Distribution Agreement

Kaveri Sheth March 27, 2019

Understanding Cross-Language Self-Derivation through Language Source Memory in Bilingual College Students

by

Kaveri K. Sheth

Patricia J. Bauer
Adviser

Department of Psychology

Patricia J. Bauer

Adviser

Lisa Dillman

Committee Member

Robyn Clarke

Committee Member

2019

Understanding Cross-Language Self-Derivation through Language Source Memory in Bilingual College Students

By

Kaveri K. Sheth

Patricia J. Bauer

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Psychology

2019

Abstract

Understanding Cross-Language Self-Derivation through Language Source Memory in Bilingual College Students
By Kaveri K. Sheth

Knowledge can be learned and also built up through the productive process of self-derivation through integration. During this process, learners acquire information through direct tuition in separate but related episodes, integrate these episodes, and form new knowledge that was never directly provided. Adults who learn information in multiple languages encounter separate episodes across languages, where episodes share low surface similarity. This low surface similarity of cross-language episodes may make integration difficult, because it is harder to recognize the relatedness between episodes. The primary goal of the present study was to understand how the presentation of information in different languages impacts self-derivation through integration. The prediction that cross-language integration may be a challenge rests on the assumption that the participant attends to and encodes the surface feature of the language in which the fact was learned. If the participant attended to the source of the fact, then they encoded or "tagged" the fact in terms of the language in which it was presented. If this is the case, then participants should be able to identify the source (the language) in which the individual facts were presented. Spanish-English bilingual college students were presented with audio recordings of fact pairs within and across languages in English and Spanish, accompanied by a visual cue. Participants were assessed for self-derivation of new knowledge through integration and the source of the individual stem facts and the perceptions of the new knowledge self-derived through integration. Participants successfully self-derived new information within and across languages. They were most successful in the within-language condition, supporting the assumption that cross-language integration is challenging. Furthermore, participants were able to successfully nominate the source of the facts in the within-language conditions and one cross-language condition. Variability in performance could be attributed to Spanish proficiency levels. The results provide evidence that cross-language integration is difficult because participants successfully encode the source, indeed paying attention to the surface features of the fact, and thus must overcome low levels of surface similarity in order to integrate separate facts and self-derive new knowledge.

Understanding Cross-Language Self-Derivation through Language Source Memory in Bilingual College Students

By

Kaveri K. Sheth

Patricia J. Bauer

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Psychology

2019

Acknowledgements

Thank you to Dr. Patricia Bauer, Dr. Alena Esposito, and Jessica Dugan for their mentorship, support, and guidance during the research process. Thank you to the Emory Independent Research Grant for helping fund this project.

Table of Contents

Introduction

Knowledge is vital for educational and career success and accumulates across different experiences and environments throughout life. Knowledge can be directly learned, or generated through productive processes, such as analogy or deduction. One such productive process is self-derivation through memory integration (e.g., Bauer & San Souci, 2010). During this process, learners acquire information through direct tuition in separate but related episodes, integrate these episodes, and form new knowledge that was never directly provided. This process is especially important for students, who are required to integrate and self-derive across different learning experiences, such as readings, lectures, and homework as they build a knowledge base over time. Indeed, self-derivation through integration is related to concurrent GPA and has been shown to predict GPA two years later (Varga, Esposito & Bauer, 2019). An increasing number of international students in U.S. colleges and universities means that many students must combine information not only between different episodes, but between different languages as well. The purpose of the current research was to expand upon our understanding of how self-derivation is accomplished with the added challenge of integration across languages. We tested whether integration and self-derivation across languages was more challenging than integration and self-derivation within the same language, and the potential source of any "cost" observed.

Self-Derivation through Integration

Learners can build their knowledge bases by successfully self-deriving new knowledge from previous information. They do so by extending beyond two facts directly acquired at different times or experiences in order to make logical inferences. In a standard laboratory paradigm to test self-derivation, participants are provided with two true but novel facts (i.e., stem facts) and then asked questions that require integration and subsequent self-derivation to successfully answer (Bauer & San Souci, 2010). For example, participants first learn Stem fact 1, *The first American woman in space was Sally Ride.* Later in the session they learn Stem fact 2, *Sally Ride was a very good tennis player.* Participants are then asked self-derivation questions, in which the answer should be the new information they derived from the two stem facts. For example, from these two related facts, participants can derive *The first American woman in space was a very good tennis player*. Previous work with adults has demonstrated that newly self-derived knowledge is rapidly incorporated into the knowledge base. In Bauer and Jackson (2015), participants read novel fact pairs, presented with intervening facts. These separate but related facts were able to be integrated in order to self-derive new knowledge, although participants were not told of their relatedness. Neural activity was measured using event-related potential (ERPs) to measure the brain's electrical activity during information processing. At the first 400ms presentation of a fact derived through integration, neural responses to the newly generated facts were in between responses to novel and well-known facts. While the integration facts were actually novel, the brain did not treat them in that way, nor did the brain treat them as wholly familiar. At the second 400ms presentation, neural responses to the newly generated facts no longer differed from well-known facts; both newly generated and well-known facts differed from novel. This transition from novel to well-known after one 400ms long presentation demonstrates that newly self-derived information is rapidly incorporated into semantic memory. This shows that semantic memory can be quickly expanded upon by a productive process, such as integration.

Self-derived knowledge has been shown to remain accessible over time. In a study by Varga and Bauer (2017b), young adults participated in the standard self-derivation paradigm. They self-derived new knowledge through integration at Session 1 and the newly derived knowledge was retained over a 1-week delay (Varga & Bauer, 2017b). This finding is consistent with the conclusion that self-derivation is a mechanism by which we build up our knowledge base. Additionally, self-derivation performance is related to measures of academic achievement as measured by SAT scores and college GPA (Varga et al., 2019). This further demonstrates the importance for learning through productive processes in order to incorporate new knowledge into our semantic memory.

Given the importance of self-derivation through integration for accumulation of knowledge, it is important to determine the conditions under which it takes place. Self-derivation performance can be challenged by low levels of surface similarity of to-be-integrated information. In a study with children, it was shown that low surface similarity interferes with self-derivation success (Bauer, King, Larkina, Varga & White, 2012). In this study, separate but related facts were presented in two passages, each of which had a distinct main character (low surface similarity) or shared a main character (high surface similarity). Only 37% of children successfully generated the integration fact in the low surface similarity condition, whereas 63% of children were successful in the high surface similarity condition. Lower surface similarity likely challenges integration by making it more difficult to recognize relatedness between information. There are numerous other sources of low surface similarity. In the present research, we investigated the potential challenge posed by the requirement to integrate across two different languages.

Cross-Language Self-Derivation

Previous research has shown that self-derivation based on integrated information may be a challenge for bilingual individuals who are required to integrate across languages (Menkes, Esposito & Bauer, 2017; Esposito and Bauer, 2017a). The requirement to integrate facts across two different languages places higher demand on the process of self-derivation, because low surface similarity interferes with integration, which is a process upon which self-derivation depends. Cross-language self-derivation through integration is especially relevant to college students who are pursuing an education in a language different from their previous education. Many foreign students come to the United States for their education. The Current Population Survey reports that the number of international students in American universities has increased from 1.5% in 1975 to 4.8% in 2015. Emory University, specifically, has 17% of its student body made up of international students, many of whom speak a native language other than English. In fact, 33% of the foreign-born population in the United States speak a language other than English, in addition to English, at a level deemed "very well," categorizing them as bilingual. Non-native English speaking students are required to process and combine information in the language of their previous education with information learned in an English speaking classroom environment.

A few studies inform our expectations regarding cross-language self-derivation through integration. For example, Menkes et al. (2017) investigated whether bilingual college students' performance differed between cross-language and same-language conditions. There were four language conditions: English/English (E/E; participants received both facts in English), Spanish/Spanish (S/S; participants received both facts in Spanish), English/Spanish (E/S; participants received the first fact in English and the second in Spanish), and Spanish/English (S/E; participants received the first fact in Spanish and the second in English). Menkes and colleagues hypothesized that integration performance would be higher in a participant's native language. Additionally, they hypothesized that greater language proficiency would predict higher integration success. Performance was strongest in the E/E condition, followed by both cross-language conditions, and the S/S condition lowest performing. Spanish vocabulary was a significant predictor of performance within the S/S condition only. Analyses were not conducted using native language due to unequal groups. The authors interpreted the patterns of results to be due to differences in language proficiency.

Esposito and Bauer (2017a) investigated cross-language self-derivation of new information through integration specifically among children participating in a Spanish-English dual language academic program. Children in fourth grade were able to successfully self-derive new knowledge through integration of separate but related facts in both cross-language and within-language conditions. However, third graders were not able to integrate as successfully in across-language conditions compared to within-language conditions. However, overall language proficiency did not differ between third and fourth grade. This suggests that it was not a group difference, but rather individual differences in language proficiency that may contribute to integration success.

Critically, the prediction that cross-language integration may be a challenge rests on the assumption that the participant attends to and encodes the surface feature of the language in which the fact was learned. If participants attend only to the meaning --and they do not attend to the surface feature of the language of presentation--then cross-language integration should be as successful as within-language integration. However, if participants attend to and encode the facts in terms of the specific languages in which they are presented, then cross-language integration should be difficult because the different languages have different surface features, resulting in lower levels of surface similarity. This, in turn, implies that the participant attended to the source of the fact, and that they encoded or "tagged" the fact in terms of the language in which it was presented. If this is the case, then participants should be able to identify the source (the language) in which the individual facts were presented.

Present Study

The primary goal of the present study was to understand how the presentation of information in different languages impacts self-derivation through integration. If the languages of the facts are different, surface similarity is low, and integration will be more challenging. However, this prediction rests on the assumption that participants attend to the language and encode the source. To date, this has never been tested, and could be one possible explanation for previous findings that cross-language integration is not that much harder than within-language integration (Menkes et al, 2017). Thus, it could be that the participants did not even notice the languages were different, perhaps due to high levels of language proficiency. Hence, the secondary purpose of the present study was to test the assumption that underlies the prediction that cross-language integration will be difficult. To accomplish this, we tested whether stem facts were encoded in the language of presentation by asking about the source of the stem facts. The third purpose of the present research was to ask about the perceptions of the source of their newly self-derived knowledge.

Hypotheses

We hypothesized that if cross-language self-derivation through integration is more difficult, then participants would self-derive more successfully in within-language conditions than cross-language conditions. Secondly, it was hypothesized that if participants attended to and encoded the facts in the language of presentation, then they would have higher source memory of the language they learned the fact in. Lastly, the third goal of the research does not have a hypothesis. Since we asked about source for stem facts, we took the opportunity to ask about the source of their newly self-derived knowledge, in order to understand the perceptions of the new knowledge.

Method

Participants

The sample consisted of 25 students (18 women, 7 men, M_{age} = 19.96, SD = 1.02, age range: 18 to 21 years) drawn from the Emory University undergraduate student population. Participants were recruited from introductory psychology courses and advanced Spanish language courses. Participants were eligible for the study if they self-identified as bilingual in Spanish and English, were native Spanish speakers, or had taken a Spanish course above the 300 level at Emory. These criteria ensured that participants were competent in Spanish verbal comprehension. English proficiency was assumed because coursework at Emory University requires a high degree of English proficiency for admission. Participants enrolled in introductory psychology received course credit upon completion of the study; participants recruited from outside this population were compensated with $20. All participants provided written informed consent prior to the study. Emory University's institutional review board (IRB) approved the study protocol and procedures.

Materials

Experimental Stimuli. The stimuli consisted of 16 novel fact pairs, each ranging from 5-15 words. All of the stimuli have been used in prior research for testing memory integration. The stimuli were created by various members of the lab. The stimuli covered a variety of domains, including science topics, such as geology and astrology. There were also stimuli about scientific inventions and famous people. The stimuli were tested in both 1-stem and 2-stem conditions in previous research. In the 1-stem condition, participants are exposed to one but not both of the stem facts necessary for self-derivation. This is done to ensure that participants cannot answer the integration question without exposure to both of the stem facts. The 2-stem self-derivation performance was higher than the 1-stem self-derivation performance, meeting our required performance criteria. The fact pairs consisted of separate but related facts (stem facts), which, when combined, could lead to the derivation of an integration fact, as stated in the previous example. The facts were presented as audio, recorded by a single native Spanish speaker and a single native English speaker, paired with visual cues through PowerPoint. The visual cues consisted of two images that represented the fact (Figure 1, Panels A and B). The facts were translated into Spanish by the experimenter and back-translated by a native Spanish speaker to ensure accuracy. The stimuli used are listed in Table 1, with the Spanish translation and corresponding visual cues.

Visual cues were used because it has been shown that the language during the test phase biases participants' nomination of the language source (Esposito & Bauer, 2018). In a study with children, they were presented with a written fact in either Spanish or English (Session 1). A week later, they were presented with the fact again in either the same or the other language (Session 2). Then, they were asked in which language they originally heard the fact (i.e. the source). The language of the second presentation influenced the participants' nomination of the language in which they originally heard the fact. For example, if the fact was presented in English at Session 2, the participants were more likely to nominate they originally heard the fact in English regardless of whether it was presented in Spanish or English at Session 1. The current research was designed to eliminate any extra verbal cues, aside from the fact itself, through the use of visual cues.

Language Measures. Two measures were used to quantify language proficiency and language experience of each participant: The Woodcock-Muñoz Language Survey®-Revised Normative Update (*WMLS®-R NU*) and The Language Exposure Questionnaire.

The *WMLS®-R NU* is a standardized test used to assess vocabulary proficiency. It is a validated measure of both Spanish and English verbal comprehension appropriate for ages 2 to 90+. Test 1, vocabulary, and Test 2, analogies, were used in both Spanish and English. The vocabulary test consists of pictures that participants are asked to identify. The analogies test requires that participants complete with the relationship with the logical vocabulary word. A raw score for English proficiency was calculated by summing the number of questions that were answered correctly for the English vocabulary and analogy tests. A raw score for Spanish proficiency was calculated by summing the number of questions that were answered correctly for the Spanish vocabulary and analogy tests.

The Language Exposure Questionnaire (Appendix A) was adapted from the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007). The LEAP-Q was shortened to alleviate participant burden and reduce session time. The questions pertaining to participant's cultural identification and health issues were eliminated from the LEAP-Q to create a modified version because they were not anticipated to be utilized in the analyses. The questionnaire asked questions only about participant's language experience, such as in what context and to what extent a language is used. This measure was used to examine the participant's language experience and self-rated proficiency across their lifetime and in different settings.

Procedure

Participants came into the laboratory for two sessions, 24 to 48 hours apart (*M* = 36.4). Participants were tested individually at both sessions; both sessions were conducted by the same female experimenter with Spanish proficiency (the author). All instructions were provided in English.

Session 1: There were four language conditions: both stem facts presented in English (E/E), both stem facts presented in Spanish (S/S), Stem 1 presented in English and Stem 2 in Spanish (E/S), Stem 1 presented in Spanish and Stem 2 in English (S/E). Each participant completed all conditions. They saw 16 fact pairs total with four fact pairs in each language condition. Facts were presented in two phases. In Phase 1, 16 individual facts were presented, blocked by language. For each fact, a visual cue was provided on the computer screen (Figure 1). For example, in Phase 1 participants listened to 8 facts in Spanish, 8 facts in English. The facts were followed by a buffer task, unrelated to the purposes of the current research. Then, in Phase 2, participants listened to the facts that were related to those presented in Phase 1, with 8 facts in Spanish and 8 facts in English. Again, all facts were accompanied by a visual image. At no time were the participants told the facts were related. The facts were presented in 1 of 4 different orders. The order of the facts were counterbalanced to ensure that a particular order did not influence participants' responses.

After listening to all of the stem facts and upon completion of Phase 2, participants were asked open-ended self-derivation questions. The correct answers to the self-derivation questions were the integration fact that participants could have derived after hearing two separate but related facts. For example, from the previously mentioned stem facts (Figure 1, Panels A and B), the participant could have derived the integration fact: *The first American woman in space was a very good tennis player* (Figure 1, Panel C).

Session 2: In the second hour-long session, participants were presented with the visual cue only, for both stem facts and integration facts, in order to assess their source memory without any language bias. The visual cues for the stem facts were the images that had accompanied the verbal fact presentation in Session 1 (Figure 1, Panels A and B). The visual cues for the integration fact were combinations of the cues from the two related stem facts (Figure 1, Panel C). Note that participants had never seen these particular combinations of visual cues for integration facts before. Participants were also presented with 8 new image pairs unrelated to those previously shown to provide the opportunity for participants to distinguish between new and previously seen images from Session 1. Thus, participants saw a total of 56 image pairs derived from stem facts that were previously presented (32), integration facts that were not presented previously but participants could have self-derived (16), and brand new images that they had not previously seen (8). Participants were told that they had either heard the facts associated with the images, self-derived a fact associated with these images, or never seen the images before. These instructions applied for all of the images, regardless of whether they had self-derived successfully in Session 1. Participants were instructed to identify, upon seeing the visual cues, whether the associated fact was new, old, or put together from the first session. When participants identified a fact as “old” or “put together,” they were asked to identify the language in which they heard the fact during the first session. If they said they heard the fact in both English and Spanish, they were asked whether the same fact was repeated in both languages or if part of the fact was in Spanish and part of it in English. All participants indicated comprehension of these instructions.

Participants then completed the Language Exposure Questionnaire and the Woodcock-Muñoz Language Survey-Revised Normative Update, starting with the Spanish tests and finishing with the English tests.

Scoring

Participants received a self-derivation score for each language condition based on the number of correct open-ended integration questions. The range for each condition was 0 (no correctly answered open-ended integration questions) to 4 (all correctly answered open-ended integration questions). Participants only received a source performance score on the correctly answered integration questions from Session 1, although they were asked the source for all of the facts. Participants received an overall source performance score for each language condition, which was the number of correctly attributed language source questions over the number of correctly answered integration questions per language condition. Language proficiency was scored using the instructions of the Woodcock-Muñoz Language Survey-Revised Normative Update.

Results

The results are reported in three sections: whether cross-language integration is more challenging than within-language integration, whether participants encode the source of the stem facts, and the perceptions of newly formed self-derived knowledge. First, we report self-derivation performance within-language conditions. We report differences in performance in each of the language conditions as a function of native language. We also report differences in performance as a function of language proficiency. Next, we report source memory performance for the stem facts. We report differences in performance in each of the language conditions as a function of native language and language proficiency. Lastly, we looked at the perceptions of the newly formed self-derived knowledge within language conditions. All statistical tests reported as significant were below an alpha level of 0.05.

Is Cross-Language Integration More Challenging than Within-Language Integration?

Self-derivation performance means and standard deviations by language condition are reported in Table 2. As can be seen in the table, participants' performance was the highest in the E/E language condition and the lowest in the S/E condition. In order to determine whether cross-language integration was more challenging than within-language integration, a one-way repeated measures ANOVA was conducted to examine the differences between the four language conditions. There was a significant main effect of the language condition, $F(3, 72) = 3.39$, $p = 0.02$. Post hoc comparisons with Bonferroni corrections indicated that performance in the E/E language condition was significantly higher than performance in the S/E language condition and the E/S language condition. There were no other significant differences between language conditions. Thus, performance in the E/E language condition was greater than cross-language conditions, regardless of whether English or Spanish was first. Performance in the S/S language condition was intermediate and did not differ from the other conditions.

Self-derivation performance means by language condition and native language are reported in Table 3. Two participants were excluded from analyses due to having a native language other than Spanish or English. To determine whether self-derivation performance varied as a function of native language, we conducted an ANOVA with native language (Spanish or English) as the between-subjects factor and language condition (E/E, S/S, S/E, E/S) as the within-subject factor. There was no main effect of native language and there was no interaction between language condition and native language, $F(3, 63) = 0.54$, $p = 0.66$. Thus, performance did not vary as a function of native language.

Native language did not interact with self-derivation performance, so we analyzed whether language proficiency in English or Spanish would be a better indicator of self-derivation performance. We used ANCOVAs with language proficiency in English or Spanish as the covariate and language condition (E/E, S/S, S/E, E/S) as the within-subjects factor. All 25 participants were included in these analyses. The ANCOVA with English revealed that performance by language condition did not vary as a function of English language proficiency, $F(3, 69) = 0.90$, $p = 0.45$. However, the ANCOVA with Spanish as the covariate revealed that performance by language condition did vary significantly as a function of Spanish language proficiency, $F(3, 69) = 2.73$, $p = 0.05$. Thus in summary, whereas self-derivation performance did not vary significantly as a function of native language or English proficiency, it did as a function of Spanish proficiency.

Do Participants Encode the Source of the Stem Facts and New Facts?

The mean percentage and standard deviation for source performance for stem facts are reported in Table 4 (Panel A). Participants were required to nominate whether the visual cues that represented the facts were new, old, or put together. Note that the correct nomination for new facts was “new” and for stem fact visual cues, the correct nomination was “old.” As the first step in the analysis, source performance for the facts was analyzed using a one-sample t-test to determine whether source nominations were significantly above chance (.33). For the new facts, we could not conduct this analysis because all participants were at ceiling for correctly nominating new facts as “new.” For stem facts, across language conditions, participants correctly nominated the stem fact visual cues as “old” at above chance levels, $t(24) = 19.03$, $p < .001$. When analyzing the correct language source nomination by language condition for the stem facts, participants were significantly above chance in E/E, S/S, and E/S language conditions. They were not significantly above chance for the S/E language condition.

In order to assess whether participants attended to and encoded the stem facts in the language of presentation, a one-way repeated measures ANOVA was conducted to compare the differences in source memory accuracy for stem facts between the four language conditions. The ANOVA revealed that there was a main effect of the language condition, $F(3, 63) = 2.92$, $p = 0.04$. Post hoc comparisons with Bonferroni corrections for multiple comparisons were made to examine the differences in the language conditions. This indicated that 2 pairwise comparisons were significant. Participants performed significantly higher in the S/S language condition than the S/E language condition ($SE = 0.13$, $p = 0.04$), and the E/S language condition ($SE = 0.12$ $p = 0.02$). There were no other significant differences between language conditions. This shows that when Spanish was involved, performance was better in the within S/S language condition than the cross-language conditions.

To determine whether source memory performance for stem facts varied as a function of native language, we conducted an ANOVA with native language (Spanish or English) as the between-subjects factor and source performance in each language condition (E/E, S/S, S/E, E/S) as the within-subject factor. Two participants were excluded from this analysis due to having a native language other than Spanish or English. The ANOVA revealed that there was no interaction between source performance and native language, $F(3, 63) = 0.71$, $p = 0.55$. This means that source performance did not vary as a function of native language.

Native language also did not interact with source memory performance for stem facts, so we analyzed whether language proficiency in English or Spanish would be a better indicator. We used two separate ANCOVAs with language proficiency in English or Spanish as the covariate and source performance in each language condition (E/E, S/S, S/E, E/S) as the within-subject factor. All 25 participants were included in these analyses. The ANCOVA with English revealed that performance by language condition did not vary as a function of English language proficiency, $F(3, 69) = 1.07$, $p = 0.37$. Likewise, the ANCOVA including Spanish language also revealed that performance by language condition did not vary as a function of Spanish language proficiency, $F(3, 69) = 0.90$, $p = 0.27$. Thus, source performance did not vary as a function of native language or as a function of English or Spanish language proficiency.

What are the Perceptions of the Newly Self-Derived Knowledge?

Source performance for the integration facts was analyzed in the same way as for the stem facts. The mean percentage and standard deviation for source performance for integration facts are reported in Table 4 (Panel B). The correct answer for the integration fact visual cues was "put together." Across language conditions, participants correctly nominated that integration fact visual cues were "put together" at above chance levels, $t(24) = 6.12$, $p < .001$. When analyzing the language perception of "put together" facts in each language condition, participants were significantly above chance in the S/S language condition, $t(24) = 3.75$, $p < .001$. Hence, they correctly nominated that both stem facts in the S/S language condition were presented in Spanish. They were not above chance for the other language conditions.

Discussion

The primary goal of the present study was to understand how the presentation of information in different languages impacts self-derivation through integration. It was found that cross-language self-derivation through integration is indeed challenging. We found that self-derivation performance was better in the English within-language condition, followed by the cross-language conditions. Thus, the hypothesis that cross-language integration would be more difficult is supported.

The prediction that cross-language integration would be challenging rests on the assumption that people attend to the language and encode the source. Hence, the secondary purpose of the present study was to test the assumption that underlies the prediction that cross-language integration will be difficult, by asking about the source of the stem facts. We hypothesized that if participants attended to and encoded the facts in the language of presentation, then they would have higher source memory of the language they learned the fact in. Participants were successful in encoding the source for the E/E, S/S, and E/S language conditions, but not in the S/E language condition. Thus, this hypothesis was supported in those conditions, but not in the S/E language condition. However, other factors such as language proficiency could still play a role in whether participants even notice the languages were different. It was found that self-derivation success did vary significantly as a function of Spanish proficiency. Hence, it may be possible that different levels of Spanish proficiency influence whether participants realize the differences in the languages of the facts, and thus whether they attend to and encode the source.

Lastly, the third purpose of the present research was to ask about the perceptions of the source of their newly self-derived knowledge. Since we asked about source for stem facts, we took the opportunity to ask about the source of their newly self-derived knowledge, in order to understand the newly formed perceptions. Participants were successful in nominating the visual cues for the integration fact as “put together.” Participants were able to nominate the language source as English, Spanish, or both languages. Participants nominated the language for both stem facts in the S/S language condition as Spanish. However, only the S/S language condition nominations were significantly different from chance. It was shown that when Spanish was involved, source performance for stem facts was better in the within S/S language condition, then the cross-language conditions. Hence, since participants successfully encoded the two stem facts in Spanish, they were led to the perception that the newly derived knowledge was in Spanish as well.

As in previous research, the present study revealed that self-derivation performance was the strongest in the E/E condition (Menkes et al, 2017). However, in Menkes and colleagues' study, performance was least successful in the S/S condition. In the present study, self-derivation was the most successful in E/E, with performance least successful in the cross-language conditions. S/S performance was intermediate. The differences could be explained by varying levels of Spanish proficiency among participants. Overall, many of Menkes and colleagues' participants acquired Spanish as a second language later in life and had lower Spanish proficiency, lowering their performance in the S/S language condition. However, in the present research, many of the participants were native Spanish speakers who had high proficiency levels, which could explain why they performed well in the S/S language condition. The present research also suggests that low surface similarity between episodes challenges integration for adults, just like it does in children (Bauer et al., 2012).

The present study examined self-derivation across languages and adults' language attributions of the source of their newly learned knowledge. It is the first study to question the idea that cross-language will only be difficult if stem facts are encoded in the specific language in which they were heard. The present research adds to the literature by demonstrating that cross-language is indeed difficult because participants attend to the source. However, other factors such as language proficiency do play a role in the success of encoding the source and, in turn, self-derivation. The findings of the current study have implications for adults who learn information in multiple languages. Previous research has shown that self-derivation through integration aids the accumulation of knowledge in adults and are related to concurrent GPA (Bauer & Jackson, 2015; Varga & Bauer, 2017b; Varga, Esposito & Bauer, 2019).

Limitations

The sample of participants created some limitations. The language of instruction is English at Emory University. The participants are all exposed to English daily. Hence, many participants who are native Spanish speakers reported that their dominant language was English. Only 6 of 25 participants reported they were dominant in Spanish. Additionally, the study limited proficiency level to either native Spanish speakers or upper level Spanish students. The results most likely do not encompass lower levels of Spanish proficiency. In a future study, in order to examine the effects of native language and language proficiency, the study should be repeated with a sample that is largely Spanish dominant.

Additionally, this study was limited by our ability to fully measure language experience. It is difficult to quantify language experience since it is very dynamic. There were many measures of interest reported in The Language Exposure Questionnaire, but since our population was a small sample of relatively homogenous skills, many of the measures of interest were restricted and we were not able to measure them. In a future study, in order to measure the full effects of language proficiency, the study should involve a population that has a wide range of both Spanish and English language proficiency.

Conclusion

In conclusion, bilingual adults can self-derive information within and across languages. This research demonstrated how bilingual adults can accumulate knowledge through episodes presented in different languages. The present research suggests that cross language integration is difficult because participants attend to and encode the source. The present research showed how language proficiency can affect self-derivation success or source attributions.

References

Bauer, P. J., & Jackson, F. L. (2015). Semantic elaboration: ERPs reveal rapid transition from novel to known. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 271.

Bauer, P. J., King, J. E., Larkina, M., Varga, N. L., & White, E. A. (2012). Characters and clues: Factors affecting children's extension of knowledge through integration of separate episodes. *Journal of Experimental Child Psychology*, *111*(4), 681–694.

Bauer, P. J., & San Souci, P. S. (2010). Going Beyond the Facts: Young Children Extend Knowledge by Integrating Episodes. *Journal of Experimental Child Psychology*, *107*(4), 452–465.

Banjong, D. N. & Olson, M. R. (2016). Issues and trends of international students in the United States. International Journal of Education, 4(1), 1–14.

Dugan, J.A., & Bauer, P. J. (2018). Relations Between Productive Processes and Metacognition. Platform paper submitted to Emory Laney Graduate School

Esposito, A. G., & Bauer, P. J. (2017a). Building a knowledge base across languages: Self-generating new factual knowledge in a dual-language classroom. Manuscript in preparation.

Esposito, A. G., & Bauer, P. J. (2018). Building a knowledge base: Predicting self-derivation through integration in 6- to 10-year olds. *Journal of Experimental Child Psychology*, *176*, 55-72.

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 (4), 940-967.

Menkes, M., Esposito, A. G., Bauer, P. J. (2017). Self-Generation of novel information through cross-language episodes. Thesis submitted to the Emory College of Arts and Sciences.

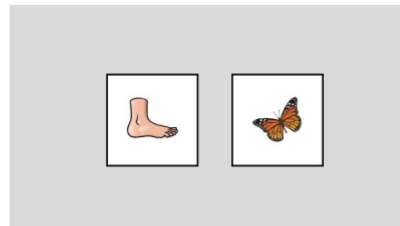
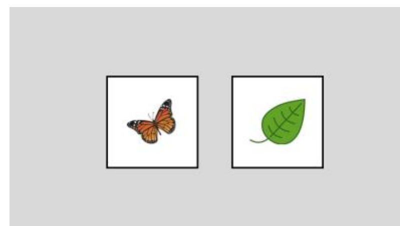
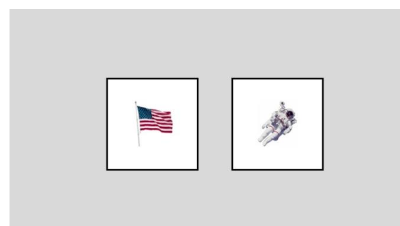
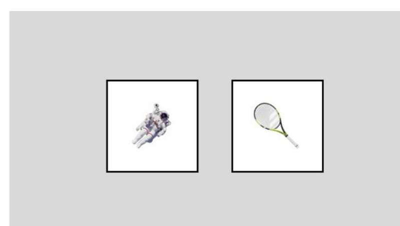
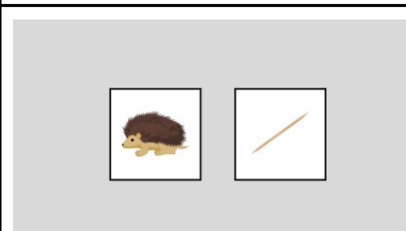
Varga, N. L., & Bauer, P. J. (2013). Effects of delays on 6-year-old children's self-generation and retention of knowledge through integration. *Journal of Experimental Child Psychology*, *115*(2), 326–341.

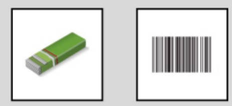
Varga, N., Esposito, A. G., and Bauer, P. J. (2019). Knowledge integration: Cognitive components and relation with measures of academic performance in children and adults. *Journal of Experimental Psychology- General.*

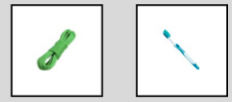
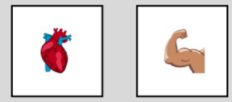
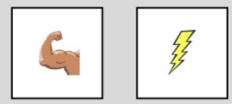
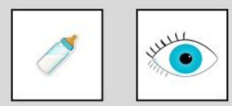
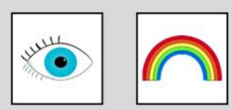
Tables and Figures

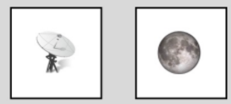
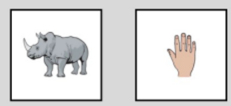
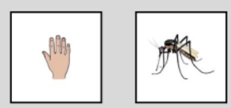
Table 1.

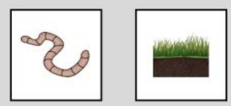
English Stem Facts with Spanish Translation and the Corresponding Visual Cues

English Stem Fact	Spanish Stem Fact	Corresponding Visual Cue
A bug that tastes with its feet is a butterfly	Un insecto que saborea con sus pies es una mariposa.	
Butterflies use taste to find leaves	Las mariposas usan sabor para encontrar hojas.	
The first American woman in space was Sally Ride	La primera mujer estadounidense en el espacio fue Sally Ride	
Sally Ride was a very good tennis player	Sally Ride era una jugadora de tenis muy buena.	
Porcupine quills were once used as toothpicks	Las púas de puercoespín alguna vez se usaron como mondadientes	

Toothpicks were the oldest dental tools	Los mondadientes eran las herramientas dentales más antiguas	
The Greeks created gum	Los griegos crearon la goma.	
Gum was the first product to have a barcode	La goma fue el primer producto en tener un código de barras	
Lice is killed by the most popular condiment	Los piojos son asesinados por el condimento más popular.	
The most popular condiment is mayonnaise	El condimento más popular es la mayonesa	
Coal is used to create nylon	El carbón se usa para crear nylon.	

Nylon was first used in toothbrushes	El nylon se usó por primera vez en cepillos de dientes	
The heart is the only muscle that never tires	El corazón es el único músculo que nunca se cansa.	
The only muscle that never tires is powered by electricity	El único músculo que nunca se cansa es motorizado por electricidad.	
A newborn baby's eyes do not have fully developed cone cells	Los ojos de un recién nacido no tienen células de cono completamente desarrolladas	
Cone cells in our eyes help tell colors apart	Las células cónicas en nuestros ojos ayudan a distinguir los colores	
Little pieces of metal make up the asteroid belt	Pequeñas piezas de metal forman el cinturón de asteroides	

The asteroid belt is located between mars and jupiter	El cinturón de asteroides se encuentra entre Marte y Júpiter	
The ocean's waves are controlled by the earth's only natural satellite	Las olas oceánicas están controladas por el único satélite natural de la tierra	
The earth's only natural satellite is the moon	El único satélite natural de la tierra es la luna	
Rhinos cover themselves with mud to protect their delicate skin	Los rinocerontes se cubren con barro para proteger su piel delicada	
Delicate skin is vulnerable to insects	La piel delicada es vulnerable a los insectos	
Cleopatra was the final pharaoh of Egypt	Cleopatra fue el último faraóna de Egipto	

The final pharaoh of Egypt used lipstick	El último faraóna de Egipto usó lápiz labial	
The slowest moving liquid is asphalt	El líquido que se mueve más lento es el asfalto	
Asphalt was used to pave roads	El asfalto se usó para pavimentar carreteras	
Worms get their food from nutrients in the soil	Los gusanos obtienen su comida de los nutrientes en la tierra	
Nutrients in the soil come from falling leaves	Los nutrientes en la tierra provienen de hojas caídas	
One of the most common dental diseases is tooth decay	Una de las enfermedades dentales más comunes es la carie dental	

Tooth Decay is prevented by chemicals in dark chocolate	La carie dental se previene con productos químicos en chocolate oscuro	 
Wallpaper was originally cleaned with playdoh	El papel pintado se limpió originalmente con playdoh	 
An important ingredient in play doh is flour	Un ingrediente importante en el playdoh es la harina	 

Table 2.

Mean and Standard Deviation of Self-Derivation Performance in each Language Condition

	Mean	*Standard Deviation*
Total E/E Open-Ended Self-Derivation	2.88	1.09
Total S/S Open-Ended Self-Derivation	2.44	0.96
Total S/E Open-Ended Self-Derivation	2.16	1.31
Total E/S Open-Ended Self-Derivation	2.28	0.89

Note: The descriptive statistics are based on N = 25

Table 3.

Mean and Standard Deviation of Self-Derivation Performance in each Language Condition based on Native Language

	Native Language	*Mean*	*Standard Deviation*
Total E/E Open-Ended Self-Derivation	English	2.69	1.18
	Spanish	3.10	0.99
Total S/S Open-Ended Self-Derivation	English	2.62	0.87
	Spanish	2.50	0.97
Total S/E Open-Ended Self-Derivation	English	2.23	1.36
	Spanish	2.30	1.25
Total E/S Open-Ended Self-Derivation	English	2.38	0.87
	Spanish	2.20	0.92

Note: The descriptive statistics are based on N = 23 (13 Native English speakers, 10 Native Spanish speakers)

Table 4.

Mean Percentage and Standard Deviation of Source Performance for Stem and Integration Facts in Each Language Condition

Panel A: Stem facts

	Mean	*Standard Deviation*
E/E	.60	.26
S/S	.63	.30
S/E	.44	.32
E/S	.52	.24

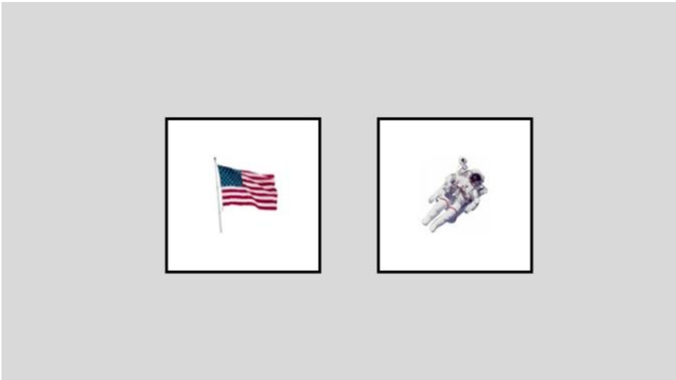
Note: The descriptive statistics are based on $N = 25$

Panel B: Integration facts

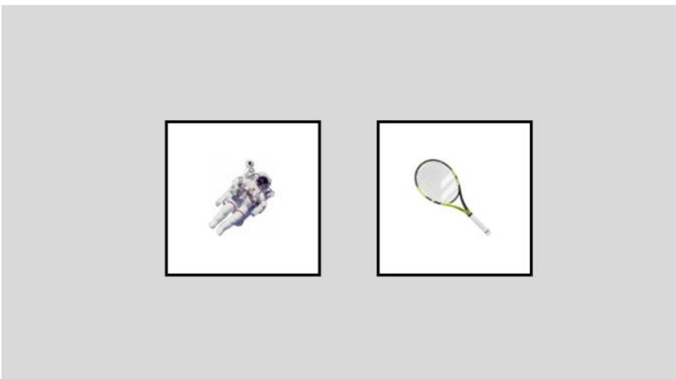
	Mean	*Standard Deviation*
E/E	.45	.35
S/S	.64	.41
S/E	.30	.36
E/S	.27	.34

Note: The descriptive statistics are based on $N = 25$

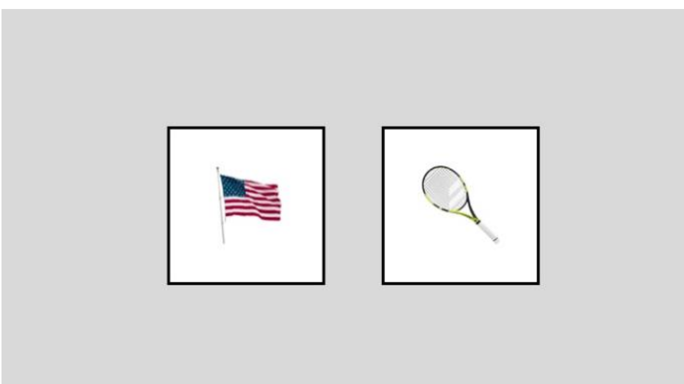
Figure 1.



Panel A: Visual cues for Stem 1 *The first American woman in space was Sally Ride.*



Panel B: Visual cues for Stem 2 *Sally Ride was a very good tennis player*.



Panel C: Visual cue for Integration fact *The first American women in space was a very good tennis player*.

Appendix A

Language Experience Questionnaire

Name _________________ **Today's Date** _____________

Date of Birth _______________ **Age:** _________

1. Please list all the languages you know in order of **dominance**:

1.	2.	3.	4.	5.

2. Please list all the languages you know in order of **acquisition**:

1.	2.	3.	4.	5.

3. When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language unknown to you.
(Your percentages should add to 100)

t your languages here:					
t your percentages here:					

4. When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak with each language?
Your percentages should add to 100)

List your languages here:					
List your percentages here:					

5. Place of birth _____________
Date of immigration to the United States (if applicable) ________________

6. Please list any college classes you have taken in any of your languages:

List your languages here:					
List your classes here:					

The following questions will refer to the language listed below.

Language: _Spanish__

This is my (native / second / third / fourth / fifth) language.

1. Age when you

Began acquiring:	Became fluent in:	Began reading in:	Became fluent reading in:

2. Please list the number of years you spent in each language environment:

	Years
A country were it is spoken:	
A family where it is spoken:	
A school or working environment where it is spoken:	

3. On a scale of zero to ten, please indicate your level of proficiency in speaking, understanding, and reading this language.

Speaking	________	Understanding	________	Reading	________

4. On a scale from zero to ten, please indicate how much of the following factors contributed to your learning of this language.

Interacting with friends	________	Language tapes / self instruction	__________
Interacting with family	________	Watching TV	__________
Reading	________	Listening to the radio / music	__________

5. In your perception, how much of a foreign accent do you have in this language:

0 1 2 3 4 5 6 7 8 9 10

none light considerable very heavy pervasive

6. Please rate how frequently other identify you as a non-native speaker based on your accent in this language:

0 1 2 3 4 5 6 7 8 9 10

never half of the time all the time

The following questions will refer to the language listed below

Language: __English_

This is my (native / second / third / fourth / fifth) language.

1. Age when you

Began acquiring:	Became fluent in:	Began reading in:	Became fluent reading in:

2. Please list the number of years you spent in each language environment:

	Years
A country where it is spoken:	
A family where it is spoken:	
A school or working environment where it is spoken:	

3. On a scale of zero to ten, please indicate your level of proficiency in speaking, understanding, and reading this language.

Speaking	________	Understanding	________	Reading	________

4. On a scale from zero to ten, please indicate how much of the following factors contributed to your learning of this language.

Interacting with friends	________	Language tapes / self instruction	________
Interacting with family	________	Watching TV	________
Reading	________	Listening to the radio / music	________

5. In your perception, how much of a foreign accent do you have in this language:

0 1 2 3 4 5 6 7 8 9 10

none light considerable very heavy pervasive

6. Please rate how frequently other identify you as a non-native speaker based on your accent in this language:

0 1 2 3 4 5 6 7 8 9 10

never half of the time all the time