The influence of language environment and individual differences (RQ2)

The results described above indicate that participants' language-control mechanisms adapt to the immediate context, in this case, the particular task or immediate context bilinguals are immersed in. However, across the tasks, language control did not appear to change between sessions in response to a change in overall language environment. None of the Bayesian analyses provided support for switching or mixing effect changes within the Move group.

Previous literature assessing the potential influence of daily-life language environment on language control during switching has either compared bilinguals differing in their language background (e.g., Bonfieni et al., 2019), manipulated the language context *within the experimental task* (e.g., Timmer et al., 2019), or provided participants with short-term (days or weeks) language-switching training (e.g., Zhang et al., 2021). However, those studies were not able to address the key question posed in this study: whether changes in daily-life language environment result in long-term language control changes. Our findings suggest that within bilinguals who are beyond the initial stages of language learning, these language-control mechanisms might not necessarily change in response to changes in the daily-life language environment. This is despite the same participants showing flexible control in response to the immediate task environment (RQ1). Together, these data show that language control is flexible but that such flexibility might be applied in response to the immediate context rather than the overall environment.

These findings align with the single-language tasks completed by these participants in the same study (reported in de Bruin et al., 2025; different from the single-language blocks included here). In those tasks, we examined proactive control during single-language production. Those tasks furthermore included both high- and low-frequency items as previous studies have suggested word frequency can influence language changes (Baus et al., 2013; Casado et al., 2023). Similar to the current study, in those single-language tasks with low- and high-frequency words, no changes in language control were observed either after the Move group changed their language environment. Overall, there thus is very little evidence that language control changed after a move to a new language environment, even though participants did indeed increase their L2 proficiency, L2 use, time in dual-language contexts, and time switching languages in their daily lives after the move (see Table 4). The absence of corresponding changes in language control could result from the fact that, in their everyday lives, bilinguals evolve through various interactional contexts. Each day, participants in the Move group likely switched from environments where they heard L2-English only (e.g., during lectures), spoke L2-English only (e.g., at the supermarket), to environments where they interacted with Mandarin-English bilingual friends. Similarly, participants in the Control group might have alternated between environments including L2 use (e.g., at university) and L1-Mandarin-dominated environments (e.g., at home). In other words, such varied language experiences in their everyday life could require bilinguals to have a language control system that is flexible and can quickly adapt to the current interactional context. As part of this adaptive language-control system, bilinguals might constantly monitor the circumstances to adjust their language control accordingly. As shown in the previous sections, this could occur through adjustments in L1 and L2 proactive activation and reactive inhibition in response to the task demands of the immediate context. If bilinguals constantly need to adjust their language control in response to the immediate contexts varying in their control demands, no overall adjustments to a *general* language environment might be possible (if there is not one constant environment bilinguals can adapt to). This might especially apply to the type of bilinguals tested here, who (as shown in Tables 3 and 4) indeed spend substantial amounts of time in different language contexts, including single-language, high-control dual-language, and dense code switching environments.

In a few specific cases, between-session changes were observed in the current study. Cued accuracy switching costs were smaller in Session B than in Session A in the Move group. This could suggest that the Move group, after living in an L2/dual-language environment, found it easier to switch languages. The finding that this only affected accuracy and not RTs could suggest these participants were better able to retrieve the picture names in a switching context, without it necessarily affecting the speed of retrieval and their language control. Within the cued RT analysis, the mixing cost showed some differences between Sessions, but this appeared largely driven by the Control group rather than the Move group (with moderate evidence from the Bayesian analyses supporting no change in switching or mixing cost RTs in the Move group).

The largest between-session difference observed was in terms of voluntary switching costs, which were smaller in Session B than A. However, this did not differ between the groups, suggesting the decrease was not related to the language environment change. Furthermore, the Bayesian analysis provided no evidence for a change in switching costs between sessions in the Move group. Instead, it appeared related to *how* participants used their languages within the task, with participants who increased their switching frequency showing a larger decrease in switching costs. Switching frequency in the task was related to daily life dense code-switching frequency. Participants who reported higher daily-life switching (see Supplementary Materials) also switched more in the task, and those who showed a larger increase in daily-life dense code-switching were also the ones who showed a larger increase in their voluntary-task switching frequency between sessions. The individual-difference analysis furthermore showed the decrease in switching costs was largest for participants who spent very little time in daily-life dense code-switching environments. This suggests that the decrease in voluntary switching costs could be driven by participants with low exposure to daily-life code switching benefiting more from increasing their initially low switching frequency.

Finally, we focused on language control but also assessed potential changes in lexical access more generally. Previous longitudinal studies on lexical access in single-language contexts have suggested language learners' L1 and L2 production can change after moving to an L2-dominant environment (e.g., Baus et al., 2013; Linck et al., 2009). In terms of changes between sessions, we had predicted that the Move group in particular would show facilitated L2 lexical access, and potentially reduced L1 access, at Session B, as a result of living in an L2-dominated environment. However, no between-session changes in lexical access were observed in the majority of tasks. That is, overall accuracy, switching frequency, and RTs stayed comparable across sessions, with a few exceptions. In the voluntary task, the Move group (but not the Control group) showed an increase in language RT differences between sessions. This was a combination of slightly faster L2 and slightly slower L1 responses in Session B. In the comprehension task, RTs were also faster in Session B, but this did not differ between Groups, suggesting a practice effect related to the animacy judgement task. The vast majority of analyses, however, showed no changes in lexical access between sessions. This could result from the fact that we tested (already) intermediate proficiency-level L2-English speakers, who were not early language learners. Longitudinal changes in lexical access might be more likely to surface in less proficient language learners (e.g., Baus et al., 2013; Linck et al., 2009).

4.3. Limitations

The study was conducted during the COVID-19 pandemic, which had substantial impacts on our study. For instance, participants of the first testing wave spent some time in quarantine when they arrived in the UK. Because of this quarantine, Session A was completed two weeks after arrival in the UK. This could have masked changes between Sessions A and B if language control changes happen immediately after arrival in a new country. However, if such immediate changes had happened (which is unlikely in quarantine, when interaction with the outside world is restricted), the Move group should have shown smaller mixing and switching costs than the Control group in Session A, which was not the case. The COVID-19 pandemic also potentially led the participants to experience fewer changes in dual-language experiences (relative to before they moved to the UK) than would have been the case in normal circumstances. That is, people potentially had fewer opportunities to meet new people and to interact in person. Although the changes in daily-life language experiences were significant and as expected, larger changes (more L2 use) could perhaps result in (larger) changes in language control. However, we recruited across two academic years and analyses comparing the groups who arrived with and without pandemic-related restrictions in place showed no differences between those groups in terms of daily-life language experiences reported in Session B (*p*s > .21) or language-experience differences relative to baseline (*p*s >.14). This suggests the restrictions did not impact the participants' immersion. Nevertheless, the pandemic, even in the second testing wave, could have impacted the social interactions of students, including in terms of engagement in online activities versus in-person activities in the UK. This in turn could also influence the amount of time spent using the L1 (e.g., in online activities) compared to the L2 (which would likely be used more during in-person activities in the UK).

We also acknowledge that, due to the COVID-19 pandemic, the sample size we reached was smaller than preregistered and power was lower than planned. There were also some small baseline differences between the Move and Control groups, in terms of overall L1 and L2 RTs. Such baseline overall RT differences are important to (attempt to) avoid, but given the absence of Session effects within the Move group itself, (a difference with) the control group is less relevant. Furthermore, the Bayesian analyses provided moderate evidence supporting null results in the Move group itself, without the Control group influencing the analyses, suggesting there was truly no change in language control.

Finally, it is important to acknowledge that we only tested one type of bilinguals. Testing different types of bilinguals could lead to different results, in particular regarding changes after a move in language environment. The bilinguals tested in our study still frequently used their L1-Mandarin while studying in the UK. It is possible that bilinguals who almost exclusively use their L2 after moving to an L2-environment undergo larger changes in their language control and/or lexical access. However, the bilinguals we tested did spend a lot of time in dual-language contexts, which included language switching. As such, any changes in language switching in particular should be more likely to occur in the bilinguals tested here than in bilinguals who predominantly use their L2 only and might not switch very often. Nevertheless, different types of bilinguals should be tested longitudinally in future research to understand the generalisability of these findings. Furthermore, such research should also consider the role of language distance. In the current sample, bilinguals' L1 and L2 differed vastly, for example, in terms of vocabulary, grammar, orthography, and phonology. Bilinguals speaking more similar languages can potentially experience more interference and might therefore develop their language control differently.

5. Conclusion

Daily life poses different linguistic and language control demands on bilinguals, be it because they interact with people with various linguistic backgrounds or because they move between countries.

Therefore, this study tested how and when bilinguals' language control might adapt to the immediate context they are in, to changes in overall language environments, and to variation in everyday language experiences. To do this, we conducted one of the first longitudinal studies that tested the same bilinguals on three different language control tasks reflecting a range of possible contexts. Our results indicated that bilinguals' language-control mechanisms can quickly adapt to the demands of the immediate context, such as whether they were producing or listening to language switches, or whether they were instructed which language to use or could use languages freely. Our findings, however, also suggest that language-control mechanisms might be less likely to change as a function of changes in language environments.

Acknowledgements: We would like to thank Tianyi Xu, Marco Kwok, Jiawen Wu, Xuanlin Fan, Jerome Pek, Danqing Liu, Qinglu Chen, Olivia Bai, Yifei Zhang, Eve Ong, Yingying Yu, Jiting Shi, and Guanyu Zhou for their help collecting and scoring data.

Author Note: The stimuli, complete dataset and analyses script are available at https://osf.io/a24xv/

Declarations of interest: none

CRediT author statement: Marion Coumel: Conceptualization, Methodology, Investigation, Data Curation, Validation, Formal Analysis, Writing- Original draft, Visualization, Project administration; **Cong Liu**: Funding Acquisition, Investigation, Writing- Review & Editing; **Danijela Trenkic:** Funding Acquisition, Conceptualization, Writing- Review & Editing; **Angela de Bruin**: Conceptualization, Methodology, Investigation, Data Curation, Validation, Formal Analysis, Writing- Original draft, Visualization, Project administration, Funding Acquisition.

Funding: This work was supported by the Economic and Social Research Council (ESRC, grant number ES/V004220/1) and Young Innovative Research Team in Shandong Higher Education Institution (2022RW030).

References

Adamou, E., & Shen, X.R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, *23*(1), 53–70. https://doi.org/10.1177/1367006917709094

Ahn, D., Abbott, M. J., Rayner, K., Ferreira, V. S., & Gollan, T. H. (2020). Minimal overlap in language control across production and comprehension: Evidence from read-aloud versus eye-tracking tasks. *Journal of Neurolinguistics*, *54*, 100885. https://doi.org/10.1016/j.jneuroling.2019.100885

Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*, 250-263. https://doi.org/10.3758/s13428-017-0867-9

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Aparicio, X., & Lavaur, J.-M. (2014). Recognising words in three languages: Effects of language dominance and language switching. *International Journal of Multilingualism*, *11*, 164–181. https://doi.org/10.1080/14790718.2013.783583

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68* (3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

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

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica*, *142*(3), 402–409. https://doi.org/10.1016/j.actpsy.2013.01.010

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(6), 1022–1047. https://doi.org/10.1037/xlm0000770

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual language control in perception versus action: MEG reveals comprehension control mechanisms in anterior cingulate cortex and domain-general control of production in dorsolateral prefrontal cortex. *The Journal of Neuroscience*, *36*(2), 290–301. https://doi.org/10.1523/JNEUROSCI.2597-15.2016

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual Language Switching in the

Laboratory versus in the Wild: The Spatiotemporal Dynamics of Adaptive Language Control. *The Journal of Neuroscience*, *37*(37), 9022–9036. https://doi.org/10.1523/JNEUROSCI.0553-17.2017

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica*, *193*, 160–170. https://doi.org/10.1016/j.actpsy.2018.11.004

Botezatu, M. R., Kroll, J. F., Trachsel, M. I., & Guo, T. (2022). Second language immersion impacts native language lexical production and comprehension. *Linguistic Approaches to Bilingualism*, *12*(3), 347-376. https://doi.org/10.1075/lab.19059.bot

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1). doi: 10.5334/joc.10

Bultena, S., Dijkstra, T., & van Hell, J. G. (2015). Switching cost modulations in bilingual sentence processing: Evidence from shadowing. *Language, Cognition and Neuroscience*, *30*, 586–605. https://doi.org/10.1080/23273798.2014.964268

Casado, A., Walther, J., Wolna, A., Szewczyk, J., Sorace, A., & Wodniecka, Z. (2023). Advantages of visiting your home country: how brief reimmersion in their native country impacts migrants' native language access. *Bilingualism: Language and Cognition*, *26*(5), 1026-1037. https://doi.org/10.1017/S136672892300024X

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: an event-related brain potential study. *Brain Research*, *1147*, 192–208. https://doi.org/10.1016/j.brainres.2007.01.137

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52-64. https://doi.org/10.1016/j.intell.2014.01.004

Coumel, M., Liu, C., Trenkic, D., & de Bruin, A. (2024). Do accent and input modality modulate processing of language switches in bilingual language comprehension?. *Journal of Experimental Psychology: Human Perception and Performance*, *50*(4), 395–415. https://doi.org/10.1037/xhp0001190

Coumel, M., Liu, C., Trenkic, D., & de Bruin, A. (2025). Moving from China to York : How do changes in language environment modulate bilingual language control?. Data set. https://doi.org/10.17605/OSF.IO/A24XV

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages

Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*, 491–511. https://doi.org/10.1016/j.jml.2004.02.002

de Bruin, A. (in press). Examining the reliability and validity of bilingual language use and switching measures. *Bilingualism: Language and Cognition.*

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00522

de Bruin, A., Kressel, H., & Hemmings, D. (2023). A comparison of language control while switching within versus between languages in younger and older adults. *Scientific Reports*, *13*(1), 16740. https://doi.org/10.1038/s41598-023-43886-1

de Bruin, A., Liu, C., Trenkic, D., Coumel, M. (2025). Bilingual language control during single-language production: Does relocation to a new linguistic environment change it? *Royal Society Open Science*, *12*(2), 241071. https://doi.org/10.1098/rsos.241071

de Bruin, A., & Martin, C. D. (2022). Perro or txakur? Bilingual language choice during production is influenced by personal preferences and external primes. *Cognition*, *222*, 104995. https://doi.org/10.1016/j.cognition.2021.104995

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28–43. https://doi.org/10.1016/j.jml.2018.07.005

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2020). Examining bilingual language switching across the lifespan in cued and voluntary switching contexts. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(8), 759–788. https://doi.org/10.1037/xhp0000746

de Bruin, A., & Shiron, V. (2024). Putting language switching in context: Effects of sentence context and interlocutors on bilingual switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(7), 1112–1132. https://doi.org/10.1037/xlm0001309

de Bruin, A., & Xu, T. (2023). Language switching in different contexts and modalities: Response-stimulus interval influences cued-naming but not voluntary-naming or comprehension language-switching costs. *Bilingualism: Language and Cognition*, *26*, 402-415. http://doi.org/10.1017/S1366728922000554

Declerck, M., & Grainger, J. (2017). Inducing asymmetrical switching costs in bilingual language comprehension by language practice. *Acta Psychologica*, *178*, 100–106. https://doi.org/10.1016/j.actpsy.2017.06.002

Declerck, M., Koch, I., Duñabeitia, J. A., Grainger, J., & Stephan, D. N. (2019). What absent switching costs and mixing costs during bilingual language comprehension can tell us about language control. *Journal of Experimental Psychology. Human Perception and Performance*, *45*, 771–789. https://doi.org/10.1037/xhp0000627

Declerck, M., Koch, I., & Philipp, A. M. (2015). The minimum requirements of language control: Evidence from sequential predictability effects in language switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 377–394. https://doi.org/10.1037/xlm0000021

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, *22*, 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

Declerck, M., Stephan, D. N., Koch, I., & Philipp, A. M. (2015). The other modality: Auditory stimuli in language switching. *Journal of Cognitive Psychology*, *27*, 685–691. https://doi.org/10.1080/20445911.2015.1026265

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175-197. https://doi.org/10.1017/S1366728902003012

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*, 808–816. https://doi.org/10.1080/17470218.2017.1310261

Duñabeitia, J. A. (2022). MultiPic: Multilingual Picture Database. Figshare. Dataset. https://doi.org/10.6084/m9.figshare.19328939.v8

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, *91*, 181–201. https://doi.org/10.1016/j.jml.2016.04.003

Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, *89*, 110-137. https://doi.org/10.1016/j.jml.2015.10.001

Gade, M., Declerck, M., Philipp, A. M., Rey-Mermet, A., & Koch, I. (2021). Assessing the evidence for asymmetrical switching costs and reversed language dominance effects—A meta-analysis. *Journal of Cognition*, *4*, 55. https://doi.org/10.5334/joc.186

Gambi, C., & Hartsuiker, R. J. (2016). If you stay, it might be easier: Switch costs from comprehension to production in a joint switching task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(4), 608–626. https://doi.org/10.1037/xlm0000190

García, P. B., Leibold, L., Buss, E., Calandruccio, L., & Rodriguez, B. (2018). Code-switching in highly proficient Spanish/English bilingual adults: Impact on masked word recognition. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2353-2363. https://doi.org/10.1044/2018_JSLHR-H-17-0399

Gavino, M. F., & Goldrick, M. (2024). The perception of code-switched speech in noise. *JASA Express Letters*, *4*(3). https://doi.org/10.1121/10.0025375

Goldrick, M., & Gollan, T. H. (2023). Inhibitory control of the dominant language: Reversed language dominance is the tip of the iceberg. *Journal of Memory and Language*, *130*, 104410. https://doi.org/10.1016/j.jml.2023.104410

Gollan, T.H., & Ferreira, V.S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*, 640–665. https://doi.org/10.1037/a0014981

Gollan, T.H., Kleinman, D., & Wierenga, C.E. (2014). What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General*, *143*, 2167– 2195. https://doi.org/10.1037/a0038006

Gosselin, L., & Sabourin, L. (2021). Lexical-semantic processing costs are not inherent to intra-sentential code-switching: The role of switching habits. *Neuropsychologia*, *159*, 107922. https://doi.org/10.1016/j.neuropsychologia.2021.107922

Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology Section A*, *39*, 295–319. https://doi.org/10.1080/14640748708401788

Grange, J.A. (2015). trimr: An implementation of common response time trimming methods. R package version 1.0.1. https://cran.r-project.org/web/packages/trimr/index.html

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81. Cambridge Core. https://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530. https://doi.org/10.1080/20445911.2013.796377

Green, W. & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience*, *29*(4), 499-511, https://doi.org/10.1080/23273798.2014.882515

Gross, M., & Kaushanskaya, M. (2015). Voluntary language switching in English–Spanish bilingual children. *Journal of Cognitive Psychology*, *27*(8), 992-1013. https://doi.org/10.1080/20445911.2015.1074242

Grunden, N., Piazza, G., García-Sánchez, C., & Calabria, M. (2020). Voluntary language switching in the context of bilingual aphasia. *Behavioral Sciences*, *10*(9), 141. https://doi.org/10.3390/bs10090141

Gullifer, J. W., & Titone, D. (2019). The impact of a momentary language switch on bilingual reading: Intense at the switch but merciful downstream for L2 but not L1 readers. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*, 2036–2050. https://doi.org/10.1037/xlm0000695

Gullifer, J., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*(2), 283-294. http://doi.org/10.1017/S1366728919000026

Han, X., Li, W., & Filippi, R. (2022). The effects of habitual code-switching in bilingual language production on cognitive control. *Bilingualism: Language and Cognition*, *25*(5), 869-889. https://doi.org/10.1017/S1366728922000244

Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition*, *150*, 10–19. https://doi. org/10.1016/j.cognition.2016.01.016

Hartanto, A., & Yang, H. (2020). The role of bilingual interactional contexts in predicting interindividual variability in executive functions: A latent variable analysis. *Journal of Experimental Psychology: General*, *149*(4), 609-633. https://doi.org/10.1037/xge0000672

Heikoop, K. W., Declerck, M., Los, S. A., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language switching. *Bilingualism: Language and Cognition*, *19*(5), 921–927. https://doi.org/10.1017/S1366728916000456

Hui, N.-Y., Fong, M. C.-M., & Wang, W. S. (2022). Bilingual prefabs: No switching cost was found in Cantonese–English habitual code-switching in Hong Kong. *Languages*, *7*(3), 198.http://dx.doi.org/10.3390/languages7030198

The International Cognitive Ability Resource Team (2014). https://icar-project.com/.

Jackson, G. M., Swainson, R., Mullin, A., Cunnington, R., & Jackson, S. R. (2004). ERP correlates of a receptive language-switching task. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *57*, 223–240. https://doi.org/10.1080/02724980343000198

Jevtović, M., Duñabeitia, J. A., & de Bruin, A. (2020). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, *23*, 401–413. Cambridge Core. https://doi.org/10.1017/S1366728919000191

Jin, Z.-L., Zhang, J.-X., & Li, L. (2014). Endogenous language control in Chinese-English switching: An event-related potentials study. *Neuroscience Bulletin*, *30*, 461–468. https://doi.org/10.1007/s12264-013-1427-7

Jylkkä, J., Lehtonen, M., Kuusakoski, A., Lindholm, F., Hut, S. C. A., & Laine, M. (2017). The role of general executive functions in receptive language switching and monitoring. *Bilingualism: Language and Cognition*, *21*, 839–855. Cambridge Core. https://doi.org/10.1017/S1366728917000384

Kaan, E., Kheder, S., Kreidler, A., Tomić, A., & Valdés Kroff, J. R. (2020). Processing code-switches in the presence of others: An ERP study. *Frontiers in Psychology*, *11*, 1288. https://doi.org/10.3389/fpsyg.2020.01288

Kałamała, P., Szewczyk, J., Chuderski, A., Senderecka, M., & Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition*, *204*, 104373. https://doi.org/ 10.1016/j.cognition.2020.104373

Kang, C., Fu, Y., Wu, J., Ma, F., Lu, C., & Guo, T. (2017). Short-term language switching training tunes the neural correlates of cognitive control in bilingual language production. *Human Brain Mapping*, *38*(12), 5859–5870. https://doi.org/10.1002/hbm.23765

Kang, C., Ma, F., & Guo, T. (2018). The plasticity of lexical selection mechanism in word production: ERP evidence from short-term language switching training in unbalanced Chinese–English bilinguals. *Bilingualism: Language and Cognition*, *21*(2), 296-313. https://doi.org/10.1017/S1366728917000037

Kleinman, D., & Gollan, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science*, *27*(5), 700-714. https://doi.org/10.1177/0956797616634633

Kootstra, G. J., Dijkstra, T., & Van Hell, J. G. (2020). Interactive alignment and lexical triggering of code-switching in bilingual dialogue. *Frontiers in Psychology*, *11*, 1747. https://doi.org/10.3389/fpsyg.2020.01747

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lehtonen, M., Fyndanis, V. & Jylkkä, J. (2023). The relationship between bilingual language use and executive functions. *Nature Reviews Psychology*, *2*, 360–373. https://doi.org/10.1038/s44159-023-00178-9

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Li, C., Midgley, K. J., Ferreira, V. S., Holcomb, P. J., & Gollan, T. H. (2024). Different language control mechanisms in comprehension and production: Evidence from paragraph reading. *Brain and Language*, *248*, 105367. https://doi.org/10.1016/j.bandl.2023.105367

Li, C., & Gollan, T. H. (2022). Language-switch costs from comprehension to production might just be task-switch costs. *Bilingualism: Language and Cognition*, *25*(3), 459-470. https://doi.org/10.1017/S1366728921001061

Liao, C. H., & Chan, S. H. (2016). Direction matters: Event-related brain potentials reflect extra processing costs in switching from the dominant to the less dominant language. *Journal of Neurolinguistics*, *40*, 79-97. https://doi.org/10.1016/j.jneuroling.2016.06.004

Linck, J.A., Kroll, J.F., & Sunderman, G. (2009) Losing access to the native language while immersed in a second language evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x

Litcofsky, K. A., & Van Hell, J. G. (2017). Switching direction affects switching costs: Behavioral, ERP and time-frequency analyses of intra-sentential codeswitching. *Neuropsychologia*, *97*, 112–139. https://doi.org/10.1016/j.neuropsychologia.2017.02.002

Liu, C., de Bruin, A., Jiao, L., Li, Z. & Wang , R. (2021) Second language learning tunes the language control network: a longitudinal fMRI study, *Language*, *Cognition and Neuroscience*, *36* (4), 462-473. https://doi.org/10.1080/23273798.2020.1856898

Liu, H., Li, W., de Bruin, A., & He, Y. (2021). Should I focus on self-language actions or should I follow others? Cross-language interference effects in voluntary and cued language switching. *Acta Psychologica*, *216*, 103308. https://doi.org/10.1016/j.actpsy.2021.103308

Liu, C., Timmer, K., Jiao, L., & Wang, R. (2020). Symmetries of comprehension-based language switching costs in conflicting versus non-conflicting contexts. *International Journal of Bilingualism*, *24*, 588–598. https://doi.org/10.1177/1367006919848487

Ma, D.S., Kantner, J. & Wittenbrink, B. (2021). Chicago Face Database: Multiracial expansion. *Behavioral Research*, *53*, 1289–1300. https://doi.org/10.3758/s13428-020-01482-5

Macizo, P., Bajo, T., & Paolieri, D. (2012). Language switching and language competition. *Second Language Research*, *28*, 131–149. https://doi.org/10.1177/0267658311434893

Mann, A., & de Bruin, A. (2022). Bilingual language use is context dependent: using the Language and Social Background Questionnaire to assess language experiences and test-rest reliability. *International Journal of Bilingual Education and Bilingualism*, *25*(8), 2886-2901. https://doi.org/10.1080/13670050.2021.1988049

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*, 25–40. https://doi.org/10.1006/jmla.1998.2602

Mosca, M., & de Bot, K. (2017). Bilingual language switching: Production vs. recognition. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00934

Olson, D. J. (2016). The gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics*, *37*(3), 725-756. https://doi.org/10.1017/S0142716415000223

Olson, D. J. (2017). Bilingual language switching costs in auditory comprehension. *Language, Cognition and Neuroscience*, *32*, 494–513. https://doi.org/10.1080/23273798.2016.1250927

Peeters, D., Runnqvist, E., Bertrand, D., & Grainger, J. (2014). Asymmetrical switching costs in bilingual language production induced by reading words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*, 284–292. https://doi.org/10.1037/a0034060

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavioral Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology*, *19*, 395–416. https://doi.org/10.1080/09541440600758812

Philipp A. M., & Huestegge, L. (2015). Language switching between sentences in reading: Exogenous and endogenous effects on eye movements and comprehension. *Bilingualism: Language and Cognition*, *18*, 614–625. https://doi.org/10.1017/S1366728914000753

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: evidence from

Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society: JINS*, *17*(4), 682–691. https://doi.org/10.1017/S1355617711000580

Project Atlast (2022). *Global Mobility Trends* https://www.iie.org/wp-content/uploads/2023/03/Project-Atlas_Infographic_2022.pdf

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859-862. https://doi.org/10.3758/BF03192979

Proverbio, A. M., Leoni, G., & Zani, A. (2004). Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia*, *42*, 1636–1656. https://doi.org/10.1016/j.neuropsychologia.2004.04.013

Reynolds, M. G., Schlöffel, S., & Peressotti, F. (2016). Asymmetric switch costs in numeral naming and number word reading: Implications for models of bilingual language production. *Frontiers in Psychology*, *6*, 2011. https://doi.org/10.3389/fpsyg.2015.02011

Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, *2*, 388. https://doi.org/10.3389/fpsyg.2011.00388

Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1477–1491. https://doi.org/10.1037/ 0278-7393.31.6.1477

Sánchez, L. M., Struys, E., & Declerck, M. (2022). Ecological validity and bilingual language control: voluntary language switching between sentences. *Language, Cognition and Neuroscience*, *37*(5), 615-623. https://doi.org/10.1080/23273798.2021.2016873

Struck, J., & Jiang, N. (2022). Language switching costs in a lexical decision task: Symmetry and cognitive correlates. *Second Language Research*, *38*, 813–838. https://doi.org/10.1177/0267658321998740

Timmer, K., Christoffels, I.K., & Costa, A. (2019). On the flexibility of bilingual language control: the effect of language context. *Bilingualism: Language and Cognition*, *22*(3), 555–568. https://doi.org/10.1017/S1366728918000329

Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Dussias, P. E. (2018). Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguistic Approaches to Bilingualism*, *8*, 98-133. https://doi.org/10.1075/lab.16011.val

Vaughan-Evans, A. (2023). External non-linguistic cues influence language selection during a forced choice task. *Bilingualism: Language and Cognition*, *26*(1), 193-201. http://doi.org/10.1017/S136672892200044X

Yim, O., & Bialystok, E. (2012). Degree of conversational code-switching enhances verbal task switching in Cantonese–English bilinguals. *Bilingualism: Language and Cognition*, *15*, 873-883. https://doi.org/10.1017/S1366728912000478

Wen, Y., Qiu, Y., Leong, C., Xiang Ru, & van Heuven, W. (2023). LexCHI: A quick lexical test for estimating language proficiency in Chinese. *Behavioral Research*, *56*(3), 2333-2352. https://doi.org/10.3758/s13428-023-02151-z

Wu, J., Kang, C., Ma, F., Gao, X., & Guo, T. (2018). The influence of short-term language- switching training on the plasticity of the cognitive control mechanism in bilingual word production. *Quarterly Journal of Experimental Psychology*, *71*(10), 2115–2128. https://doi.org/10.1177/1747021817737520

Zhang, H., Kang, C., Wu, Y., Ma, F., & Guo, T. (2015). Improving proactive control with training on language switching in bilinguals. *NeuroReport*, *26*(6), 354–359. https://doi.org/10.1097/WNR.0000000000000353

Zhang, H., Diaz, M. T., Guo, T., & Kroll, J. F. (2021). Language immersion and language training: Two paths to enhanced language regulation and cognitive control. *Brain and Language*, *223*, 105043. https://doi.org/10.1016/j.bandl.2021.105043

Zheng, X., Roelofs, A., Erkan, H., & Lemhöfer, K. (2020). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study. *Neuropsychologia*, *140*, 107387. https://doi.org/10.1016/j.neuropsychologia.2020.107387

Zhu, J. D., Blanco-Elorrieta, E., Sun, Y., Szakay, A., & Sowman, P. F. (2022). Natural vs forced language switching: Free selection and consistent language use eliminate significant performance costs and cognitive demands in the brain. *NeuroImage*, *247*, 118797. https://doi.org/10.1016/j.neuroimage.2021.118797