

Engaging Proactive Control: Influences of Diverse Language Experiences Using Insights From Machine Learning

Jason W. Gullifer and Debra Titone
McGill University

We used insights from machine learning to address an important but contentious question: Is bilingual language experience associated with executive control abilities? Specifically, we assess proactive executive control for over 400 young adult bilinguals via reaction time (RT) on an AX continuous performance task (AX-CPT). We measured bilingual experience as a continuous, multidimensional spectrum (i.e., age of acquisition, language entropy, and sheer second language exposure). Linear mixed effects regression analyses indicated significant associations between bilingual language experience and proactive control, consistent with previous work. Information criteria (e.g., AIC) and cross-validation further suggested that these models are robust in predicting data from novel, unmodeled participants. These results were bolstered by cross-validated LASSO regression, a form of penalized regression. However, the results of both cross-validation procedures also indicated that similar predictive performance could be achieved through simpler models that only included information about the AX-CPT (i.e., trial type). Collectively, these results suggest that the effects of bilingual experience on proactive control, to the extent that they exist in younger adults, are likely small. Thus, future studies will require even larger or qualitatively different samples (e.g., older adults or children) in combination with valid, granular quantifications of language experience to reveal predictive effects on novel participants.

Keywords: bilingualism, language entropy, individual differences, interactional context, proactive control


Supplemental materials: <http://dx.doi.org/10.1037/xge0000933.supp>

Bilingualism, that is, the knowledge and active use of two or more languages, exhibits several hallmarks of an ideal form of cognitive training that should generalize to domains besides language (Diamond & Ling, 2016). Bilingual language use is rigorous and cognitively taxing: bilinguals experience constant activation and competition from the nontarget language that they are thought to manage, in part, by applying domain general executive control (Gullifer & Titone, 2019b; Kroll, Gullifer, & Zirnstein, 2016; Lauro & Schwartz, 2017; Pivneva, Mercier, & Titone, 2014; Whitford, Pivneva, & Titone, 2016). Bilingualism is quite often continual and unrelenting: onset of bilingualism can occur from birth, and there may be no upper limit on the number of hours per day that people can spend using language if activities like thinking and dreaming are included as valid forms of language experience (Bialystok, 2017). Bilingualism is also highly socioculturally relevant: language use has an impact on economic

outcomes (Saiz & Zoido, 2005), social-emotional well-being (Chen, Benet-Martínez, & Harris Bond, 2008; Doucerain, Varnaamkhaasti, Segalowitz, & Ryder, 2015; Han, 2010), and political attitudes (Heller, 1992), particularly in parts of the world where multiple languages are in contact with one another. Bilingualism is not a single, dichotomous construct (e.g., Luk & Bialystok, 2013), and bilinguals vary along many dimensions of language ability, exposure, and usage (e.g., Baum & Titone, 2014; Tiv, Gullifer, Feng, Titone, 2020). Our goal here is to assess whether individual differences in bilingual language experience in younger adults, along multiple dimensions predict individuals' engagement of executive control, at a behavioral level. To this end, we analyze proactive executive control data from a sample of over 400 bilinguals from Montréal, Québec, Canada between the ages of 18 and 35.

Editor's Note. Alessandra S. Souza served as the guest editor for this article.—NC

This article was published Online First October 1, 2020.

 Jason W. Gullifer and Debra Titone, Department of Psychology and Centre for Research on Brain, Language and Music, McGill University.

Data and code will be made available here: <https://osf.io/eaks2/>.

This work was presented previously at the 2019 Annual Meeting of the Psychonomic Society. This work was supported by the Natural Sciences and Engineering Research Council of Canada (individual Discovery Grant, 264146 to Debra Titone); the Social Sciences and Humanities Research

Council of Canada (Insight Development Grant, 01037 to Jason W. Gullifer and Debra Titone); the National Institutes of Health (Postdoctoral training Grant, F32-HD082983 to Jason W. Gullifer and Debra Titone), and the Centre for Research on Brain, Language & Music. Jason W. Gullifer designed research, contributed analytic tools, analyzed data, and drafted the manuscript. Debra Titone designed research, analyzed data, and revised the manuscript. Both authors approved the final version of the manuscript for submission.

Correspondence concerning this article should be addressed to Jason W. Gullifer or Debra Titone, Department of Psychology, McGill University, 2001 McGill College Avenue, Montréal, QC H3A 1G1, Canada. E-mail: jason.gullifer@mail.mcgill.ca or debra.titone@mcgill.ca

There is now substantial experimental evidence to suggest that bilinguals exhibit differential executive control performance relative to monolinguals (e.g., Bialystok, 1992; Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2012; Bialystok & Martin, 2004; Costa, Hernández, & Sebastián-Gallés, 2008; Peal & Lambert, 1962; Prior & MacWhinney, 2010) and that aspects of bilingual experience explain modulations in executive control performance (Dash, Berroir, Ghazi-Saidi, Adrover-Roig, & Ansaldo, 2019; Donnelly, Brooks, & Homer, 2019; Grundy, Chung-Fat-Yim, Friesen, Mak, & Bialystok, 2017; Gullifer et al., 2018; Hartanto & Yang, 2016; Hofweber, Marinis, & Treffers-Daller, 2016; Jylkkä et al., 2017; Kousaie, Chai, Sander, & Klein, 2017; Morales, Gómez-Ariza, & Bajo, 2013; Navarro-Torres, García, Chidambaram, & Kroll, 2019; Prior & Gollan, 2011; Soveri, Rodríguez-Fornells, & Laine, 2011; Verreyt, Woumans, Vandelandotte, Szmalec, & Duyck, 2016; Zimstein, van Hell, & Kroll, 2018a, 2018b). The primary theoretical argument is that cognitive and linguistic systems and their corresponding brain networks adapt to meet the demands of the environment (i.e., the adaptive control hypothesis; Abutalebi & Green, 2016; Green & Abutalebi, 2013; Green & Wei, 2014). In particular, environments where two or more languages are used may be particularly demanding for a variety of control processes including active goal maintenance and conflict monitoring (Green & Abutalebi, 2013). However, the relationship between bilingualism and executive control has been contentious because of high variance in the outcomes of individual studies and meta-analytical reports of small effect sizes that do not survive corrections for publication bias (de Bruin & Della Sala, 2019; de Bruin, Treccani, & Della Sala, 2015; Donnelly et al., 2019; Lehtonen et al., 2018; Paap, 2019, 2014; Paap, Johnson, & Sawi, 2015).

There are at least three root causes of conflicting results, not yet exhaustively assessed, that we address in the present study. First, people vary continuously along many dimensions in terms of language exposure, use, and ability, but there is often a failure to assess these experiences as a continuous dimension (though this point is garnering attention within the field; e.g., Anderson, Mak, Keyvani Chahi, & Bialystok, 2018; Baum & Titone, 2014; Bice & Kroll, 2019; de Bruin, 2019; DeLuca, Rothman, Bialystok, & Pliatsikas, 2019; Gullifer et al., 2018; Gullifer & Titone, 2019a; Hartanto & Yang, 2016; Paap, Anders-Jefferson, Mikulinsky, Masuda, & Mason, 2019; Poarch, Vanhove, & Berthele, 2019; Surrain & Luk, 2019; Takahesu Tabori, Mech, & Atagi, 2018; Titone, Gullifer, Subramaniapillai, Rajah, & Baum, 2017). Instead participants are often dichotomized into groups of potentially heterogeneous bilinguals and potentially heterogeneous monolinguals. These groups are then compared statistically. However, dichotomization ignores the nuances that can exist even within the same sample of speakers (e.g., Gullifer & Titone, 2019a) and can lead to other statistical issues, including reduced effect sizes, reduced power, spurious effects, and difficulties in comparing effects across studies (MacCallum, Zhang, Preacher, & Rucker, 2002).

As a case in point, a recent study took a dichotomization approach and failed to find a relationship between bilingualism and executive control abilities in a sample of over 11,000 people (though the comparison involving relatively matched groups in the Canada, the United States, and Australia involved a sample of approximately 500; Nichols, Wild, Stojanoski, Battista, & Owen, 2020). Crucially, dichotomization into monolingual and bilingual groups was based on a single item from a questionnaire ("How many languages do you speak?"). This is problematic, because people may differentially interpret what it means to "speak" different languages in the context of

highly culturally monolingual countries (i.e., the United States and Australia) versus culturally bilingual countries (i.e., Canada, where English and French have official status federally, and the extent of bilingualism varies dramatically across provinces). As well, this question does not differentiate the many relevant ways in which bilinguals or multilinguals could differ, including historical or current language exposure, the distribution of language use socially, whether they are proficient bilinguals, learners, code-switchers, heritage language speakers, and so forth.

To address the issue of dichotomization, we assess bilingualism through joint interactions of multiple, continuous measures. Specifically, we include a continuous measure of the age at which the L2 was acquired (L2 AoA), an index of sheer L2 exposure, and a measure of bilingual language diversity. L2 AoA is a core measure of historical language experience that is related to language abilities, executive control abilities, and brain organization (see e.g., Johnson & Newport 1989; Flege, Munro, & MacKay, 1995; Gullifer et al., 2018; Kousaie et al. 2017; Luk, De Sa, & Bialystok, 2011; Tao, Marzecová, Taft, Asanowicz, & Wodniecka, 2011). However, L2 AoA is not sufficient to fully capture the bilingual language experience, which is rich and diverse throughout the life span. Current L2 exposure is also crucially important for issues related to language and executive control (Subramaniapillai, Rajah, Pasvanis, & Titone, 2019; Whitford & Titone, 2012, 2015) but this measure fails to fully account for the usage of other languages. Our research group has recently devised a sophisticated measure of bilingual language experience using the concept of entropy (Gullifer et al., 2018; Gullifer & Titone, 2018, 2019a, 2019b; for similar applications of entropy see, e.g., Eleta & Golbeck, 2014; Krein, MacLean, Delorey, Knutson, & Eggett, 2009), and this measure incorporates information about all languages that a bilingual or multilingual reports in a language history questionnaire. Entropy comes from physics and information theory, and it provides a measure of diversity and uncertainty when the relative proportion of occurrences for a set of "states" is known. Here, we take states to be the usage of a particular language (e.g., English or French) within some communicative context (e.g., at home or at work), and we elicit the proportion of usage from participants on a language background questionnaire. From this, we compute a measure of *language entropy* within various communicative contexts, which is thought to track diversity in language usage.

Individuals with low language entropy are *compartmentalized* and tend to use a single language in a context. Individuals with high language entropy are *integrated* and tend to use multiple languages in a balanced fashion within a context. In our view, language entropy provides a means to test predictions of the adaptive control hypothesis related to bilinguals who immerse themselves in language contexts where two (or more) languages are used (dual language contexts) relative to those who immerse themselves in contexts where only one language is used (single language contexts).¹ To assess the construct

¹ Critically, language entropy, as quantified here, does not isolate the extent to which participants engage in dense code-switching. Dense code-switching is a core interactional context that the adaptive control hypothesis predicts should require an open control mode, in which there is no competition between languages (Abutalebi & Green, 2016; Green & Abutalebi, 2013; Green & Wei, 2014). There is mixed evidence for this aspect of the hypothesis in the literature (e.g., Adler, Valdes Kroff, & Novick, 2020), and we cannot test it here because we do not have information code-switching behavior of these participants.

validity of this measure, we have shown that language entropy can be decomposed into two latent components, reflecting general language entropy across several communicative contexts (i.e., general entropy) and language entropy in professional contexts (i.e., work entropy). We further show that these components differentially relate to self-reported second language (L2) outcome measures such as accentedness and L2 ability ratings (Gullifer & Titone, 2019a).

A second issue with respect to bilingualism and executive control is that while there is an intense focus on control in a general sense and on particular control tasks, most research does not distinguish between the various mechanisms, or modes, of executive control. Executive control can be decomposed into at least two core modes: reactive control that occurs in the moment to suppress an irrelevant stimulus or attend to a relevant stimulus measured, and proactive control that is applied to resolve conflict ahead of time (Braver, 2012; Braver, Gray, & Burgess, 2007). It is likely that all control tasks recruit both modes of control, yet proactive control is only recently being assessed in literature on bilingualism (Beatty-Martínez et al., 2020; Berry, 2016; Declerck, Koch, Duñabeitia, Grainger, & Stephan, 2019; Gullifer et al., 2018; Ma, Li, & Guo, 2016; Morales et al., 2013; Morales, Yudes, Gómez-Ariza, & Bajo, 2015; Zhang, Kang, Wu, Ma, & Guo, 2015; Zirnstein et al., 2018a, 2018b), and few studies assess the relationship with bilingualism by way of continuous indices of bilingual language experience.

Proactive control requires, among other things, active maintenance of a goal throughout a task. Goal maintenance (together with conflict monitoring) is thought to be in particularly high demand and exercised for bilinguals with high language entropy (i.e., bilinguals immersed in dual language contexts; Green & Abutalebi, 2013). High entropy bilinguals experience the activation of multiple languages simultaneously across multiple speakers in a communicative context. To the extent that any given conversation for a high entropy bilingual occurs primarily in a single language (i.e., without code-switching), effective communication may benefit from the ability to maintain the communicative goal (i.e., speak in the intended language with the intended interlocutor) by allowing a speaker to preemptively regulate the unintended language (e.g., Declerck et al., 2019; Ma et al., 2016). Such tracking and regulation might depend on the integration of multiple cues from multiple modalities, including visual cues, such as the face of an interlocutor; coarse auditory cues, such as a speaker's tone of voice; or subtle auditory cues such as phonetic markers of an upcoming language switch. Thus, if executive control abilities adapt to meet the demands of a communicative context, then high entropy bilinguals might be expected to rely more on this component than low entropy bilinguals.

We and others have begun to assess the extent to which bilingual language experience is associated with proactive executive control using an AX continuous performance task (AX-CPT; Beatty-Martínez et al., 2020; Berry, 2016; Gullifer et al., 2018; Morales et al., 2013, 2015; Zhang et al., 2015; Zirnstein et al., 2018a, 2018b). In the AX-CPT, goal-relevant information is encoded in cue items that must be used to respond to upcoming target items within each trial. At the group level, studies on healthy younger adults commonly show evidence of attention to this goal-relevant information (Barch et al., 1997; Braver & Barch, 2002; Carter et al., 1998; Locke & Braver, 2008), marked by

slower and less accurate responses on trials where target items are unlikely given the cue. In other words, people tend to plan their responses ahead of time, proactively, through the active maintenance of goal-relevant information. If expectations are violated, then people must revise their planned responses.

At the group level, there is some evidence that bilinguals adjust proactive and reactive control more efficiently relative to monolinguals (Morales et al., 2013, 2015), whereby bilinguals (relative to monolinguals) evidence fewer errors on trials in which proactive control should hinder performance in the absence of group differences in RTs. Among bilinguals, we have shown that language entropy and L2 AoA are associated with independent patterns of brain connectivity among areas involved in the engagement of executive control, and in turn, these patterns of brain connectivity are associated with RTs on the AX-CPT (Gullifer et al., 2018). For example, we show that high entropy bilinguals (relative to low entropy bilinguals) exhibit more connectivity between regions implicated in goal maintenance and articulation. In turn, they exhibit slower RTs on trials where proactive control is expected to hinder performance relative to trials where it is expected to facilitate performance. This pattern suggests that individual differences in bilingual language experience are associated with individual differences in the maintenance of goal-relevant cue items. Thus, proactive control appears to be a crucial component of executive control that may be molded by both historical and current language experiences.

A third issue generally raised in the social sciences, applied to the domain of bilingualism and executive control, is the potential for publication bias (e.g., Easterbrook, Gopalan, Berlin, & Matthews, 1991) or other questionable research practices (such as p-hacking; Nuzzo, 2014), as well as an overreliance on small sample sizes among published studies (de Bruin & Della Sala, 2019; de Bruin et al., 2015; Donnelly et al., 2019; Lehtonen et al., 2018; Paap, 2019, 2014). According to this line of argument, the culture within the social sciences has resulted in a publication bias where studies with significant positive results to be published over studies with negative or null results. Moreover, many studies in this domain have traditionally relied on small sample sizes, which can result in misestimated effect sizes. (see e.g., Type-M or Type-S errors; Gelman & Carlin, 2014). It is well known that the effect sizes for effects related to bilingual experience may, in fact, be quite small, particularly when considering samples of young adults who are arguably operating at peak cognitive capacity (Bialystok, Martin, & Viswanathan, 2016). Thus, when all of these factors are considered together, one could argue that evidence for effects of bilingualism on executive control are due to researchers overfitting their data or publishing only positive results.

To help address these theoretical and methodological issues, we take two steps here. First, we report data on a large sample (relative to past studies) of young adult bilinguals, which should provide a more accurate estimate of the effect sizes. Second, following suggestions of Yarkoni and Westfall (2017), we adopt key insights from machine learning approaches. In many machine learning contexts data sets are often large, effects may be small, and models may have many degrees of freedom. Thus, the potential for models to overfit data (and find many significant small effects) is highly prevalent. A key insight in machine learning contexts is the importance of acknowledging the potential for overfitting, identifying its presence, and constructing models in ways to avoid it.

Overfitting can be assessed by measuring or estimating the ability of a model to predict new, unmodeled data. A model's predictive performance on new data can be evaluated in three ways. Predictive performance of models that have been fit to existing data sets can be evaluated by collecting a new set of data (validation set), and measuring how well predictions from those models match the new data. Unfortunately, this approach is highly resource intensive, as it requires immense amounts of data to ensure appropriate sampling. However, predictive performance on new data can also be estimated, either directly by performing cross-validation, or indirectly through the use of an information criterion (see, e.g., [Friedman, Hastie, & Tibshirani, 2001](#); [James, Witten, Hastie, & Tibshirani, 2013](#); [McElreath, 2018](#)). Cross-validation involves breaking a dataset up into several subsets (i.e., folds), fitting models on some of these folds, and assessing fit or prediction of the models on the remaining folds that were not used for fitting. A related insight from machine learning is that parsimonious (or simple models), can reduce overfitting and lead to better predictive performance on new, unmodeled data. In other words, fitting data less closely with a parsimonious model may yield better prediction of new data. Various methods are available to yield parsimonious models, including penalization of the number of parameters in a model (e.g., through information criteria) and applying regularization to regression (e.g., [Friedman et al., 2001](#); [Friedman, Hastie, & Tibshirani, 2010](#); [James et al., 2013](#)). One such model that uses regularization is least absolute shrinkage and selection operator (i.e., LASSO) regression ([Tibshirani, 1996](#)). [Yarkoni and Westfall \(2017\)](#) argue that incorporating insights like these from machine learning can help Psychology become a predictive science and allow us to better understand human behavior by eliminating spurious effects that are not predictive.

Thus, the goal of the present study is to assess whether a constellation of continuous individual differences in bilingual language experience predict engagement of proactive control, at a behavioral level. We extract AX-CPT RT data for a sample of over 400 multilingual adults, and we model that data as a function of several features related to bilingual language experience, including L2 AoA, L2 exposure, and two language entropy components. We present an analysis of trial-level AX-CPT data using linear mixed effects regression (Analysis 1) and regularized LASSO (least absolute shrinkage and selection operator) regression (Analysis 2).

For linear mixed effects models, we compare theoretically relevant model parametrizations via several approaches, including likelihood ratio tests (LRTs), Akaike Information Criterion (AIC), and leave-one-out cross-validation. LRTs constitute a traditional approach to model evaluation in the psychological and language sciences (e.g., [Baayen, 2008](#); [Cunnings, 2012](#); [Jaeger, 2008](#)). However, there is some contention over whether LRTs can lead to overfitting, presumably because they represent an approach concerned with goodness of fit rather than prediction, and because the selection procedure relies on p values. Some have argued that LRTs can overfit when they are used to select fixed effects ([Pinheiro & Bates, 2009](#)), while others argue that LRTs do not necessarily increase Type I error rate ([Baayen, Davidson, & Bates, 2008](#); [Barr, Levy, Scheepers, & Tily, 2013](#)). Thus, in addition to reporting LRTs, we also use AIC and cross-validation as ways to estimate predictive performance and help control overfitting. For regularized LASSO regression, we supply the largest model parametrization and use cross-validation to determine an optimal value of a regularization parameter (λ) that leads to the best

predictive performance. The regularization parameter effectively controls how simple or complex a model is by pushing model coefficients (i.e., fixed effects) toward zero as the parameter is increased.

If individual differences in bilingual language experience are key factors related to proactive control, then these factors should be retained in the model by selection procedures. Ideally, these features would be deemed important by approaches that emphasize model fit (i.e., LRTs) and those that balance model fit and complexity (i.e., AIC and cross-validation). Given previous work on a smaller sample of bilinguals from Montréal ([Gullifer et al., 2018](#)), we expect that L2 AoA and language entropy should each interact with AX-CPT trial type in predicting RTs. In particular, we expect these variables to interact with the critical trial type contrast that taps into goal-maintenance and conflict monitoring (the AY-BX contrast, see the section *Assessing proactive control using the AX continuous performance task*, below).

Using linear mixed effects regression models, we show that both LRTs and AIC converge on the same model specification that includes interactions between trial type and language entropy components (though there were no interactions with L2 AoA), partially in line with predictions. This suggests that the inclusion of language entropy components results in the best fit to AX-CPT data and may be of use in predicting new data. When we directly estimate predictive performance via cross-validated linear mixed effects regression and via regularized LASSO regression, there is still support for the inclusion of these features related to bilingual language experience. However, the support is weakened because these procedures indicate that similar predictive performance could be obtained through the use of simpler models that do not include individual differences in bilingual language experience. Thus, although there is a suggestion that bilingual language experience provides signal for the estimation of proactive executive control, these effects may be too small to sufficiently predict novel, unmodeled data.

Materials and Method

Participants

Approximately 459 bilingual or multilingual participants were tested who reported detailed language history information (including the relative exposure to and use of two or more languages) and AX-CPT data. The language history data of these participants was previously analyzed and reported in [Gullifer and Titone \(2019a\)](#), and these data are reanalyzed in the Supplemental Information with a machine learning approach. Here 58 participants were excluded from the analyses because they did not complete the AX-CPT. The sample was roughly equally split between bilinguals with English as their first language (L1; $n = 222$) and French as their L1 ($n = 237$). A number of participants ($n = 127$) reported current exposure to a third language (L3) besides English or French. See [Table 1](#) for a summary of demographic and language history data. Before participation in the experiments, all participants gave informed consent to have their data from experiments and questionnaires collected, stored, and analyzed. The McGill Research Ethics Board Office approved this research.

Some of the bilingual language experience variables that we extracted refer to participants' work environments. The sample con-

Table 1
Participant Demographics

Measure	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	Skew	Kurtosis
Age (years)	22.754	3.633	18	35	1.3	1.2
L2 AoA (years)	6.752	4.334	0	28	1	2.24
Years bilingual	16.002	5.34	1	34	0.01	0.42
L1 exposure (/100)	63.69	20.576	5	98	-0.48	-0.43
L2 exposure (/100)	33.89	20.07	2	95	0.6	-0.21
L3 exposure (/100)	2.42	5.648	0	40	3.22	11.67
Reading entropy	0.595	0.41	0	1.571	-0.16	-1.14
Speaking entropy	0.695	0.417	0	1.585	-0.3	-0.74
Home entropy	0.602	0.456	0	1.585	-0.05	-1.06
Work entropy	0.756	0.367	0	2	-0.82	0.25
Social entropy	0.944	0.276	0	1.585	-0.39	1.79
L2 accentedness (/7)	3.625	1.787	1	7	0.16	-1.09
L2 ability (/10)	7.319	1.847	1	10	-0.62	0.53

Note. This information includes self-report information about chronological age, age of acquisition of the L2, current exposure to each language (L1–L3), language entropy in various communicative contexts (reading, speaking, home, work, and social), L2 accentedness (7 = *strong accent*), and L2 abilities (10 = *high ability*).

sisted mainly of students from the Montréal area, and the questionnaire did not differentiate between work and school contexts. However, participants reported their current occupation. Approximately 300 reported “student” as their current occupation, 132 participants reported a different occupation besides student, 16 reported student together with another occupation, and 11 reported no occupation.

Materials

Assessing language experience. All participants included in this sample completed a language background questionnaire adapted from the LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007) and LHQ 2.0 (Li, Zhang, Tsai, & Puls, 2014). This allowed us to probe language usage and experience within the Montréal context. For the purposes of the analysis, we extracted or computed several background measures, outlined below.

Basic demographic information. In the questionnaire, participants reported basic information about their demographics and language use. We extracted two classic measures of L2 experience: L2 AoA (based on the onset of learning) and global exposure to the L2. Global L2 exposure is frequently used as a covariate in our lab (Gullifer & Titone, 2019b; Pivneva et al., 2014; Subramaniapillai et al., 2019), and did not factor into the computation of language entropy.

Language exposure in different usage contexts. Participants reported the extent to which they used or preferred to use the L1, L2, and L3 in a variety of communicative contexts in the home, at work, in social settings, for reading, and for speaking. The questionnaire assessed language use at home, work, and in social settings via Likert scales (e.g., “Please rate the amount of time you use each language at home”), with a score of 1 indicating “no usage at all” and a score of 7 indicating “usage all the time” or “a significant amount.” Labels were not provided for intervening scores. We baselined Likert data at 0 by subtracting 1 from each response. Thus, a value of 0 reflects “no usage at all.” We converted these data to proportions of usage by dividing a given language’s score by the sum total of the scores within context. For example, a participant who reported (after baselining) the follow-

ing data for language usage at home, L1: 6, L2: 5, L3: 0, would receive the following proportions for the home context, L1: 6/11, L2: 5/11, L3: 0/11.

Language use for reading and speaking were collected through percentage of use (“What percentage of time would you choose to read/speak each language?”), which totaled to 100% within any given context. We converted percentages to proportions, and we used this proportional usage data to compute language entropy in each context.

Computing language entropy. For each usage context (see “Language exposure in different usage contexts,” above), we computed language entropy (H) with the following equation $H = -\sum_{i=1}^n P_i \log_2(P_i)$ using the methods available in the languageEntropy R package (Gullifer & Titone, 2018). Here, n represents the total possible languages within the context (e.g., two or three) and P_i is the proportion that language _{i} is used within a context. Thus, if a bilingual reported using French 80% of the time and English 20% of the time within the work context, one would compute language entropy by summing together $0.80 \cdot \log_2(0.80)$ and $0.20 \cdot \log_2(0.80)$ and then multiplying by -1 to yield a positive language entropy value. This hypothetical individual’s language entropy in the work context would be approximately 0.72.

Theoretically, the entropy distribution has a minimum value of 0 that occurs when the proportion of usage for a given language is 1.0, representing a completely compartmentalized context. The distribution has a maximum value equal to $\log n$ (1 for two languages and approximately 1.585 for three languages) when the proportion of use for each language is equivalent, representing a completely integrated context. This procedure resulted in five entropy scores for each participant that pertained to language entropy in the following communicative contexts: home, work, social, reading, and speaking.

Assessing proactive control using the AX continuous performance task. In the AX-CPT (see, e.g., Braver, 2012; Servan-Schreiber, Cohen, & Steingard, 1996 for introductions), participants view a continuous series of letters, and are instructed to respond “yes” if and only if the current letter is an X and the prior letter was an A; otherwise they respond “no.” AX trial types occur 70% of the time, establishing a strong impulse to respond yes on

all trials, and particularly when the prior letter is an A or the current letter is an X. Two critical no conditions provide an index of proactive control. In AY trial types, people first see an A, and then see a non-X letter (here, sampled from the letters F, K, L, N, P, U, V, W, X, Y, and Z). If participants use proactive control to prepare a yes response for the subsequent trial upon seeing an A, their performance should suffer when a non-X letter appears and a no response is required. In BX trial types, by contrast, people first see a non-A letter (though in our version, participants always see the letter B), and then see an X.² If participants use proactive control to prepare a no response for the subsequent trial upon seeing the non-A letter cue, performance should improve when an X appears and a no response is required. Unlike proactive control, a reactive control strategy (i.e., reduced maintenance of the cue) should benefit performance in the moment for AY trials, but it may hinder performance in on BX trials. Thus, one can obtain a continuous measure from the AX-CPT that reflects how “proactive” a given individual is by comparing AY performance to BX performance.

Data from the AX-CPT were collected through multiple waves of data collection over many years. Thus, some specific details regarding the nature of the task are not available. The available procedures and parameters were as follows. The task was programmed and presented with E-Prime Version 1 software (Psychology Software Tools, 2020). Participants saw a series of letters presented individually, and we instructed them to press a key responding to yes if a character was an X immediately following an A; otherwise they were instructed to press a key corresponding to no. Letters were presented in continuous cue–target pairs, and each letter was followed by a blank screen. All screens had a duration of 1,000 ms. Responses were collected for 1,000 ms after the onset of each letter. If a response was not made within this window, participants were scored with an error. Of note, there were three versions of the task that differed in the number of trials (100 trials vs. 200 trials) or in the ratio of AX trials (72/100 trials vs. 70/100 trials). The general pattern of result presented below did not change when we controlled for the experiment version.

Results

Data were processed and analyzed in R, Version 3.6.0 (R Core Team, 2019). Before modeling, we preprocessed the data in the following manner. For subject-level language history data, we z-scored numeric features (i.e., L2 AoA and L2 exposure). We also conducted principal component analysis (PCA) on the five entropy features, following procedures outlined in Gullifer and Titone (2019a). In that article, we identified two oblique principal components. The first component primarily included loadings from reading, speaking, home, and social entropy, and this component explained 44% of the variance in the data. The second component primarily included loadings from work entropy with some cross-loading from social entropy, and this component explained 21% of the variance in the data. The correlation between the two components was 0.33. We extracted the component scores for each participant to serve as indices of language entropy at work and language entropy everywhere else.

For trial-level AX-CPT data, we removed RTs below 40 ms as outliers (1.5% of the data). Outlier detection was based on visual inspection of a density plot. This outlier threshold is low compared with the typical approach in the cognitive sciences (e.g., 200 ms).

However, given that the goal of the task is to prepare responses proactively, it is likely that participants could anticipate upcoming trials and program behavioral responses very quickly. There were no RTs above 1,000 ms because of the programming of the AX-CPT experiment (nonresponses were coded as errors), and, thus, we did not set an upper threshold for outlier detection. We also identified four participants with mean RTs that were greater than 2.5 *SD* above the sample average, and we removed these participants from further analysis. Before modeling, we separated the AX-CPT data into $k = 455$ folds, such that each participant was assigned a fold, and all trials associated with a participant were kept together within the fold. Trial type was treatment coded as a three-level factor with “BX” as the reference level.

Aggregated RT data at the group level are illustrated in Figure 1 (Panel A) and corresponding accuracy data are illustrated in Figure 1 (Panel B). Visual inspection of this figure indicates that AY trials were approximately 111 ms slower and 4% less accurate compared with BX trials. This pattern of results indicates that, overall, participants approached the task with a proactive strategy. In the [online supplemental materials](#) we provide scatterplots to illustrate these data as they relate to individual differences in bilingual language experience.

Analysis 1: AX-CPT, Mixed-Effects Regression

Using linear mixed effects regression, we fit several theoretically driven, nested models to unstandardized RTs of entire AX-CPT dataset.³ Specifically, we fit a set of four nested models using the *lme4* package (Bates, Machler, Bolker, & Walker, 2015). The first model included a main effect of trial type (three-level variable, BX, AY, BY; treatment coded with BX as the reference level). The second model additionally included two-way interactions between trial type and L2 AoA, and trial type and sheer L2 exposure. The third model additionally included two-way interactions between each entropy component and trial type. The fourth model additionally included all three-way interactions with L2 AoA. Of note, we also fitted an alternative second model that included condition by entropy interactions without L2 AoA and sheer L2 exposure. This specification allowed us to perform an additional series of model comparisons to assess whether entropy components explain unique variance compared with L2 AoA and sheer L2 exposure (by substituting this model in for the second model). Model specifications are illustrated in Table 2.

² Typically, in the AX-CPT, B cues are randomly selected from various letters of the alphabet on each trial whereas A cues consist only of the letter A, resulting in an exceptional status for the A cue. Our version of the AX-CPT was simplified such that B cues were not sampled, and instead, consisted of the fixed B character. This version of the task has been used elsewhere (Gullifer et al., 2018; Kam, Dominelli, & Carlson, 2012; Ličen, Hartmann, Repovš, & Slapničar, 2016). However, we note that the fixed B cue could function to mask the exceptional status of the A cue, which could decrease reliance on proactive control relative to a task with sampled B cues.

³ Accuracy data were imbalanced in the sense that participants were overall highly accurate (90% accurate across all “no” trials). This imbalance presents challenges in modeling accuracy data through logistic regression. Notably, models can achieve good fit by predicting accurate responses for all trials. For these reasons, we have restricted the analyses in the body of the article to reaction time data. For completeness, we present analyses of accuracy data in the [online supplemental materials](#).

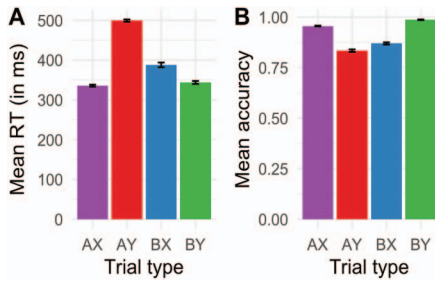


Figure 1. Illustration of the group-level, aggregate effects for AX continuous performance task (AX-CPT) reaction time (RT) in milliseconds (A) and percent accuracy (B). Error bars illustrate 1 *SEM*. AY trials were overall slower and less accurate compared with BX trials, suggesting that, overall, participants approached the task with a proactive control strategy. See the online article for the color version of this figure.

To compare the models, we took three approaches. The first approach involved using LRTs to test whether restricted models fit the data significantly worse than richer models. This approach does not consider predictive accuracy of the models. This approach has become standard procedure in the psychological and language sciences but may lead to overfitting of the data (Pinheiro & Bates, 2009). We also used two prediction centric approaches: we estimated predictive performance of the models indirectly using AIC, and we then estimated predictive performance directly using leave-one-out cross-validation.⁴ For leave one out cross-validation, we constructed an algorithm following the procedures outlined in (James et al., 2013). Thus, we wrote a function to fit models to the participants in the *k*-1 training folds. For each model, predictive accuracy (root mean square error, RMSE) and model fit (R^2) were computed for all participants in the training folds, and, more importantly, for the participant in the holdout test fold. The process was repeated such that each participant participated once as a holdout fold. We then averaged RMSEs and R^2 for each of the training sets (to compute $RMSE_{train}$ and R^2_{train}) and for each of the testing sets (to compute $RMSE_{test}$ and R^2_{test}). Thus, the metrics for the testing sets provide more direct estimates of the predictive accuracy and fit to unmodeled data. All linear mixed effects regression models included a random intercept for subject. We did not add a random slope for trial type by subject, as our purpose was to assess fixed effects related to individual differences, and these would have been completely captured by a random slope by subject. *p* values were computed using degrees of freedom approximated using the Satterthwaite approximation (Kuznetsova, Brockhoff, & Christensen, 2017).

Model comparisons via LRTs indicated that the best fitting model was the one that included interactions between trial type and each of the two entropy components, which improved fit over the model with a main effect of trial type alone, $\chi^2(6) = 19.09$, $p = .004$. Model selection via AIC indicated that the same model was the best fitting model (AIC = 254,746.0; $\Delta AIC = 7.1$ vs. model with trial type alone) suggesting this model should evidence the best predictive performance as well. Results for LRTs and AIC metrics are illustrated in Table 3.

Model selection via cross-validation provided further support that the best fitting, predictive model was the same as the one identified via LRTs and AIC (i.e., interactions between trial type

and each of the two entropy components; $M_{RMSE_{test}} = 161.362$, $SEM_{RMSE_{test}} = 1.639$, $M_{R^2_{test}} = 0.207$). Critically, however, visual inspection of the cross-validation results warrants some skepticism toward this finding. All models performed similarly in terms of R^2_{test} and $RMSE_{test}$ performance on the hold-out fold. A common, though somewhat arbitrary, approach in machine learning is to choose a parsimonious model: the simplest model (fewest DF) that evidences performance within 1 *SEM* of the best performing model. Under this criterion, the model that contains trial type as the only fixed effect would be selected as the parsimonious model ($M_{RMSE_{test}} = 161.433$, $SEM_{RMSE_{test}} = 1.623$, $M_{R^2_{test}} = 0.207$). Cross-validation results are illustrated in Table 4.

We inspected the best fitting model with interactions between trial type and language entropy components to assess specific direction of effects. See Table 5 for a summary of this model, including effect sizes and standardized β coefficients. Inspection of the model, $B_{intercept} = 384.98$, 95% confidence interval (CI) [377.23, 392.74], $t(607.86) = 97.28$, $p < .001$, indicated the following: AY trials were responded to slower than BX trials, $B_{AY} = 113.83$, 95% CI [108.72, 118.95], $t(19,364.32) = 43.61$, $p < .001$, and BY trials were responded to more quickly than BX trials, $B_{BY} = -42.14$, 95% CI [-47.04, -37.24], $t(19,351.92) = 16.85$, $p < .001$. The AY and BX difference suggests that participants approached the task with a proactive strategy. However, there were also significant effects and interactions related to language entropy.

There was no statistical evidence for a simple effect for the general entropy component, $B_{general} = -4.24$, 95% CI [-12.62, 3.77], $t(605.25) = 1.05$, $p = .29$; however, it did interact with the AY, $B_{AY} \times general = 9.27$, 95% CI [3.89, 14.65], $t(19,363.65) = 3.37$, $p = .001$, and BY, $B_{BY} \times general = 7.45$, 95% CI [2.29, 12.61], $t(19,352.413) = 2.82$, $p = .005$, levels of trial type. Specifically, as general entropy increased, so did the difference between AY and BX, suggesting that an increase in general entropy was associated with a greater engagement of proactive control via a widening of the AY BX contrast. See Figure 2 (Panel A) for an illustration of the effects related to general entropy. There was a simple main effect of the work entropy component, $B_{work} = 9.54$, 95% CI [1.34, 17.43], $t(608.076) = 2.27$, $p = .023$. There was evidence for an interaction between this component and the AY levels of trial type, $B_{AY} \times work = -5.45$, 95% CI [-10.86, -0.37], $t(19,361.25) = 1.97$, $p = .048$, suggesting that an increase in work entropy was associated a reduced engagement of proactive control via a narrowing of the AY BX contrast. See Figure 2 (Panel B) for an illustration of the effects related to work entropy.

Analysis 2: AXCPT, Leave-one-out LASSO Regression Approach

We used LASSO regression with leave-one-out cross-validation to assess RT performance on the AX-CPT in relation to trial type, individual differences in bilingual language experience, and interactions between them (model specification: L2 AoA \times Trial type \times L2_exposure + L2 AoA \times Trial type \times General Entropy

⁴ While we focus on AIC in text for the purposes of model selection, we also report BIC in the tables of results. For the results here, BIC prefers the simplest model specification because of heavier penalization.

Table 2
Model Specifications

Model	df	Fixed effects
1. Trial type only	5	Trial type
2a. Trial type by L2 AoA and exposure	11	Trial type \times aoa + Trial type \times L2_exposure
2b. Trial type by entropy	11	Trial type \times PCA_Work + Trial type \times PCA_General
3. Trial type by L2 AoA, exposure, entropy	17	Trial type \times aoa + Trial type \times L2_exposure + Trial type \times PCA_Work + Trial type \times PCA_General
4. All three-ways with L2 AoA	26	aoa \times (Trial type \times L2_exposure + Trial type \times PCA_Work + Trial type \times PCA_General)

Note. Random effects included random intercepts by subject (1 | Subject).

Component + L2 AoA \times Trial type \times Work Entropy Component). Models were fit using the *glmnet* package (Friedman et al., 2010), and the model standardizes predictors and outcomes before fitting.

The cross-validation procedure and model fitting procedure are implemented in the *cv.glmnet* function, and we called this function with the *k* parameter set to 459 (to run leave-one-out cross-validation) and the alpha parameter set to 1 (to apply the LASSO penalty). The model fitting procedures are described in more depth elsewhere (Friedman et al., 2001, 2010; James et al., 2013; Tibshirani, 1996), but we outline the procedure here. Generally, the cross-validation procedure tunes a regularization parameter (λ). This parameter applies a constraint to the linear regression model that functions to minimize the absolute value of the model coefficients in the model in addition to minimizing the sum of squared errors between data points and the best fit line. This procedure yields two models of interest: a model that minimizes cross-validation error and a more parsimonious Model 1 *SEM* model. The minimizing model yields the lowest error metric when predicting out-of-sample data (e.g., RMSE). However, a common guideline specifies that, for parsimony, one should select the simplest model that performs within 1 *SEM* of the minimizing model's performance (Friedman et al., 2001). In other words, if a set of models achieve similar performance in predicting out-of-sample data, it is best to select the model with the simplest parametrization. Thus, the parsimonious model is fit with a stronger value of λ that results in additional regularization but that still falls within 1 *SEM* of the minimizing model. Here, we report both models. Throughout this section, the *SEM* is reported in brackets for RMSE (as opposed to the 95% CI).

Figures 3 and 4 illustrate the impact of the regularization parameter, λ , on RMSE (Figure 3A) and model coefficients (Figure 3B, Figure 4). Under a high regularization parameter, model coefficients converge to zero and the model makes high prediction errors. However, as the regularization parameter is lessened, coefficients become nonzero. As the model fits the data, the errors decrease to a point, and may gradually increase as the model begins to overfit the data. Leave-one-out cross-validation identified the minimizing model ($\log \lambda = 0.26$, RMSE = 164.60 [162.84, 166.33]) and the parsimonious model within 1 *SEM* of the minimizing model ($\log \lambda = 2.96$, RMSE = 166.05 [164.31, 167.76]).⁵ The minimizing model retained features related to trial type and one feature related to bilingual language experience. However, the parsimonious model retained only features related to trial type.

The minimizing model, $B_{\text{intercept}} = 380.82$, contained an effect of the AY trial type, $B_{\text{AY}} = 118.89$, suggesting that AY trials were

slower than BX trials, and an effect of the BY trial type, $B_{\text{BY}} = -36.56$, suggesting that BY trials were responded to more quickly than BX trials. There was no simple main effect of the general language entropy component; however, this component interacted with the AY trial type, $B_{\text{AY}} \times \text{general} = 2.43$, and the BY trial type, $B_{\text{BY}} \times \text{general} = 1.95$. This suggests that high general entropy is associated with greater engagement of proactive executive control. There was also an interaction between the general entropy component and L2 AoA, $B_{\text{aoa}} \times \text{general} = -2.66$, suggesting that high general entropy and late L2 AoA were associated with faster RTs overall. There was a simple main effect of the work language entropy component, $B_{\text{work}} = 6.24$, and no indication for an interaction between this component and the levels of trial type, suggesting that an increase in work entropy was associated with slower performance overall. There was also an interaction between the work language entropy component and L2 AoA, $B_{\text{aoa}} \times \text{work} = 4.20$, suggesting that the combination of late L2 AoA and high work entropy was associated with slower RTs. Finally, the model also included an interaction between L2 exposure and the AY trial type, $B_{\text{aoa}} \times \text{exposure} = 0.31$, and no main effect for L2 exposure, similarly suggesting that high L2 exposure was associated with greater engagement of proactive executive control. These effects are illustrated in Figure 5. All other effects were regularized to 0. In comparison, the parsimonious model contained only main effects of trial type, $B_{\text{intercept}} = 380.27$, $B_{\text{AY}} = 92.66$, $B_{\text{BY}} = -11.51$. The cross-validated R^2 of the minimizing model was 0.14 and the cross-validated R^2 of the parsimonious model was 0.12. The results of the LASSO regression are illustrated in Table 6.

Summary of Results

Using a traditional model comparison approach, we found evidence for main effects of, and interactions between features related to bilingual language experience on executive control performance. Increased work entropy component scores were associated to slower RT overall and a barely significant reduction in AY-BX differences, suggesting that individuals with high work entropy engaged proactive control to a lesser degree relative to individual with low work entropy. Increased general entropy component scores were associated with larger AY-BX differences in RT,

⁵ Of note, we also conducted 10-fold cross-validation, where participants are randomly divided into 10 folds. The results of this process were highly similar to the results of leave-one-out cross-validation. We opted to report leave-one-out cross-validation because, unlike 10-fold, it is not subject to issues of randomization as every participant becomes a fold.

Table 3

Model Comparison Results for Linear Mixed Effects Regression Models Fit on Reaction Time Data (Full Dataset)

Model	df	LogLik	Deviance	χ^2	Chi df	Pr(> χ^2)	AIC	BIC
1. Trial type only	5	-127371.6	254743.1				254753.1	254792.6
2a. Trial type by L2 AoA and exposure	11	-127366.1	254732.1	11.012288	6	0.0597755	254754.1	254840.9
3. Trial type by L2 AoA, exposure, entropy	17	-127359.1	254718.2	13.877692	6	0.0384592	254752.2	254886.4
4. All three-ways with L2 AoA	26	-127356.1	254712.2	6.087412	9	0.8551724	254764.2	254969.4
1. Trial type only	5	-127371.6	254743.1				254753.1	254792.6
2b. Trial type by entropy	11	-127362.0	254724.0	19.095428	6	0.0039103	254746.0	254832.9
3. Trial type by L2 AoA, exposure, entropy	17	-127359.1	254718.2	5.794552	6	0.3958833	254752.2	254886.4
4. All three-ways with L2 AoA	26	-127356.1	254712.2	6.087412	9	0.8551724	254764.2	254969.4

Note. The top half of the table illustrates comparisons with entropy components added after L2 exposure, and the bottom half illustrates comparisons with exposure added after entropy components. Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are reported for each model, providing an indirect estimate of a model's predictive performance on new data. Bolding illustrates the best fitting models for each metric. Likelihood ratio tests (LRTs) and AIC agree that the best model contained interactions between trial type and the entropy components. BIC prefers the model with trial type only.

suggesting that individuals with high general entropy engaged in a proactive strategy to a greater extent than individuals with low general entropy. These results were confirmed when we used other metrics for model comparison that estimate predictive performance on unmodeled data (e.g., AIC). These results were also partially confirmed when we estimated predictive performance more directly through cross-validation. However, the cross-validation approach also indicated that a model more parsimonious model without individual differences performed reasonably well compared with the most predictive model.

A similar story emerged when considering LASSO regression. The model with the best predictive performance for AX-CPT RTs included effects and interactions between trial type and effects related to language entropy and L2 exposure. The model also included main effects and joint effects related to L2 AoA. However, a less complex, more parsimonious without this effect performed similarly in terms of predictive accuracy.

Taken together, these results suggest that individual differences in bilingual language experience provide signal when estimating proactive executive control, although models with these features perform similarly to models without these factors (within 1 *SEM*).

Discussion

Our primary goal was to test whether individual differences in bilingual language experience predict engagement of proactive

control by analyzing behavioral performance on the AX-CPT for a large sample of over 400 bilinguals and multilinguals from Montréal, Québec, Canada. We extracted or computed several features that index bilingual language experience on a continuum, including L2 AoA, amount of current L2 exposure, and two latent language entropy components that index diversity of language experience among multilinguals. We then modeled task performance as a function of these features using linear mixed effects regression (Analysis 1) and penalized LASSO regression (Analysis 2). In Analysis 1, we compare theoretically relevant models using LRTs, AIC, and cross-validation, which variously provide estimates of how well the models would predict performance on new, unmodeled data. In Analysis 2, LASSO models essentially perform model selection through the tuning of the penalty parameter via cross-validation.

In our first analysis, LRTs indicated a significant relationship between language entropy and the propensity to engage proactive control. Specifically, high general language entropy was associated with larger differences in RT between AY and BX trials, suggesting that integrated bilinguals who use their two languages in a balanced manner rely more on proactive control than compartmentalized bilinguals with low language entropy. We observed the reverse pattern as a function of work entropy, though this interaction was barely significant. LRTs have been argued to lead to overfit models that will not generalize to new data. Thus, we

Table 4

Leave-One-Out Cross-Validation Performance for Linear Mixed Effects Regression Models Fit on Reaction Time Data

Model	df	Mean RMSE training	Mean RMSE testing	SEM RMSE testing	Mean R2 training	Mean R2 testing
1. Trial type only	5	144.8892275	161.4332774	1.623262248	0.333458411	0.207406183
2a. Trial type by L2 AoA and Exposure	11	144.8473449	161.6123077	1.626639281	0.333842535	0.207198753
2b. Trial type by entropy	11	144.834644.0	161.3616435	1.638838663	0.333957718	0.207376613
3. Trial type by L2 AoA, exposure, entropy	17	144.8129985	161.5420277	1.642476628	0.334156151	0.207057256
4. All three-ways with L2 AoA	26	144.8015603	161.7891097	1.650414930	0.334260165	0.206478092

Note. Cross-validation provides a direct estimate of a model's predictive performance on new data. Bolding illustrates the best fitting models for the root mean square error (RMSE) metric on the hold-out testing sample. The model that minimized RMSE on the test sample (i.e., evidenced best prediction performance) contained interactions between trial type and the entropy components. However, all models displayed similar performance, and indeed model with only trial type was within 1 *SEM* of the minimizing model. Thus, in terms of predictive performance all simpler models performed similarly to most predictive model.

Table 5

Regression Coefficients (Unstandardized B and Standardized β) for the Best Fitting Linear Mixed Effects Regression Model That Included Interactions Between Trial Type and Language Entropy Components (i.e., Model 2b)

Effect	B	95% CI B		d	β	t -value	df	p
		LL	UL					
(Intercept = Trial type: BX)	384.985	377.229	392.742		−0.11	97.283	607.866	0
Trial type: AY	113.831	108.715	118.947	0.69	0.64	43.608	19364.317	0
Trial type: BY	−42.14	−47.041	−37.24	0.26	−0.24	−16.854	19351.923	0
PC: Work entropy	9.54	1.337	17.743	0.06	0.05	2.279	608.076	.023
PC: General entropy	−4.424	−12.618	3.77	0.03	−0.03	−1.058	605.252	.29
Trial type: AY \times PC: Work entropy	−5.449	−10.86	−0.037	0.03	−0.03	−1.974	19361.25	.048
Trial type: BY \times PC: Work entropy	−1.739	−6.924	3.446	0.01	−0.01	−0.657	19351.85	.511
Trial type: AY \times PC: General entropy	9.271	3.889	14.654	0.06	0.05	3.376	19363.649	.001
Trial type: BY \times PC: General entropy	7.448	2.288	12.608	0.05	0.04	2.829	19352.413	.005

Note. Effect sizes (d) were computed by dividing the slope estimate by the square root of the sum of the random effect variances, following (Brysbaert & Stevens, 2018; Westfall, Kenny, & Judd, 2014). Random effects: Participant (variance: 5621, SD : 74.97), residual (variance: 21,427, SD : 146.38). p values were computed using Satterthwaite df .

also estimated the predictive performance of each of the models indirectly through the use of AIC and directly through cross-validation. Both estimates of predictive performance converged upon the same best model as LRTs. Of note, however, a stricter measure of performance from the cross-validation procedure indicated that a model without individual differences (i.e., with trial type alone) performed similarly to the best model (within 1 SEM). This suggests that some skepticism is warranted as to whether effects of individual differences on AX-CPT would truly generalize to a completely novel sample of participants.

Despite potential skepticism over the present results, we highlight the fact that patterns related to general language entropy are consistent with previous work that assesses RT on the AX-CPT for different, smaller sample of bilinguals from Montréal (Gullifer et al., 2018). There, we found that increases in entropy (and corresponding increases in corresponding functional connectivity with the anterior cingulate cortex) were associated with increased reliance on proactive control, mirroring the results here. To-

gether, these results suggest that bilinguals with high general language entropy may have adapted their cognitive systems to place greater reliance on active goal maintenance to manage conflict, in line with predictions of the adaptive control hypothesis (Abutalebi & Green, 2016; Green & Abutalebi, 2013; Green & Wei, 2014). High relative to low entropy bilinguals are more likely to participate in diverse communicative contexts that strongly activate all of the languages they know, but where the language demands may shift at any moment depending on the interlocutor. Thus, for high entropy bilinguals, attention to goal relevant information (e.g., knowledge about who speaks what languages or attention to ambient cues about which language will come next) may be crucially important to help manage cross-language conflict.

We also found evidence that high work entropy was associated with decreased reliance on goal maintenance and proactive control, and instead reliance on reactive control. However, this relationship

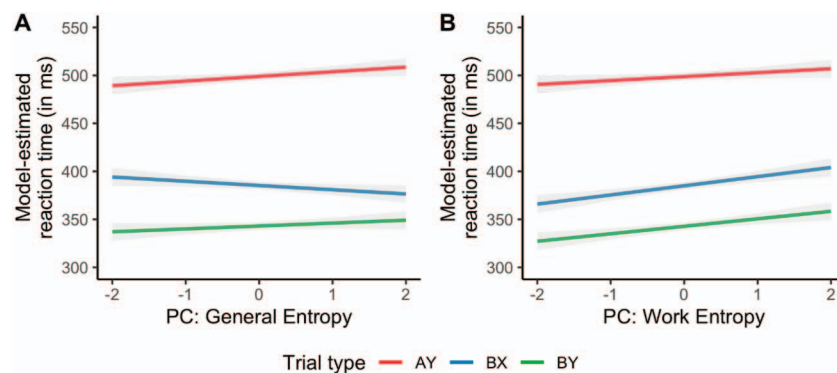


Figure 2. Illustration of the linear mixed effects regression model-estimated effects for the best fitting linear mixed effects regression model that included interactions between trial type and language entropy components (i.e., Model 2b). The plot illustrates AX continuous performance task (AX-CPT) reaction time (RT) as a function of language entropy components: general entropy (A) and work entropy (B). An increase in general entropy was associated with a divergence between the AY and BX trial types, consistent with a greater engagement of proactive control. An increase in work entropy was associated with a general increase in RT. Error bands illustrate 1 SEM . See the online article for the color version of this figure.

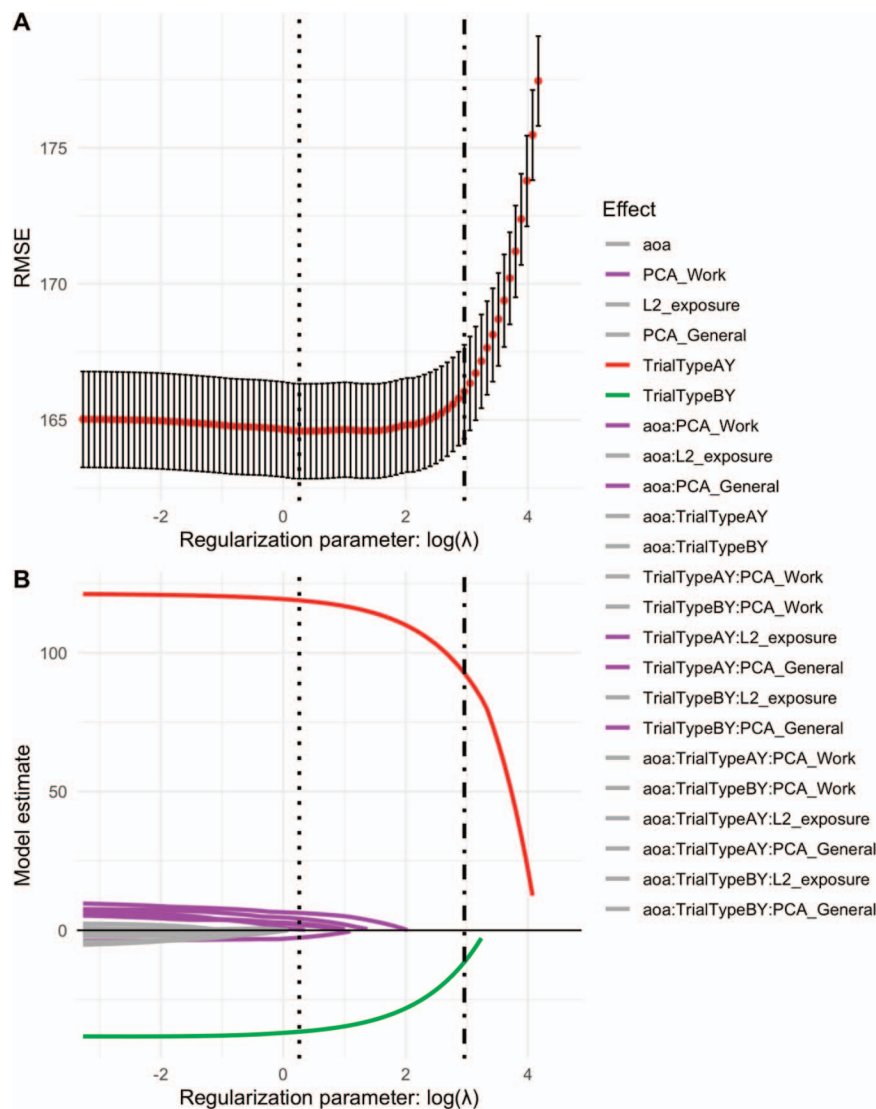


Figure 3. Illustration the LASSO regression results for AX continuous performance task (AX-CPT) reaction time (RT) using leave-one-out cross-validation. (A) Illustrates prediction error (root mean square error, RMSE) as a function of the regularization parameter. The left-most dotted vertical line illustrates the value that minimized RMSE, and the right-most dotted line illustrates the parameter resulting in the simplest model within 1 *SEM* of the minimizing model (i.e., the most parsimonious, predictive model). Effects that survive the minimizing value of λ are colored in purple (i.e., a main effect of work entropy). Error bands illustrate 1 *SEM*. (B) Illustrates model estimates (unstandardized B) as a function of the penalizing parameter. As the value of the parameter increases (left-to-right) effects are regularized to zero. Dotted lines again indicate minimizing value of λ and the value leading to the most parsimonious, predictive model. The model that minimized RMSE contained several effects and interactions related to individual differences in bilingual language experience, but the most parsimonious model contained only effects of trial type. See the online article for the color version of this figure.

was weak; it was barely significant at $p = .048$. This pattern of results is opposite of what would be predicted by adaptive control model. For these reasons, we do not wish to speculate on the origin of this pattern, but we only point out only that the work environment is somewhat unique in Montréal. Despite being a highly multilingual city, language use in most workplaces is mandated by the provincial government: people working in customer-facing jobs are mandated to use French (at least initially before switching to another language). We also point

out that a number of participants conflated work and school contexts when answering the questionnaire. As such, the true pattern of results related to language use in the work environment may be clouded by these issues. Future work and questionnaires should more clearly distinguish between the school and work communicative contexts, at least for samples that include students.

The results of the traditional analysis related to L2 AoA are inconsistent with our previous work (Gullifer et al., 2018). In that study, we

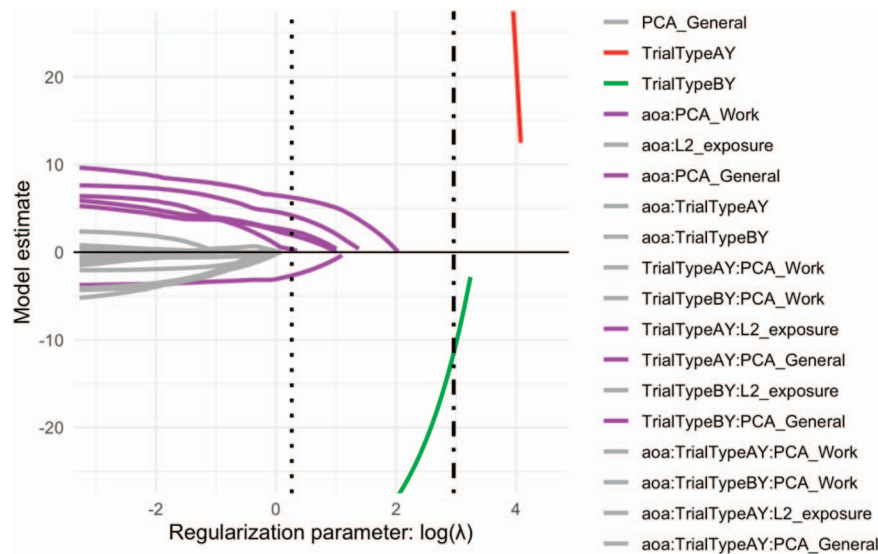


Figure 4. A zoomed-in reproduction of Figure 3B. This figure illustrates model estimates (unstandardized B) as a function of the penalizing parameter. See the online article for the color version of this figure.

found that participants with late AoA (and corresponding decreases in between bilateral inferior frontal gyri) also had increased engagement of proactive control. In contrast here we found no evidence for interaction effects between trial type and L2 AoA in RT.

Thus, it would seem that there is convergent evidence across studies that high general language entropy is associated with increased engagement of proactive control, but lack of convergence regarding L2 AoA. Effects of L2 AoA on executive control are not always consistently observed. For example, some studies show that early L2 AoA results in better executive control perfor-

mance (Kousaie et al., 2017; Luk et al., 2011), others show that late L2 AoA results in better performance (Tao et al., 2011), and others no effect of L2 AoA (Pelham & Abrams, 2014). Going forward, the best approach may be to attempt to account for historical language experiences together with current language experiences jointly using multiple regression. This approach may help build a clearer picture of the associations between various types of language experience and behavior.

Although individual differences in bilingual language experience improved estimates proactive control as measured by AX-CPT in a

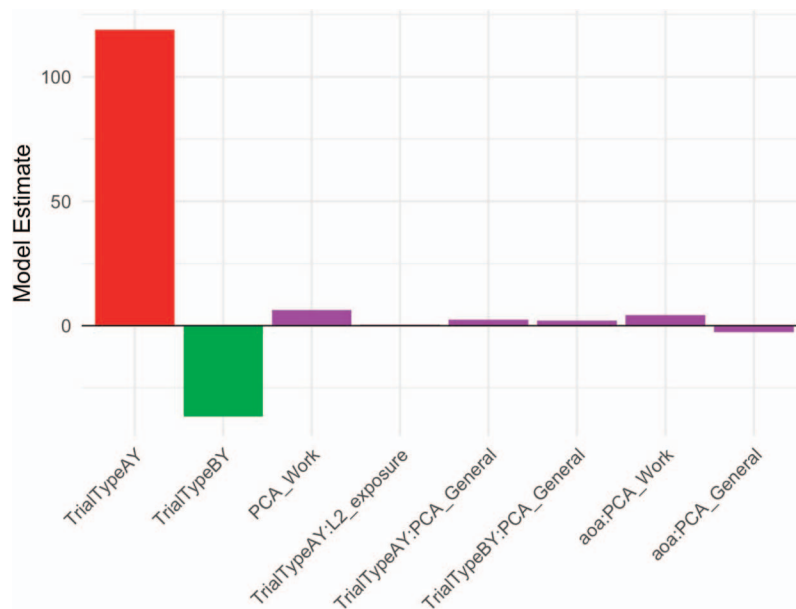


Figure 5. Illustration of magnitude of the model estimates (unstandardized B) for cross-validated LASSO regression at the minimizing value of λ . See the online article for the color version of this figure.

Table 6
Regression Coefficients (Unstandardized B and Standardized β) for the LASSO Regression

Effect	Minimizing model		1 SEM parsimonious model	
	B	β	B	β
(Intercept)	380.82	-0.13	380.28	-0.14
Trial type: AY	118.89	0.67	92.67	0.52
Trial type: BY	-36.56	-0.21	-11.51	-0.06
PC: Work entropy	6.24	0.04		
Trial type: AY \times L2 exposure	0.31	0.00		
Trial type: AY \times PC: General entropy	2.43	0.01		
Trial type: BY \times PC: General entropy	1.95	0.01		
L2 AoA \times PC: Work entropy	4.20	0.02		
L2 AoA \times PC: General entropy	-2.66	-0.02		

Note. Estimates are included for both the minimizing model and the more parsimonious 1 SEM model.

manner consistent with theoretical models and previous empirical work, there was not consistent evidence that individual differences were able to predict novel, unmodeled data on this task. The results of cross-validated regressions showed that although the models that minimized prediction error did include features related to bilingual language experience, simpler models with only effects of trial type performed comparably well (within 1 SEM). In other words, features related to bilingual experience did not substantially improve the predictive accuracy of models, at least for this sample of young adult bilinguals who performed this specific executive control task. There are several potential explanations for these paradoxical results.

One potential argument for the lack of predictive power on the AX-CPT could be that our indices of bilingual language experience are poor measures that are not sufficient to cross-validate. However, we believe this explanation is unlikely for several reasons. In our previous work on a comparable sample of bilinguals, we showed that the same features related to bilingual experience were associated with self-report language outcome measures, including self-report L2 accentedness and L2 abilities (Gullifer & Titone, 2019a). We reanalyzed these results (see online supplemental materials) using the same methods in this study (i.e., cross-validated LASSO), and we found that these features could indeed predict the outcome measures, even in the more conservative parsimonious machine learning models. The nature of the effects identified by those models were highly similar to those identified by our traditional analysis of that data (Gullifer & Titone, 2019a). Thus, in our view, the indicators of bilingual language experience are sound, though further validation is always necessary.

At the same time, we acknowledge there are other unmeasured features and dependent measures that could plausibly enhance predictive accuracy of the machine learning models. For example, whereas L2 AoA provides a coarse measure of an individual's duration being bilingual, and language entropy provides information about the extent of language usage in communicative contexts, we did not have access to information about the duration of use in these contexts. Nor did we have access to information about participants' code-switching behavior. All of these factors are predicted to be important drivers of neurocognitive plasticity (DeLuca et al., 2019; Green & Wei, 2014; Li, Legault, & Litcofsky, 2014). Moreover, we did not have access to neural data for these participants, and it is possible that neural changes may be the most sensitive measures neurocognitive changes.

A second possibility for the lack of predictive power on the AX-CPT is that this version of the task is not sufficiently sensitive to make good predictions related to individual difference on new data. The results of cross-validation indicated that trial type was by far the most important variable to retain in the model. This is important because it suggests the task manipulation consistently affects behavior in a predictable manner. At the same time, the magnitude of the prediction errors (e.g., approximately 75 ms for RT) is quite high relative to differences between trial types (e.g., approximately 86 ms for the AY-BX contrast), suggesting high variability in task performance. Given that the effect sizes of trial type by individual difference interactions were quite small when they were observed in the mixed effects regression model (unstandardized effect sizes: were less than 10 ms; standardized effect sizes were between 0.01 and 0.06), it may be difficult for any model to make accurate predictions of this manner. Moreover, when models are further penalized by way of a regularization parameter (as in LASSO models) and through application of the stringent 1 SEM rule, it is unsurprising that these small effects are completely removed. It is possible that a more demanding or otherwise robust version of the AX-CPT (Morales et al., 2013) might be more sensitive in eliciting difference between trial types related to proactive executive control, allowing for better predictions as a function of individual differences.

A third possible explanation for the lack of predictive power with AX-CPT is that aspects of bilingual language experience do not modulate proactive control as measured by the AX-CPT, at least for this sample of younger adult bilinguals. This explanation is consistent with some of the variable results between studies and between analyses in this paper that include individual differences. For example, we previously mentioned the lack of agreement between the traditional approach here and the results of Gullifer et al. (2018) in terms of whether L2 AoA interacts with trial type on the AX-CPT. Additionally, there is also some lack of agreement between various models reported in this paper as far as the which effects are identified important. While the core interactions between language entropy and trial type (e.g., AY-BX contrast) in the traditional analysis of RT were identified by both minimizing cross-validated mixed effects regression models and LASSO models, there were interactions and main effects that were not consistently identified by the LASSO and mixed effects regression models (e.g., interaction between trial type and

work entropy). High variance in model estimates when models are fit to different subsets of data is a hallmark of overfitting, and a key benefit of the machine approaches is that they acknowledge that overfitting (and underfitting) are possible and attempt find a balance between model simplicity and model complexity (James et al., 2013; Yarkoni & Westfall, 2017).

However, making this assumption would invalidate several consistencies identified between studies (e.g., related to general language entropy) and theoretical perspectives. For example, we would have to assume that the consistencies in (a) traditional models here, (b) the minimizing cross-validated models here, (c) the traditional models reported in Gullifer et al. (2018), and (d) in the neural models reported in Gullifer et al. (2018) are all simply because of overfitting noise in the data. This assumption would also invalidate several predictions of the adaptive control hypothesis, the corresponding evidence that was used to construct that theoretical perspective, and the various empirical tests of this perspective that have emerged since its publication. In our view, it would be a grave scientific error to completely disregard published evidence and theoretical perspectives to simply conclude on the basis of a single study that all previously reported findings are wrong. Thus, we offer a final alternative: that the adaptive effects of language experience on broader cognitive processes may exist, and if so, are very small and, therefore, require an even larger sample size with increasingly validated measures to consistently identify them.

To conclude, we conducted analyses on a large sample of bilingual and multilingual speakers immersed in a highly diverse language environment, a situation that might be predicted to yield robust evidence for an association between language experience and executive control abilities. These analyses yielded somewhat paradoxical results. On the one hand, features related to bilingual experience seemingly provide sufficient signal to estimate patterns of individual variation in the AX-CPT in ways that are theoretically interesting and consistent with previous studies. For example, we replicated a pattern of results in which bilinguals who integrate the usage of multiple languages across various domains displayed greater reliance on proactive executive control processes. On the other hand, these features are not uniquely predictive of performance on the task for novel participants, likely because of the small effect size. Moreover, there are some patterns that are inconsistent with previous studies and theoretical accounts. Thus, the best interpretation of these data we can offer is that relationships between bilingual language experience and proactive executive control, to the extent that they exist, are small. We likely would require a larger sample size together with valid estimates of bilingual language experience to uncover predictive effects related to these features of bilingual experience. Future studies with large sample sizes should absolutely ensure that they use validated measures of bilingual language experience. Otherwise, potential patterns may remain obscured because of high variability within samples or groups.

Crucially, the import of this investigation is not limited to the domain of bilingualism and executive control. Rather, but the methods used here could similarly help to address empirical controversies in other domains. Thus, we agree with Yarkoni and Westfall (2017) that novel analytic methods that assess prediction should be used throughout the behavioral and neural sciences when the sample size is sufficiently large. Such approaches may better help us to identify the signal in the noise.

Research Context

Bilingual language experience is thought to impact domain general executive control abilities. However, the precise quantification of the relationship between these two constructs is not well understood. We believe this has occurred for at least three reasons. First, the field has not converged on a set of best practices for measuring the diverse array of bilingual language experiences, instead individuals are sorted into groups of bilinguals and monolinguals, when in fact the individuals in these groups may be quite heterogeneous in terms of their language experiences. Second, many prior studies have ignored a crucial “mode” of executive control thought to be strengthened by bilingual language experience: on one’s ability to maintain and proactively apply goal-relevant information to preempt responses. Finally, there have been suggestions that prior findings on bilingualism and executive control are because of publication bias, questionable research practices, or reliance on small sample sizes. Here we use sophisticated, continuous measures of bilingual language experience as predictors of proactive executive control within a group of over 400 bilinguals. Our analysis is highly rigorous as it involves metrics that estimate model performance for unmodeled subjects. We find that there is indeed signal related to estimating proactive executive control, however, these effects are small.

References

- Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, 19, 689–698. <http://dx.doi.org/10.1017/S1366728916000225>
- Adler, R. M., Valdes Kroff, J. R., & Novick, J. M. (2020). Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46, 741–759. <http://dx.doi.org/10.1037/xlm0000755>
- Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, 50, 250–263. <http://dx.doi.org/10.3758/s13428-017-0867-9>
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511801686>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <http://dx.doi.org/10.1016/j.jml.2007.12.005>
- Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia*, 35, 1373–1380. [http://dx.doi.org/10.1016/S0028-3932\(97\)00072-9](http://dx.doi.org/10.1016/S0028-3932(97)00072-9)
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, 35, 857–894. <http://dx.doi.org/10.1017/S0142716414000174>
- 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, 1022–1047. <http://dx.doi.org/10.1037/xlm0000770>
- Berry, G. M. (2016). Processing linguistic variation through dual mechanisms of cognitive control. *Linguistics Vanguard: Multimodal Online Journal*. Advance online publication. <http://dx.doi.org/10.1515/lingvan-2016-0033>
- Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. *Developmental Psychology*, 28, 654–664. <http://dx.doi.org/10.1037/0012-1649.28.4.654>
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143, 233–262. <http://dx.doi.org/10.1037/bul0000099>
- Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19, 290–303. <http://dx.doi.org/10.1037/0882-7974.19.2.290>
- Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, 16, 240–250. <http://dx.doi.org/10.1016/j.tics.2012.03.001>
- Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, 7, 325–339. <http://dx.doi.org/10.1111/j.1467-7687.2004.00351.x>
- Bialystok, E., Martin, M. M., & Viswanathan, M. (2016). Bilingualism across the lifespan: The rise and fall of inhibitory control. *The International Journal of Bilingualism*, 9, 103–119. <http://dx.doi.org/10.1177/13670069050090010701>
- Bice, K., & Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language*, 196, 104644. <http://dx.doi.org/10.1016/j.bandl.2019.104644>
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16, 106–113. <http://dx.doi.org/10.1016/j.tics.2011.12.010>
- Braver, T. S., & Barch, D. M. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neuroscience and Biobehavioral Reviews*, 26, 809–817. [http://dx.doi.org/10.1016/S0149-7634\(02\)00067-2](http://dx.doi.org/10.1016/S0149-7634(02)00067-2)
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 76–106). New York, NY: Oxford University Press.
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1, 9. <http://dx.doi.org/10.5334/joc.10>
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280, 747–749. <http://dx.doi.org/10.1126/science.280.5364.747>
- Chen, S. X., Benet-Martínez, V., & Harris Bond, M. (2008). Bicultural identity, bilingualism, and psychological adjustment in multicultural societies: Immigration-based and globalization-based acculturation. *Journal of Personality*, 76, 803–838. <http://dx.doi.org/10.1111/j.1467-6494.2008.00505.x>
- Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, 106, 59–86. <http://dx.doi.org/10.1016/j.cognition.2006.12.013>
- Cummings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28, 369–382. <http://dx.doi.org/10.1177/0267658312443651>
- Dash, T., Berroir, P., Ghazi-Saidi, L., Adrover-Roig, D., & Ansaldo, A. I. (2019). A new look at the question of the bilingual advantage: Dual mechanisms of cognitive control. *Linguistic Approaches to Bilingualism*. Advance online publication. <http://dx.doi.org/10.1075/lab.18036.das>
- de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Science*, 9, 33. <http://dx.doi.org/10.3390/bs9030033>
- de Bruin, A., & Della Sala, S. (2019). The bilingual advantage debate. In J. W. Schwieter (Ed.), *The handbook of the neuroscience of multilingualism* (pp. 736–753). New York, NY: Wiley, Ltd. <http://dx.doi.org/10.1002/9781119387725.ch35>
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, 26, 99–107. <http://dx.doi.org/10.1177/0956797614557866>
- Declerck, M., Koch, I., Duñabeitia, J. A., Grainger, J., & Stephan, D. N. (2019). What absent switch 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. <http://dx.doi.org/10.1037/xhp0000627>
- DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 7565–7574. <http://dx.doi.org/10.1073/pnas.1811513116>
- Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, 18, 34–48. <http://dx.doi.org/10.1016/j.dcn.2015.11.005>
- Donnelly, S., Brooks, P. J., & Homer, B. D. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review*, 26, 1122–1147. <http://dx.doi.org/10.3758/s13423-019-01567-z>
- Doucerein, M. M., Varnaamkhaasti, R. S., Segalowitz, N., & Ryder, A. G. (2015). Second language social networks and communication-related acculturative stress: The role of interconnectedness. *Frontiers in Psychology*, 6, 1111. <http://dx.doi.org/10.3389/fpsyg.2015.01111>
- Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, 337, 867–872. [http://dx.doi.org/10.1016/0140-6736\(91\)90201-Y](http://dx.doi.org/10.1016/0140-6736(91)90201-Y)
- Eleta, I., & Golbeck, J. (2014). Multilingual use of twitter: Social networks at the language frontier. *Computers in Human Behavior*, 41, 424–432. <http://dx.doi.org/10.1016/j.chb.2014.05.005>
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97, 3125–3134. <http://dx.doi.org/10.1121/1.413041>
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Series in Statistics. New York, NY: Springer.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33, 1–22. <http://dx.doi.org/10.18637/jss.v033.i01>
- Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9, 641–651. <http://dx.doi.org/10.1177/1745691614551642>
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25, 515–530. <http://dx.doi.org/10.1080/20445911.2013.796377>
- Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience*, 29, 499–511. <http://dx.doi.org/10.1080/23273798.2014.882515>
- Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L., & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, 163, 42–55. <http://dx.doi.org/10.1016/j.cognition.2017.02.010>
- Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain con-

- nectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia*, 117, 123–134. <http://dx.doi.org/10.1016/j.neuropsychologia.2018.04.037>
- Gullifer, J. W., & Titone, D. (2018). Compute language entropy with {languageEntropy}. Retrieved from <https://github.com/jasongullifer/languageEntropy>
- Gullifer, J. W., & Titone, D. (2019a). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, 23, 283–294. <http://dx.doi.org/10.1017/s1366728919000026>
- Gullifer, J. W., & Titone, D. (2019b). 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. <http://dx.doi.org/10.1037/xlm0000695>
- Han, W.-J. (2010). Bilingualism and socioemotional well-being. *Children and Youth Services Review*, 32, 720–731. <http://dx.doi.org/10.1016/j.childyouth.2010.01.009>
- 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. <http://dx.doi.org/10.1016/j.cognition.2016.01.016>
- Heller, M. (1992). The politics of codeswitching and language choice. *Journal of Multilingual and Multicultural Development*, 13, 123–142. <http://dx.doi.org/10.1080/01434632.1992.9994487>
- Hofweber, J., Marinis, T., & Treffers-Daller, J. (2016). Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism*, 6, 648–668. <http://dx.doi.org/10.1075/lab.15052.hof>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59, 434–446. <http://dx.doi.org/10.1016/j.jml.2007.11.007>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York, NY: Springer. <http://dx.doi.org/10.1007/978-1-4614-7138-7>
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Jylkkä, J., Soveri, A., Wahlström, J., Lehtonen, M., Rodríguez-Fornells, A., & Laine, M. (2017). Relationship between language switching experience and executive functions in bilinguals: An Internet-based study. *Journal of Cognitive Psychology*, 29, 404–419. <http://dx.doi.org/10.1080/20445911.2017.1282489>
- Kam, J. W., Dominelli, R., & Carlson, S. R. (2012). Differential relationships between sub-traits of BIS-11 impulsivity and executive processes: An ERP study. *International Journal of Psychophysiology*, 85, 174–187. <http://dx.doi.org/10.1016/j.ijpsycho.2012.05.006>
- Kousaie, S., Chai, X. J., Sander, K. M., & Klein, D. (2017). Simultaneous learning of two languages from birth positively impacts intrinsic functional connectivity and cognitive control. *Brain and Cognition*, 117, 49–56. <http://dx.doi.org/10.1016/j.bandc.2017.06.003>
- Krein, J. L., MacLean, A. C., Delorey, D. P., Knutson, C. D., & Eggett, D. L. (2009, June). *Language entropy: A metric for characterization of author programming language distribution*. Paper presented at the 4th Workshop on Public Data about Software Development.
- Kroll, J. F., Gullifer, J. W., & Zirnstein, M. (2016). Literacy in adulthood: Reading in two languages. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 225–245). Washington, DC: American Psychological Association.
- 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. <http://dx.doi.org/10.18637/jss.v082.i13>
- Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language*, 92, 217–233. <http://dx.doi.org/10.1016/j.jml.2016.06.010>
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, 144, 394–425. <http://dx.doi.org/10.1037/bul0000142>
- Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, 58, 301–324. <http://dx.doi.org/10.1016/j.cortex.2014.05.001>
- Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, 17, 673–680. <http://dx.doi.org/10.1017/S1366728913000606>
- Ličen, M., Hartmann, F., Repovš, G., & Slapničar, S. (2016). The impact of social pressure and monetary incentive on cognitive control. *Frontiers in Psychology*, 7, 93. <http://dx.doi.org/10.3389/fpsyg.2016.00093>
- Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: Behavior, brain activation, and individual differences. *Cognitive, Affective & Behavioral Neuroscience*, 8, 99–112. <http://dx.doi.org/10.3758/CABN.8.1.99>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25, 605–621. <http://dx.doi.org/10.1080/20445911.2013.795574>
- Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, 14, 588–595. <http://dx.doi.org/10.1017/S1366728911000010>
- Ma, F., Li, S., & Guo, T. (2016). Reactive and proactive control in bilingual word production: An investigation of influential factors. *Journal of Memory and Language*, 86, 35–59. <http://dx.doi.org/10.1016/j.jml.2015.08.004>
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40. <http://dx.doi.org/10.1037/1082-989X.7.1.19>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940–967. [http://dx.doi.org/10.1044/1092-4388\(2007\)067](http://dx.doi.org/10.1044/1092-4388(2007)067)
- McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. New York, NY: CRC Press. <http://dx.doi.org/10.1201/9781315372495>
- Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, 25, 531–546. <http://dx.doi.org/10.1080/20445911.2013.807812>
- Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia*, 66, 157–169. <http://dx.doi.org/10.1016/j.neuropsychologia.2014.11.014>
- Navarro-Torres, C. A., Garcia, D. L., Chidambaram, V., & Kroll, J. F. (2019). Cognitive control facilitates attentional disengagement during second language comprehension. *Brain Sciences*, 9, 1–23. <http://dx.doi.org/10.3390/brainsci9050095>
- Nichols, E. S., Wild, C. J., Stojanoski, B., Battista, M. E., & Owen, A. M. (2020). Bilingualism affords no general cognitive advantages: A population study of executive function in 11,000 people. *Psychological Science*, 31, 548–567. <http://dx.doi.org/10.1177/0956797620903113>
- Nuzzo, R. (2014). Scientific method: Statistical errors. *NATNews*, 506, 150–152. <http://dx.doi.org/10.1038/506150a>
- Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual ad-

- vantages in executive functioning. *Journal of Cognitive Psychology*, 26, 242–255. <http://dx.doi.org/10.1080/20445911.2014.891597>
- Paap, K. (2019). The bilingual advantage debate. In J. W. Schwieter & M. Paradis (Eds.), *The handbook of the neuroscience of multilingualism*. Advance online publication. <http://dx.doi.org/10.1002/9781119387725.ch34>
- Paap, K. R., Anders-Jefferson, R., Mikulinsky, R., Masuda, S., & Mason, L. (2019). On the encapsulation of bilingual language control. *Journal of Memory and Language*, 105, 76–92. <http://dx.doi.org/10.1016/j.jml.2018.12.001>
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265–278. <http://dx.doi.org/10.1016/j.cortex.2015.04.014>
- Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs*, 76, 1–23. <http://dx.doi.org/10.1037/h0093840>
- Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 313–325. <http://dx.doi.org/10.1037/a0035224>
- Pinheiro, J. C., & Bates, D. (2009). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.
- Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 787–796. <http://dx.doi.org/10.1037/a0035583>
- Poarch, G. J., Vanhove, J., & Berthele, R. (2019). The effect of bidialectalism on executive function. *The International Journal of Bilingualism*, 23, 612–628. <http://dx.doi.org/10.1177/1367006918763132>
- 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*, 17, 682–691. <http://dx.doi.org/10.1017/S1355617711000580>
- Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13, 253–262. <http://dx.doi.org/10.1017/S1366728909990526>
- Psychology Software Tools. (2020). E-Prime (Version 1) [Computer software]. Pittsburgh, PA: Author. Retrieved from <https://www.pstnet.com>
- R Core Team. (2019). R: A language and environment for statistical computing. Retrieved from <https://www.R-project.org/>
- Saiz, A., & Zoido, E. (2005). Listening to what the world says: Bilingualism and earnings in the United States. *The Review of Economics and Statistics*, 87, 523–538. <http://dx.doi.org/10.1162/0034653054638256>
- Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context. A test of a theoretical model. *Archives of General Psychiatry*, 53, 1105–1112. <http://dx.doi.org/10.1001/archpsyc.1996.01830120037008>
- Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group Analysis Approach. *Frontiers in Psychology*, 2, 183. <http://dx.doi.org/10.3389/fpsyg.2011.00183>
- Subramaniapillai, S., Rajah, M. N., Pasvanis, S., & Titone, D. (2019). Bilingual experience and executive control over the adult lifespan: The role of biological sex. *Bilingualism-Language and Cognition*, 22, 733–751. <http://dx.doi.org/10.1017/S1366728918000317>
- Surraín, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, 22, 401–415. <http://dx.doi.org/10.1017/S1366728917000682>
- Takahesu Tabori, A. A., Mech, E. N., & Atagi, N. (2018). Exploiting language variation to better understand the cognitive consequences of bilingualism. *Frontiers in Psychology*, 9, 1686. <http://dx.doi.org/10.3389/fpsyg.2018.01686>
- Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology*, 2, 123. <http://dx.doi.org/10.3389/fpsyg.2011.00123>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B. Methodological*, 58, 267–288. <http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Titone, D., Gullifer, J. W., Subramaniapillai, S., Rajah, N., & Baum, S. (2017). History-inspired reflections on the bilingual advantages hypothesis. In E. Bialystok & M. D. Sullivan (Eds.), *Growing old with two languages: Effects of bilingualism on cognitive aging* (Vol. 53, p. 265). Amsterdam, the Netherlands: John Benjamins Publishing.
- Tiv, M., Gullifer, J., Feng, R., & Titone, D. (2020). Using network science to map what Montréal bilinguals talk about across languages and communicative contexts. *Journal of Neurolinguistics*, 56, 100913. <http://dx.doi.org/10.1016/j.jneuroling.2020.100913>
- Verreyt, N., Woumans, E. V. Y., Vandelandotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, 19, 181–190. <http://dx.doi.org/10.1017/S1366728914000352>
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143, 2020–2045. <http://dx.doi.org/10.1037/xge0000014>
- Whitford, V., Pivneva, I., & Titone, D. (2016). Eye movement methods to investigate bilingual reading. In R. Heredia, J. Altarriba, & A. Cieślicka (Eds.), *Methods in bilingual reading comprehension research* (pp. 183–211). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4939-2993-1_8
- Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, 19, 73–80. <http://dx.doi.org/10.3758/s13423-011-0179-5>
- Whitford, V., & Titone, D. (2015). Second-language experience modulates eye movements during first- and second-language sentence reading: Evidence from a gaze-contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 1118–1129. <http://dx.doi.org/10.1037/xlm0000093>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12, 1100–1122. <http://dx.doi.org/10.1177/1745691617693393>
- Zhang, H., Kang, C., Wu, Y., Ma, F., & Guo, T. (2015). Improving proactive control with training on language switching in bilinguals. *NeuroReport*, 26, 354–359. <http://dx.doi.org/10.1097/WNR.0000000000000353>
- Zirnshtein, M., van Hell, J. G., & Kroll, J. F. (2018a). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, 176, 87–106. <http://dx.doi.org/10.1016/j.cognition.2018.03.001>
- Zirnshtein, M., van Hell, J. G., & Kroll, J. F. (2018b). Cognitive control and language ability contribute to online reading comprehension: Implications for older adult bilinguals. *The International Journal of Bilingualism*, 23, 971–985. <http://dx.doi.org/10.1177/1367006918763143>

Received August 7, 2019

Revision received May 14, 2020

Accepted May 21, 2020 ■