Comparing Inhibitory Control in Monolingual and Bilingual Sentence Processing

A Thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
the requirements for
the Degree

Master of Arts

In
Psychology: Mind, Brain, and Behavior

by

Yunyun Liu

San Francisco, California

August 2015
CERTIFICATION OF APPROVAL

I certify that I have read Comparing Inhibitory Control in Monolingual and Bilingual Sentence Processing by Yunyun Liu, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Master of Arts: Psychology at San Francisco State University.

Mark W. Geisler, Ph.D.
Professor of Psychology

Ken Paap, Ph.D.
Professor of Psychology
Monolinguals outperform bilinguals in a sentence acceptability judgment task, but not a grammaticality judgment task. A recent study concludes that this pattern of results is evidence for a bilingual advantage in conflict resolution during sentence processing. This study attempts to replicate the previous study with crucial methodological changes, and finds no evidence for a bilingual advantage, yet does offer an alternative interpretation of the previous study's results. Finally, the authors discuss the importance of matching bilingual and monolingual participants on a variety of measures and the effects of confirmation bias on bilingual research.
# TABLE OF CONTENTS

List of Tables....................................................................................................................... v
List of Figures .................................................................................................................... vi
List of Appendices ............................................................................................................ vii
Introduction ....................................................................................................................... 1
Method............................................................................................................................. 11
  Participants.................................................................................................................. 11
  Materials.................................................................................................................... 11
  Procedure................................................................................................................... 14
Results .............................................................................................................................. 14
  Accuracy (Percent Correct) .................................................................................. 15
  Reaction Time ........................................................................................................ 16
  Reading Time ........................................................................................................... 17
Discussion......................................................................................................................... 18
  Difficulties of Bilingual Research ........................................................................ 22
  Problem of Confirmation Bias ............................................................................... 27
  Conclusion ............................................................................................................... 29
Reference ......................................................................................................................... 29
Appendices ....................................................................................................................... 36
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Table 1</td>
<td>12</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figures</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Figure 1</td>
<td>19</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Self-rating Instructions</td>
<td>36</td>
</tr>
<tr>
<td>2. IRB Consent Form</td>
<td>37</td>
</tr>
</tbody>
</table>
Life experiences can change the form and function of the brain - this is known as functional neuroplasticity. A brain-changing life experience that affects most of the world's population is bilingualism or multilingualism (Bialystok, Craik, & Luk, 2012; Kroll & Bialystok, 2013). The first studies of the cognitive effects of bilingualism dating from the early twentieth century were poorly designed with insufficient controls and unmatched participant groups (Peal & Lambert, 1962). These early studies generally found disadvantages to bilingualism. More modern studies find that there are a few disadvantages to bilingualism, but also a variety of possible cognitive benefits.

Disadvantages of bilingualism include weaker lexical access and a smaller vocabulary in each language. These disadvantages have been shown repeatedly in experiments and are indicated by weaker category fluency and more tip-of-the-tongue states (Bialystok, 2009; Bialystok et al., 2012). This is explained by the fact that a bilingual uses each of his or her languages less (since the bilingual switches between the two languages), resulting in weaker links between concepts and words.

There are advantages to bilingualism other than the obvious ones of being able to communicate in two languages and most likely being bicultural. There is some evidence from all age groups for cognitive advantages that native bilinguals have over their monolingual peers (See Bialystok (2009) for an overview, see Paap & Greenberg (2013), Paap, Johnson, & Sawi (2015), and Hilchey, Saint-Aubin, & Klein (n.d.) for an alternate
These reported advantages include enhanced ability to inhibit distracting or irrelevant information (Bialystok, Craik, & Luk, 2008), enhanced ability for conflict resolution (Costa, Hernández, & Sebastián-Gallés, 2008), earlier development of metalinguistic awareness (Bialystok, 1986), better control of processing in children (Bialystok & Martin, 2004), and slower onset and less severe symptoms of dementia and mild cognitive impairment in elderly adults. (Bialystok, Craik, & Freedman, 2007; Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok et al., 2012; but see Fuller-Thomson, 2015 for an alternate viewpoint).

These advantages can be explained by the Inhibitory Control Model - a model for the control of the bilingual lexico-semantic system (Green, 1998). According to this model the mental representation bilinguals have of their two languages consists of shared concepts and separate lexicons. Activation of a concept spreads activation to the corresponding word in each lexicon. Since only one word can be produced at a time (in spoken languages), Green assumes that one of these word forms is inhibited in order for the other to be spoken. This allows the bilingual to speak in the language that he or she intends to speak (since the activated but unintended translation equivalent is inhibited).

In bimodal bilinguals (people who know a sign language and a spoken language), it is possible to use both languages simultaneously. Therefore these bilinguals have less practice in inhibition and perform similarly to monolinguals in measures of inhibition.
Emmorey et al. compared bimodal bilinguals, unimodal bilinguals, and monolinguals on performance in the Flanker Task. There were no group differences in accuracy, but unimodal bilinguals responded faster. Most interestingly, there were no differences in performance between bimodal bilinguals and monolinguals. This suggests that the cognitive advantages associated with bilingualism occur not simply because bilinguals have mental representations of two languages, but rather because of competition between translation equivalents for expression and the necessity of inhibiting one of them. However, since Emmorey et al., there have been many failures to replicate the advantage of unimodal bilinguals over monolinguals (e.g., Kousai & Phillips, 2012; Paap & Greenberg, 2013; Paap & Sawi, 2014; Anton et al., 2014.)

Inhibition plays a major role in executive function. According to the Unity/Diversity Framework of executive function (Miyake, 2000; Miyake & Friedman, 2012), inhibition is a key component of executive function in itself and contributes to other separable components of executive function. Miyake et al. divided tasks commonly used to measure executive functioning (e.g. Stroop Task, Color-Shape Task, Letter Memory Task) into three groups: those logically requiring inhibition, those requiring shifting, and those requiring updating. The authors defined inhibition as the ability to "override a dominant or pre-potent response," shifting as the ability to "switch flexibly between tasks or mental sets," and updating as the ability for "constant monitoring and rapid addition/deletion of mental sets." Using the statistical techniques of confirmatory factor
analysis and structural equation modeling, the authors determined that not only is inhibition a core component of executive function but it is also involved in the shifting and updating components. Most reported bilingual advantages do in fact appear in tasks that involve control of processing, conflict resolution, and inhibition, rather than tasks involving representation and working memory (Bialystok & Martin, 2004; Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012).

Researchers can assess individuals' abilities for control of processing, conflict resolution, and inhibition using a set of sentence judgment tasks known as the Acceptability Task and the Grammaticality Task. In the Acceptability Task the participant should respond "no" if there is anything wrong with the sentence, that is, if there is either a syntax violation or something that doesn't make sense. In the Grammaticality Task, the participant is instructed to say "no" only when there is a syntax violation and should still respond "yes" to sentences with a semantic anomaly. For example, in Bialystok (1986) children were presented with sentences containing a semantic anomaly, such as "Apples on noses.", that while grammatically correct, had an anomalous meaning. The children were then asked whether sentences such as these were grammatically correct. Bilingual children outperformed monolingual children on this task. This result suggests earlier development of metalinguistic awareness; that is, the bilingual children understand that semantic and syntactic information are different. Moreover, they are able to inhibit
salient semantic information in order to make judgments based on syntactic information alone.

Sentences containing syntactic violation or semantic anomaly have also been used in Event-Related Brain Potential (ERP) studies to establish that syntactic violation and semantic anomaly elicit distinct waveforms (Kutas & Hillyard, 1980; L. Osterhout & Holcomb, 1992). These studies established that a component of the ERP known as N400 is elicited by semantic anomaly (a senseless but grammatically sound sentence) and that another component of the ERP known as P600 is reliably elicited by the presence of a grammatical violation. Researchers speculate that the N400 reflects the effort of attempting to integrate the meaning of a word into the surrounding context. The P600 reflects a process of reanalysis and repair of syntactic structure, as it is also elicited in garden path sentences. In garden path sentences, an initial syntactic analysis must be reevaluated upon reaching a certain word in the sentence that indicates a less commonly used, but still correct, syntactic structure is required.

Moreno, Bialystok, Wodniecka, & Alain (2010) combine the ERP methodology with the Acceptability and Grammaticality Tasks in a study to determine whether the bilingual advantage in these judgment tasks (shown in children in Bialystok (1986)) extend to adults. Given that adults have much greater executive function and metalinguistic awareness than children, the use of ERP’s provide additional data that is more
informative than task performance alone. Additionally the sentences are presented in a timed manner using rapid serial visual presentation (RSVP) to increase the difficulty of the task.

The behavioral results of Moreno et al. (2010) show that while monolinguals outperform (i.e. are more accurate than) bilinguals on the Acceptability Task, monolinguals and bilinguals perform comparably on the Grammaticality Task. The authors interpret their results to support the existence of a bilingual advantage in inhibition (and by extension, executive function) in adults. Crucially, they claim that the Grammaticality Task is more difficult than the Acceptability Task because the Grammaticality Task requires more inhibition. In the Acceptability Task any sort of violation makes a sentence incorrect (i.e. judgments are made based on both grammatical and semantic information). In the Grammaticality Task, semantic information (which the authors suppose to be more salient) must be inhibited and a judgment made solely on grammatical information. On the basis of this assumption, the authors claim that bilinguals do not perform as well on the Acceptability Task because of their general lower proficiency in English (established through performance on the Peabody Picture Vocabulary Test). However bilinguals' accuracy matches that of monolinguals on the Grammaticality Task because of the bigger role inhibition plays in performance of the Grammaticality Task combined with the bilinguals' enhanced ability
for inhibition. Thus, bilinguals are able to perform comparably despite the disadvantage of lower English proficiency.

It is notable that Moreno et al. do not report a Group x Task interaction in their study. A significant Group x Task interaction is important because it would demonstrate that monolinguals and bilinguals differ in performance depending on the task. This would further support the conclusion that monolinguals outperform bilinguals on the Acceptability Task, but bilinguals and monolinguals perform comparably on the Grammaticality Task. Testing for a Group x Task interaction is a more sensitive and appropriate statistical test than the test Moreno et al. chose to do (t-tests for group differences separately for each of the two tasks).

A bilingual advantage which only appears in tasks requiring relatively high levels of inhibition is a pattern that has been seen before. Previous studies have also found bilingual advantages which only occurred in high-conflict conditions, that is, conditions requiring inhibition of distracting or irrelevant information (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Engel de Abreu et al., 2012). Given these previous findings, the fact that bilinguals do relatively better in the high-conflict Grammaticality Task compared to the lower-conflict Acceptability Task is not surprising.

Moreno et al. further claim that the ERP data collected in the study support their interpretation of the behavioral data. Monolinguals and bilinguals showed a N400 of
equal amplitude in the Acceptability Task, but bilinguals showed a higher N400 than monolinguals in the Grammaticality Task. According to the authors, this indicates that bilinguals are more efficient at inhibition. They were able to perform comparably with monolinguals on the task (which involved processing grammatical information) but still had the cognitive resources to process additional semantic information. An alternative interpretation for the N400 differences is presented in the discussion. Bilinguals also showed a smaller amplitude P600 than monolinguals in the Grammaticality Task. This was interpreted to mean that the "repair and reanalysis" process upon meeting with syntactic violation is less effortful and more efficient in bilinguals than in monolinguals. Moreno et al. thus conclude that there is a bilingual advantage in conflict resolution (which consists of the inhibition required to suppress the irrelevant semantic anomaly and to base the judgment only on grammaticality) during sentence processing.

Not only is Moreno et al.'s interpretation of their results questionable, but there are several problems of methodology. One could question the validity of the assumption that the Grammaticality Task is more difficult than the Acceptability Task. There is no experimental data cited to support this nor do the authors examine this hypothesis in their study. The behavioral data from the study seem to indicate that the Grammaticality Task is the easier task. According to Moreno et al., both monolinguals and bilinguals performed near ceiling (greater than 93% accuracy) on the Grammaticality Task. When accuracy is analyzed by sentence type it is seen that both
monolinguals and bilinguals are significantly less likely to answer correctly when the sentence is semantically anomalous (regardless of task). This would seem to suggest that semantically anomalous sentences are particularly difficult to process. The assumption that the Grammaticality Task is difficult because it requires suppression of semantic anomalies is the basis of the Moreno et al.'s entire interpretation of the results and so it is surprising that more care was not taken in establishing its validity.

Another potential problem is the lack of counterbalancing of the Grammaticality Task and the Acceptability Task. These tasks were done by all participants in the same order (Grammaticality first, Acceptability second). This can create undesired transfer effects. Additionally, the same set of stimulus sentences was seen by participants in each task, only with small modifications to allow for counterbalancing of sentence types. For instance during the first task, a participant might read "The sea lions bask on the beach all day." During the second task, the participant would see an altered version such as "The sea lions edit on the beach all day." (a semantically anomalous sentence). The participant might expect to see the same version of the sentence in the second task and performance could be affected by the new (and surprising) form of the sentence. Potentially more problematic, is the possibility of systematically handicapping performance on the second task (Acceptability Task). During the first task, the participants become familiar with the instructions and gain practice on the task. Then in the second task, they are required to learn and follow new instructions. This requires
inhibition of the old set of instructions and inhibition of a practiced response. By choosing not to counterbalance task order, Moreno et al. may have unintentionally made the second task (Acceptability Task) more difficult (require more inhibition). This might explain the overall better performance in the Grammaticality Task.

The current study aims to replicate the behavioral portion of Moreno et al. with key methodological and design changes that address the previously expressed concerns. These changes include self paced presentation of stimulus sentences, recording of response times and reading times, and counterbalancing of task order. Moreno et al. chose not to record response times as a precaution against excessive motor artifacts in the ERP recordings. However response times and reading times, which a self-paced method of presentation allows us to record, should be particularly informative for determining which task and which sentence type is most difficult for participants. Response times and reading times have been used in previous studies to examine processing of different types of sentences (Ditman, Holcomb, & Kuperberg, 2007; Lee Osterhout & Nicol, 1999). We expect longer response times to indicate a more difficult task and longer reading times to indicate greater difficulty of processing a certain type of sentence. In addition to allowing us to collect reading times, self-paced presentation of sentences (in contrast to experimenter controlled presentation of sentences) provides participants with a more natural reading experience and accommodates different reading speeds (Ditman et al., 2007a). Ditman et al. have also shown in the
same study that self-paced presentation does not interfere with ERP recording. The final change to the original experiment is counterbalancing of task order to prevent the previously discussed transfer effects. The introduction of these changes makes for a more rigorous and well-designed, as well as more informative experiment.

**Method**

**Participants**

Participants consisted of twenty-four monolinguals and twenty-four bilinguals recruited from undergraduate psychology courses at San Francisco State University. Each participant first completed a language history and use questionnaire. This questionnaire included a self-rated proficiency portion in which participants rated their proficiency in each of their known languages on a scale of 1 to 7. Refer to the Appendix for a key to the scale used. Previous research has established a high degree of correlation between self-rated proficiency and performance on language proficiency tests such as picture naming (Marian, Blumenfeld, & Kaushanskaya, 2007). These proficiency ratings were used to determine whether a participant was considered a bilingual or a monolingual. Participants with self-ratings of 4 or higher in two or more languages were considered to be bilingual.

**Materials**
Participants completed the Grammaticality Task, Acceptability Task, and a distracter task (Ravens advanced matrices test). The tasks were computer controlled using DirectRT. Stimuli consisted of sentences of four types: 1) correct 2) syntactic violation 3) semantically anomalous 4) both syntactic violation and semantically anomalous. The last type of sentence was used as filler and responses were not analyzed. In the Acceptability Task, participants were instructed to decide whether a sentence was "acceptable". An acceptable sentence was defined as one that did not contain a semantic anomaly or a syntactic violation (i.e. correct). In the Grammaticality Task participants decided whether a sentence was "grammatical". A grammatical sentence was defined as a sentence that did not contain a syntactic violation. The challenge of the Acceptability Task is that it requires ignoring semantically anomalous sentences that are syntactically well-formed. For examples of each type of anomaly, how sentences of each type were created from a set of sentence frames taken from Lee Osterhout & Nicol (1999) and whether each type of sentence was "acceptable" or "grammatical," see Table 1.

Table 1

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Sentence</th>
<th>Acceptable?</th>
<th>Grammatical?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not anomalous (correct)</td>
<td>The sea lions can bask on the beach all day.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Syntactically anomalous | The sea lions can basking on the beach all day. | No | No
Semantically anomalous | The sea lions can edit on the beach all day. | No | Yes
Syntactically and Semantically Anomalous | The sea lions can editing on the beach all day. | No | No

a. An example of a sentence frame taken from Osterhout and Nichols (1999)
b. Table containing the three sentence types created from the sentence frame in A) and the correct responses to each sentence in the Acceptability and Grammaticality tasks.

As in Moreno et al., we used the set of 90 sentences frames from Osterhout and Nichols (1999) to create the stimuli. These sentence frames were divided into three lists of 30. The sentence frames composing each list were used to create the correct, ungrammatical, and semantically anomalous sentences for each participant’s Grammaticality Task and Acceptability Task. The lists were rotated between tasks so participants never saw the exact same version of a sentence more than once. For instance, during the first task, sentence frames in List 1 might be used to create correct sentences, List 2 sentence frames to create ungrammatical sentences, and List 3 sentence frames to create anomalous sentences. Then during Task 2, the lists are rotated so that List 2 is used for correct sentences, List 3 for ungrammatical sentences, and List 1 for anomalous sentences. For the next participants, the lists are rotated again, with List 3 providing the correct sentence in the first task. The purpose of this is to avoid
any item-specific effects. An additional thirty novel filler sentences that were both syntactically and semantically anomalous were presented along with the other three lists. These sentences were presented during each task and never changed – bringing the total number of sentences a participant read in each task to 120.

Procedure

After filling out the questionnaire, participants were seated in front of the computer and given the instructions for the first task. This was followed by a block of 32 practice sentences to make sure instructions were understood. Half of the participants completed the Grammaticality Task first and the other half completed the Acceptability Task first. Sentences were presented word-by-word, followed by an answer prompt. Participants progressed though the sentences by pressing the space bar on the keyboard with their right hand and responded to the answer prompt by either pressing Z to indicate “Yes” or X to indicate “No” with the middle and index finger of their left hand. Participants received performance feedback in the form of a beep from the computer after an incorrect response. Between Task 1 and Task 2, participants completed the Ravens Progressive Matrices. No performance feedback was given for this task.

Results
Analysis

The purpose of our experiment is to examine the effects of language group (bilingual or monolingual), task, sentence type, and task order on accuracy, reaction time, and reading time. In order to do this, a Language Group x Task X Sentence Type x Order mixed factorial ANOVA was conducted for each of the dependent variables.

Accuracy

Each factor had a significant main effect. Across both tasks monolinguals (M = 95.3%) were more accurate than bilinguals (M = 90.3%), F(1, 44) = 20.51, p < .001, partial $\eta^2 = .318$. Participants were more likely to respond incorrectly to sentences containing a semantic anomaly (M = 87.7% correct) than to correct sentences (M = 94.7%) or sentences containing syntactic violations (95.9%), F(2,88) = 27.39, p < .001, partial $\eta^2 = .384$. Participants were also more accurate at the Grammaticality Task (M = 93.8%) than the Acceptability Task (M = 91.7%), F(1, 44) = 4.60, p = .037, partial $\eta^2 = .095$. Finally, participants that did the Grammaticality Task first (M = 94.7%) were more accurate than those who did the tasks in the reverse order (M = 90.8%) (when the acceptability task was done first), F(1, 44) = 9.81, p = .003, partial $\eta^2 = .182$.

A significant Group X Task interaction would support Moreno et al.'s claim that bilinguals catch up to monolinguals on the Grammaticality Task. The Language Group X Task interaction was not significant in the present study.
Reaction Times

There was a significant main effect of Task, $F(1,44) = 7.629$, $p = .008$, partial $\eta^2 = .148$. Participants had faster reaction times in the Grammaticality Task ($M = 629$ ms) than in the acceptability task ($M = 681$ ms). There was also a significant main effect of sentence type $F(2,88) = 50.46$, $p < .001$, $\eta^2 = .534$. Participants responded quickly to syntactic violations ($M = 541$ ms) and responded slower to semantic anomaly ($M = 764$ ms).

There was a significant Task x Sentence interaction involving correct sentences, $F(2, 88) = 12.697$, $p < .001$, partial $\eta^2 = .224$. Participants were faster at responding to correct sentences in the Grammaticality Task than in the Acceptability Task. There was also a significant Task X Order interaction, $F(1, 44) = 23.186$, $p < .001$, partial $\eta^2 = .345$. When the Grammaticality Task comes first, response times on both tasks are about the same. When the Acceptability Task comes first, response times to the Acceptability Task are slower. It appears that the Grammaticality Task acts as a warm-up or practice for the Acceptability Task since doing the Grammaticality Task first improves response times during the Acceptability Task. This is further evidence that the Acceptability Task is more difficult than the Grammaticality Task.

Reading Times

Reading times were adjusted for word length and individual reading speed. For each participant, an average reading time was calculated for words of each length. For
instance, an average reading time for four-letter words, five-letter words and so on is calculated for Participant A. The same is done for participant B and the rest of the participants. These average times were used to calculate an equation for a best-fitting straight line for each individual participant. The best-fitting straight line is then used to predict a reading time for the critical word and sentence-final word. The adjusted reading time used in the data analysis is the result of an observed reading time minus the predicted reading time. Thus, positive adjusted times reflect longer reading times adjusted for both the readers average reading speed and the length of the critical word.

Reading time to critical word

The critical word of each sentence is the word that determines whether a sentences is correct, grammatically incorrect, or semantically anomalous (e.g. in the sentence "The sea lions can basking on the beach all day.", the critical word is "basking").

There was a significant main effect of sentence type, $F(2,88) = 3.62, p = .031$, partial $\eta^2 = .076$. Adjusted reading times to the critical word in sentences containing a syntactic violation were longer (M=+28.6ms) than the adjusted reading times in correct sentences (M=-5.6ms) and semantically anomalous sentences (M=+.2ms).

Reading time to final word
There was a significant main effect of sentence type, \( F(2, 88) = 40.20, p < .001, \text{ partial } \eta^2 = .477 \). Reading time to the final word was faster in sentences containing syntactic violations (\( M = -54 \text{ ms} \)) than in correct sentences (\( M = +123 \text{ ms} \)) and in sentences containing semantic anomalies (\( M = +163 \)). There was a significant Task x Sentence Type interaction involving correct sentences, \( F(2,88) = 6.18, p = .003, \text{ partial } \eta^2 = .123 \). In correct sentences, reading times were longer for the Acceptability Task than for the Grammaticality Task.

Discussion

The results show that the Acceptability Task is more difficult than the Grammaticality Task. Both bilinguals and monolinguals are more accurate on the Grammaticality Task than on the Acceptability Task. The reaction time data supports this conclusion as reaction times are slower on the Acceptability Task than on the Grammaticality Task. The main effect of sentence type in the reaction time data shows that it is semantically anomalous sentences themselves that cause slower response times. As can be seen in Figure 1, regardless of task, participants responded faster to syntactic violations than to semantic anomalies. This indicates that there is something about semantic anomalies that is difficult to process and the difficulties associated with the Acceptability Task are not necessarily task-specific.
Regardless of task, participants responded faster to syntactic violations than to semantic anomalies.

A possible reason for this difficulty in processing is the relative unfamiliarity of semantic anomaly compared to syntactic violations. Students have a lot of experience identifying and correcting syntactic violations in sentences from school. Whereas semantically anomalous sentences are rarer and usually interpreted as metaphor (instead of judged as acceptable or unacceptable). It could be that participants performed better on the Grammaticality Task and responded more quickly to sentences containing syntactic violation because it was a more familiar task.
We did not replicate the finding of Moreno et al. (2010) that bilinguals "catch-up" in performance on the Grammaticality Task. In both tasks, monolinguals were more accurate than bilinguals. In the original study, performance on the Grammaticality task, especially for monolinguals, was very near ceiling (above 95% accuracy). It is possible that this might have obscured the data that would have shown a monolingual advantage. Another possibility is that participants were able to pick up on a pattern and heuristically use it to identify ungrammatical sentences in the Grammaticality Task. All sentences containing a syntactic violation had the same sort of violation consisting of an auxiliary verb followed by a progressive tense (e.g, *can basking*). Therefore it is easy to identify ungrammatical sentences by the presence of the -ing tense marker. It is possible that in Moreno et al., (2010) bilinguals used this strategy to catch up in the Grammaticality Task. In a future study, post-experiment questionnaires asking participants whether they used any sort of strategy while doing the tasks may clarify this issue.

Reading times replicated the findings of Ditman et al. (2007). This is reassuring as it suggests that the stimuli, equipment, and environment elicited the expected type of sentence processing. Both reading time to the critical word and reading time to the final word were affected by sentence type and none of the other factors examined. Sentences containing syntactic violations had longer reading times to the critical word than semantically anomalous or correct sentences. This suggests that syntactic
violations are immediately detected and that the extra time taken reflects the time needed for attempted repair and reanalysis. Semantically anomalous and correct sentences had longer reading times to the final word than sentences containing syntactic violations. This supports the idea that syntactic errors are identified immediately. In both the Acceptability and Grammaticality task, finding a syntactic violation always means the correct response is "no". Therefore it makes sense that readers skim quickly to the end after spotting one. It is interesting that semantically anomalous sentences do not have longer reading times than correct sentences since accuracy data and reaction times suggest that semantic anomaly is difficult to process.

It is possible that the Acceptability and Grammaticality tasks reveal no bilingual advantage because mechanism behind the bilingual advantage in executive function is not inhibition, but in fact facilitation or a combination of both (Colzato et al., 2008). Colzato et al., (2008) compared the performance of bilinguals and monolinguals on tasks that measured different aspects of inhibition. Bilinguals and monolinguals performed similarly on a Stop Signal Task, which is considered a measure of active inhibition, and on an Inhibition of Return (IOR) task - another task associated with inhibition. Bilinguals and monolinguals differed in their performance on an Attentional Blink task. Bilinguals exhibited more of an attentional blink than did monolingual - meaning that bilinguals were less accurate than monolinguals in identifying the second of two temporally close targets. This means that bilinguals invested more mental resources into identifying the
first target at the cost of their ability to identify the second target. The authors believe this result implies that bilinguals are better able to maintain attention on goal-relevant information (in this case, information relevant to identifying the first target) and inhibit goal-irrelevant information (information relevant to identifying the second target).

Colzato et al., (2008) conclude that the mechanism behind the bilingual advantage is not enhanced inhibitory ability, but improved ability to attend to goal-relevant information (i.e. facilitating activation of goal relevant information in addition to or instead of reactively inhibiting goal-irrelevant information). It is possible that researchers studying the possible bilingual advantages in executive function are starting from the wrong premises and that the underlying theory needs to be reexamined and reevaluated. Grosjean (1998) illustrates through examples that inconsistent results often arise because researchers base their hypotheses on an inaccurate or inappropriate model.

Difficulties of Bilingual Research

A final possibility for not finding a bilingual advantage is that "bilingual" is a very complicated label that contains a lot of variation. This variation can consist of other factors that are easily conflated with bilingualism. For instance, socioeconomic status (SES), immigration status, and culture have been found to affect executive function. A study comparing English monolinguals and English-French bilinguals, where the language groups were matched on ethnicity and SES, found no difference in
performance on the Simon Task (Koussaie & Phillips, 2012). In a study (Carlson & Choi, 2009) that compared American populations of English monolinguals and English-Korean bilinguals on measures of EF, there appeared to be a bilingual advantage. However in a comparison between the American bilinguals and Korean monolinguals in Korea, the bilingual advantage disappeared. Carlson and Choi (2009) point out that bilinguals are often also bicultural. It is possible that in many studies of bilinguals the effects of bilingualism and culture are conflated. However this study used 6 year olds. Executive function (EF) improves throughout childhood. It could be that all these populations end up with comparable levels of EF. Or that if this study was repeated with young adults or adults, we might see a bilingual advantage again.

Another study taking place in Scotland comparing older English-Gaelic bilinguals to English monolinguals of the same age group found no evidence of a bilingual advantage in the Simon Task (Kirk, Scott-Brown, & Kempe, 2013). The authors of the study suggest that factors other than bilingualism, such as immigrant status or culture, may be behind the bilingual advantage found in other studies. Grosjean (1998) additionally makes a distinction between bilinguals that are still in the process of acquiring one of their languages and bilinguals that are relatively stable language proficiencies. In future studies it may be helpful to analyze bilingual data based on immigrant status and years spent in an English speaking environment.
Bilinguals can vary in other ways such as languages spoken, proficiency in each language, age of acquisition (AoA), and frequency of switching. These are all factors that can obscure or complicate the results. Ideally all bilingualism research would be conducted using populations such as Spanish-Basque bilinguals in Spain, English-French bilinguals in Canada, or English-Gaelic bilinguals in Scotland and then results compared to those of monolingual Spanish, Canadians, and Scottish. Populations such as these tend not to vary greatly in factors like AoA, immigration status, culture, frequency of language switching, and proficiency. There would be a minimal amount of variation within the bilingual group studied as well as minimal variation between bilinguals and monolinguals -- the only difference between the two groups being that one has native knowledge of two languages and switches between them with relative frequency, while the other has native knowledge of one language and does not switch languages.

It is interesting to note that two recent large-scale studies (with up to 650 participants) were conducted on populations meeting these strict guidelines and found no effect of bilingualism on EF. The first study (Gathercole et al., 2014) involves English-Welsh bilinguals and English monolinguals from North West Wales who "grew up in the same context". The researchers compared monolinguals and bilinguals (across seven age groups ranging from three years old to over sixty) on performance on a switching task, the Simon task, and a linguistic judgment task involving grammatically correct but semantically anomalous sentences. Overall results suggested no substantial evidence for
a bilingual advantage in performance in any of the three tasks. This negative evidence for the existence of a bilingual advantage is especially compelling because usual predictors of performance on these tasks such as age, congruent or incongruent condition, and sentence type remained. For instance, participants' performance on all tasks improved with age until adulthood and decreased in older age groups. Likewise participants were faster to respond to congruent than incongruent trials in the Simon Task and grammaticality judgments were better on grammatical than ungrammatical sentences.

The second study (Duñabeitia et al., 2014) involved 252 carefully matched Basque-Spanish bilingual children and Spanish monolingual children from Spain. The bilinguals were either simultaneous bilinguals or had started to acquire L2 by age four. The children were tested on verbal and nonverbal versions of the Stroop tasks. Results showed no overall difference between monolingual and bilingual children in performance on the Stroop task. The results of these large scale studies involving hundreds of participants which find no effect of bilingualism on measures of EF is especially striking when compared to the small n's used in many of the studies that do report a bilingual advantage. These small n's result in underpowered studies where small perturbations in the data can affect the significance of the hypotheses. It may be the case that if studies reporting a bilingual advantage had used larger sample sizes, those bilingual advantages would disappear.
"Ideal" populations of balanced bilinguals, who switch languages every day and have similar life experiences are not readily accessible to most bilingualism researchers. Furthermore studies based on cases such as these don't have much ecological validity, since most populations of bilinguals are not so homogenous and perfectly balanced bilinguals are unusual. Researchers must carefully design their experiments and control for as many of confounding factors as possible by carefully matching participants on SES, immigration status, proficiency, AoA, and frequency of switching.

Paap, Johnson, & Sawi, (2014) try to isolate the effect of bilinguals' AoA and relative proficiencies on the occurrence of the bilingual advantage in a linguistically diverse group of bilinguals at San Francisco State University and find that these factors do not significantly affect EF. In order to examine the effect of AoA, bilinguals were divided into three groups and compared to matched monolingual groups across four tasks requiring inhibitory control, monitoring, and switching. Results indicated no bilingual advantage and no consistent differences in performance between the bilingual groups.

A second part of the same study used the same four tasks to examine the effect a bilingual's proficiency in his or her two languages on any potential bilingual advantage in EF. Participants' self-rated L2 proficiency scores (using the scale in appendix A) were used to divide participants into five groups ranging from high proficiency to no
proficiency in L2. The results again showed no consistent difference between the
groups. This study would suggest that AoA and relative proficiencies do not affect
measures of EF and so increase the validity of studies that involve bilinguals with a range
of AoAs and proficiencies.

Problem of Confirmation Bias

Confirmation bias is another possible pitfall in the study of the bilingual advantage. A
recent review (Paap, Johnson & Sawi 2015) identifies the causes of confirmation bias as
institutional (relating to the established practices of academic research) and
methodological (related to the personal practices of certain researchers). Journals and
conferences are less likely to accept studies reporting null results or replications
because they are looking for novel or newsworthy items. Mahoney (1977) shows that
reviewers were biased in favor of positive results. Furthermore de Bruin, Treccani, &
Della Sala (2015) found in a survey of 104 conference abstracts that out of the abstracts
that later became published only 29% reported no evidence for a bilingual advantage;
whereas 68% of the published abstracts reported evidence supporting a bilingual
advantage. The prevalent bias for positive results in reviewers and editors discourages
researchers from attempting to publish null results and leads to the "file drawer
problem." Researchers file away null results instead of trying to get them published.
Paap & Liu (2014) suggest that the bilingual advantage is an "undead theory" as described by Ferguson & Heene (2012) where published positive results give a theory the appearance of viability even though there might be copious unpublished results that provide evidence against this theory. The "file drawer problem" and "undead theories" are general problems in the field of psychology, but specifically they have an impact on the theory of the bilingual advantage. Together they create biases both in what journals publish and what researchers choose to report, which causes the theory of the bilingual advantage to appear stronger than it actually is.

The way researchers report and interpret their data is influenced by the preference for positive results. For instance according to John, Loewenstein, & Prelec (2012), 48% of researchers only report studies that worked. Additionally researchers sometimes only report the statistical tests and interactions that support a positive result while downplaying or ignoring those that do not. When it comes to interpreting the data, there is a tendency to interpret any difference between monolinguals and bilinguals as evidence for a bilingual advantage. This can especially be the case in neuroscience data. For instance, Moreno et al. (2010) interpret higher N400s in bilinguals during the Grammaticality Task as evidence for a bilingual advantage in EF. The reasoning behind this interpretation is that bilinguals have the mental resources to process semantic information in addition to the task-relevant grammatical information. However Paap & Liu (2014) argue that the same ERP results are better interpreted as evidence against
the existence of a bilingual advantage. The bilinguals' higher N400 amplitudes during the Grammaticality Task could indicate that bilinguals are not adequately inhibiting task-irrelevant semantic information. In the Grammaticality Task, semantic information is not relevant to the final judgment of whether a sentence is grammatically correct or not. The bilinguals' higher N400 amplitudes indicate that task-irrelevant processing intrudes on the task-relevant processing of bilinguals.

Conclusion

Bilingualism manifests itself in a variety of ways. For this reason, results have to be replicated in multiple populations of bilinguals before they can be said to apply to all bilinguals. Bilingualism is a complicated phenomena, so carefully designed experiments that are replicable are especially important (Refer to Grosjean (1998) for an overview of some of the subtleties of bilingual research). Peal & Lambert (1962), one of first studies to suggest that bilingualism could be beneficial to cognitive function, emphasized the importance of carefully controlling for confounds. Like Peal and Lambert we stress the importance of careful experimental design, as well as replicability in bilingual research.

References


http://doi.org/10.1037/0278-7393.34.2.302


*Bilingualism: Language and Cognition, 1*(2), 131–149.  
http://doi.org/10.1017/S136672899800025X

Hilchey, M. D., Saint-Aubin, J., & Klein, R. M. (n.d.). Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks?


Kirk, N. W., Scott-Brown, K., & Kempe, V. (2013). Do older Gaelic-English bilinguals show an advantage in inhibitory control? Retrieved from  

http://doi.org/10.1016/j.brainres.2012.01.052


Appendix A

Instructions for self-rating of language proficiencies shown to participants

Use the 1 to 7 scale shown at the bottom of the page. The description next to each rating-scale number will help you to select the level that best describes your level of proficiency. CONSIDER ONLY YOUR SPEAKING AND LISTENING SKILLS. Please disregard your ability to read or write in the designated language.

**Scale: Proficiency in Speaking and Listening (do not consider reading and writing skills)**
1- Beginner: Know some words and basic grammar.
2- Advanced beginner: Can converse with a native speaker only on some topics and with quite a bit of difficulty.
3- Intermediate: Can converse with a native speaker on most everyday topics, but with some difficulty.
4- Advanced Intermediate: Can converse with little difficulty with a native speaker on most everyday topics, but with less fluency than a native speaker.
5- Near Fluency: Almost as good as a typical native speaker on both everyday topics and specialized topics I know about.
6- Fluent: As good as a typical native speaker.
7- Super Fluency: Better than a typical native speaker.
PURPOSE AND BACKGROUND
The primary purpose of this research study is to compare the relative difficulty of two types of judgments about English sentences. In the “grammaticality” task you must decide if the sentence you just read is “grammatical” or has a “grammar error”. In the “acceptability” task you must decide if the sentence you just read is “OK” or “not OK”. Sentences may be “not OK” because they have a grammar error or because the meaning is odd or makes little sense. We also want to test the hypothesis that performance in these tasks may differ for bilinguals compared to English-speaking monolinguals.

The researcher, Yunyun Liu, is a second-year graduate student in the psychology program working under the supervision of Dr. Kenneth Paap, a professor at San Francisco State University

PROCEDURES
If you agree to participate in this research study, the following will occur:

- Activity 1. You will fill out a background questionnaire that takes about 10 minutes. Several of the questions are about the different languages you speak. We want to include participants who are both monolingual (e.g., speak only English) and participants who are bilingual or multilingual. Two questions ask about your educational level and that of your parents. Several questions ask how often you do specific types of multi-tasking (e.g., do homework while listening to music, etc.).
- Activity 2. You will do the first sentence evaluation task. On each trial you will read a sentence one word at a time by pushing the space bar on the computer keyboard. After the last word you will be asked a question and respond either “yes” or “no” with a key press. (Takes about 15 minutes)
- Activity 3. You will take a computer-controlled version of the Raven’s Advanced Matrices Test. Each trial asks you to select the visual pattern that
best completes the series of patterns that you are looking at. (Takes about 10 minutes)

- Activity 4. You will do the other sentence-evaluation task. Half of the participants do the “grammaticality” task first and the others do the “acceptability” task first.
- Activity 5. You will be given a debriefing document to take with you that provides more details about the purpose of the experiment. You will also have a chance to ask your experimenter questions about the study.

The study should take 45 minutes to an hour. It takes place in Room 513 of the Ethnic Studies and Psychology Building.

RISKS
The risks are minimal. If you experience any discomfort or anxiety due to the nature of the questions you can skip those questions. Some people find the computer-controlled tasks somewhat repetitive or boring. You can take a break when you want to. You can stop participating completely at any time.

D. CONFIDENTIALITY
We will use the last four digits of your SFSU student identification number as an identifier that links your responses to the questionnaire to your responses in the computer-controlled tasks. No names or identifiers will be used in any informal or published reports of the research. The research data will be kept in a secure location and/or password protected. Only the researcher will have access to the data. The background questionnaires will be destroyed at the end of the study.

E. DIRECT BENEFITS
There will be no direct benefits to the participant. Indirect benefits may include a greater understanding of how research is conducted in psycholinguistics.

F. COSTS
There will be no cost to you for participating in this research.

G. COMPENSATION
There will be no compensation for participating in this research.

H. ALTERNATIVES
The alternative is not to participate in the research.

I. QUESTIONS
You have spoken with the researcher about this study and have had your questions answered. If you have any further questions about the study, you may contact Ken Paap (kenp@sfsu.edu). Questions about your rights as a study participant, comments, or complaints about the study may also be addressed to the Human and Animal Protections at 415-338-1093 or protocol@sfsu.edu.

J. CONSENT
You have been offered a copy of this consent form to keep.
PARTICIPATION IN THIS RESEARCH IS VOLUNTARY. You are free to decline to participate in this research study, or to withdraw your participation at any point, without penalty. Your decision whether or not to participate in this research study will have no influence on your present or future status at San Francisco State University.

Signature ___________________________ Date: __________
Research Participant

Print Name ___________________________
Research Participant

Signature ___________________________
Researcher

Date: __________