

Divergence and overlap in bilingual conceptual representation: does prior language brokering experience matter?

BELEM G. LÓPEZ
University of Texas at Austin
 JYOTSNA VAID
Texas A&M University

(Received: June 26, 2014; final revision received: August 27, 2016; accepted: September 06, 2016; first published online 11 October 2016)

The present research examined whether conceptual divergence is reduced in bilinguals with extensive informal translation experience. Across two experiments, Spanish–English bilinguals (brokers vs. non-brokers) generated exemplars for 10 categories, using the same or different language across sessions. Both groups demonstrated more divergence for different than same language responses across sessions a week apart. More convergence was found in both groups for no delay compared to delayed responses. Brokers showed significantly more convergence in exemplars than non-brokers; for both immediate and delayed sessions. Findings indicate exemplars are differentially accessible depending on language and timing of response, but also individual differences in brokering experience. Extensive brokering experience may lead to a more integrated conceptual representation for features of concepts shared across languages. Findings support concept models that emphasize the dynamic and distributed nature of concepts, and underscore the need to consider the cognitive impact of systematic sources of variability among bilinguals.

Keywords: bilingual, language brokering, category structure, translation, semantic fluency

In many language contact situations across the world, bilingual children of immigrants or refugees are often enlisted as linguistic and/or cultural intermediaries (“brokers”) for family members and majority language users. There is a growing literature characterizing the nature and socio-cultural aspects of language brokering (e.g., Guan, Nash & Orellana, 2016; Morales & Hanson, 2005; Tse, 1995; Villanueva & Buriel, 2010). A separate body of research has found that formal training in translation has distinct cognitive repercussions (e.g., Garcia, Ibanez, Huepe, Houck, Michon, Lezama, Chadha & Rivera-Rei, 2014; Tzou, Eslami, Chen & Vaid, 2012). However, only a few studies to date have explored

cognitive correlates of informal translation expertise (e.g., López & Vaid, 2016; Vaid & López, 2014; Vaid, López & Martínez, 2015). The present research sought to contribute to this emerging line of inquiry by examining the impact of language brokering experience on bilinguals’ conceptual representation. In particular, it examined whether informal translation experience is associated with a greater convergence across languages in category structure.

The ability to categorize is a central aspect of human cognition, influencing how events are encoded, interpreted, and retrieved, and how new ideas are formed. Categories have a graded structure, whereby some category exemplars are considered more typical members of a category and come to mind earlier and more easily than other. The graded nature of categories has been shown to affect decision making, reasoning, memory, language, and creative thinking (e.g., Hull, Tosun & Vaid, 2016; Lichtenberk, Vaid & Chen, 2011; Ward, Kolomyts, Chu, Vaid & Heredia, 2009; Ward, Patterson, Sifonis, Dodds & Saunders, 2002).

Current models of conceptual representation assume that categories are not static and unvarying but are dynamic and malleable, depending on a number of factors, such as the language used to encode them (Malt & Majid, 2013), and the life experiences of language users,

A number of student research assistants at Texas A&M University and the University of Texas provided invaluable help in data collection and coding: Kelly Milliken, Eleazar Montes, Zaira Cortes, Katerin Cerrato, Mayra Chantal Bonilla, Estefania Lezama, Gerardo Morales Mendez, Katarina Antolovic, and Melissa Solano. We are grateful to Roberto Heredia, Anna Cieslicka and Monica Muñoz for facilitating access to participants at TAMIU, where part of the data were collected for Experiment 2. We also thank Dr. Michel Paradis, Dr. Debra Jared, and three reviewers for their comments on the manuscript. Portions of this research were presented at the 2013 meeting of the Psychonomics Society, Toronto. Preparation of this manuscript was supported in part by a Faculty Development Leave awarded to the second author by Texas A&M University.

Address for correspondence:

Belem G. López, Department of Mexican American and Latina/o Studies, University of Texas at Austin, Austin, Texas 78712
bglopez@austin.utexas.edu

which can change over time and across different contexts (see Paradis, 2004; 2014a, for further discussion). For example, Zinszer, Malt, Ameel, and Li (2014) found that Chinese–English users with more second language (L2) immersion resembled monolinguals more in their lexical categorization (naming of pictured objects) than did users with less L2 exposure. Malt, Ping, Pavlenko, Zhu, and Ameel (2015) found that bilinguals who have become more proficient in their L2 begin to lose the less frequent category labels in their L1, as compared to monolingual users of each language. Pavlenko and Malt (2011) noted that bilinguals who completely acquire their first language before being exposed to a second language are more likely to show lexical categorization convergence across their languages. Thus, research on how non-native users arrive at their choice of category names for pictured objects shows an influence of various factors, including age of L2 exposure and degree of immersion in an L2 environment.

Aside from differences in proficiency and immersion in a particular language environment, second language users and bilinguals are also known to differ in the degree to which they engage in informal translation. How might the linguistic and pragmatic skills gained through prolonged experience in informal translation, or so-called language brokering (Tse, 1995), affect the way concepts are accessed and/or represented in these bilinguals, as compared to bilinguals with little or no translation experience? Research in lexical categorization suggests that object features in categorical representation may activate more than one label name at one time and that these labels will compete in activation and later production. However, if a bilingual has had to translate across English and Spanish in order to comprehend, interpret, and reformulate the meaning of an utterance in another language, then it is possible that bilinguals with extensive translation experience (language brokers) may be able to more quickly activate and generate category exemplars than bilinguals who have not had to translate informally (non-brokers). Furthermore, the types of exemplars that brokers generate may overlap more across their two languages than those generated by non-brokers. The present research tested this issue.

Of relevance to this issue are early studies that used a word association generation paradigm (e.g., Kolers, 1963). This paradigm was seen as a way of exploring how bilinguals organize and represent their experiences in their mental lexicon. In particular, two possibilities were examined. In one view, events and experiences are thought to be encoded in a single, shared conceptual store that is equally accessed by each language. An alternative possibility is that events are encoded in the language in which they are experienced and, thus, will elicit different associations depending on the language used in a task. If experiences are tied to the language in which they were first encoded, one would expect few instances of

translation equivalents in the associations elicited to concepts presented in each language. In a study of word associations of three different groups of bilinguals, Kolers (1963) found that only a third of the responses across languages were translations of each other, lending support to this view. Other studies corroborated these findings and further showed that word association patterns across languages tended to diverge more in bilinguals who had acquired and used their two languages in different sociocultural contexts than in bilinguals who had acquired their languages in similar contexts (Berney & Cooper, 1969; Lambert & Moore, 1966).

Aside from differences arising from language use, word association patterns of bilinguals have also been found to vary as a function of word type. Concrete words typically elicit more similar associates across languages than abstract words (Kolers, 1963; van Hell & de Groot, 1998), cognates elicit more similar associates across languages than non-cognates (see Taylor, 1976; van Hell & de Groot, 1998), and words referring to certain domains (e.g., work) show less overlap across languages than words referring to other domains (e.g., family) (Berney & Cooper, 1969; Ward et al., 2009).

The issue of convergence vs. divergence in bilingual conceptual representation is also relevant when considering performance in category exemplar generation (also termed category fluency task, semantic fluency, object verbal fluency, list generation, or generative naming). This task is increasingly being used because of its clinical neuropsychological utility for diagnosing and assessing patients with dementia or aphasia in speakers of different languages. Participants are given a small set of common categories (e.g., ANIMALS, FOOD, CLOTHING) and asked to say or write down as many exemplars of those categories as they can think of in a specified interval. Differences in the number and type of exemplars produced, and in the semantic clustering of the exemplars, have been noted in relation to age, gender, educational experience, geographic location, and urban vs. rural experience, among other variables (see Pekkala, 2012, for a review).

A subset of studies of category exemplar generation have examined performance of second language users or bilinguals (e.g., Kastenbaum, 2015). In one of the earliest such studies, Roberts and LeDorze (1997) found an equal number of exemplars generated to common categories (e.g., ANIMALS) in French and in English, but the responses in French showed more semantic clustering than those in English. Only a subset of responses showed translation equivalence. Similarly, Peña, Bedore, and Zlatic-Giunta (2002) found that 5-6-year-old Spanish–English bilingual children showed only 28.2% of “doublet” (translation equivalent) responses.

The majority of studies of category listing have used a very small set of categories, typically, two or three per

study. Moreover, with the exception of the studies noted above, the majority of studies conducted with bilinguals did not systematically examine differences within the same bilinguals across their two languages. However, even in studies that did test bilinguals in both languages across two temporally separated sessions, it is not clear whether the pattern of divergence noted reflects language-related divergence or just underlying fluctuation that may have arisen even if the same language were tested at two different time points. Prior work on category exemplar listing in monolinguals (Bellezza, 1984) has shown that responses can fluctuate when participants are tested in the same language on two separate occasions separated by one or two weeks. As such, it would be important to establish whether the divergence in responses elicited to different languages presented in different sessions is greater than the fluctuation that might be expected to occur over time even within a given language. Thus, testing performance of bilinguals on the same language twice can provide a baseline measure against which to evaluate performance when different languages are used across test sessions. To date, only one previous study of bilingual category listing included this aspect in its design (Ward, Chu, Vaid & Heredia, 2005).

In their study, Ward et al. (2005) tested category exemplar generation in Chinese–English bilinguals for a set of 10 common categories that were presented twice in test sessions separated by a week. At the initial session, half of the categories were to be responded to in Chinese and half in English. In the second session, the participants were shown the same categories as before but this time half of the participants were to respond in the same language as the one used before for a given category and the other half were to respond in the other language than that used previously for a given item. Ward et al. (2005) found greater convergence in the responses in the same-language condition than in the different-language condition. The finding of greater divergence in exemplars listed when a change in response language is involved may be interpreted within the Distributed Feature Model of bilingual memory representation (de Groot, 1992; van Hell & de Groot, 1998), which posits that word meanings are represented over a network of interconnected units or features. It may also be interpreted in terms of models of conceptual representation, such as the reduced activation threshold approach, that assume that concepts are not fixed or static, but are comprised of underlying sets of features that shift as a function of experience and context (Paradis, 1997, 2007, 2014b).

The present study

The present research sought to extend the paradigm used by Ward et al. (2005) to see if there may be differences as a function of language brokering experience in the degree of

divergence noted in generating category exemplars across languages. Situating our hypothesis within the framework of the distributed feature model of bilingual memory representation (van Hell & de Groot, 1998) as well as the reduced activation threshold hypothesis of Paradis (2007), we propose that extended experience in informal translation leads to faster retrieval of underlying semantic elements that overlap across languages, reflecting a reduced activation threshold for accessing conceptual features that are shared for translation pairs with similar meanings. In particular, we suggest that the practice of informal translation experience will facilitate the retrieval of shared conceptual features of words across languages even when no translation is explicitly required, and will thus lead brokers to produce more converging category exemplars across languages than non-brokers, whose prior language experience does not lead them to seek translation equivalence.

Two experiments were designed. Both tested proficient Spanish–English bilingual adults who differed in whether they had or did not have prior brokering experience. The first experiment was aimed at testing whether the finding observed by Ward et al. (2005), of greater divergence of exemplars produced in different language than same language testing conditions across sessions, may have been due to the structural (and/or cultural) distance between the two languages of the bilinguals rather than by the fact of a shift in response language. Taylor (1976) had proposed, based on evidence from French–English speakers, that the degree of spoken or surface similarity between languages may affect the proportion of overlapping word association responses produced across languages. The languages used in our study (Spanish–English) are structurally (and culturally) more similar than Chinese–English, studied by Ward et al. (2005). Thus, if we still get the effect observed by Ward et al. (2005), it would rule out the structural/cultural distance explanation, and lend support to the shift in response language explanation for the divergence.

Our study also aimed at studying whether language brokering experience would affect the degree of divergence. We hypothesized that the degree of divergence for exemplars generated in a different language would be reduced for brokers than non-brokers. To test whether this proposed difference arises from a reorganization in conceptual representation or from a conscious strategy of looking for translation equivalents of category exemplars, we examined performance across two sessions that were either presented in immediate succession (where use of a conscious translation strategy could be expected) or separated by a week (where a group difference, if obtained, could not be attributable to a conscious strategy to seek translation equivalents of previously generated exemplars).

Experiment 1. Category exemplar generation for same vs. different language response conditions

Spanish–English bilinguals were asked to generate category exemplars to 10 categories, with half of the items tested in the same language twice and the other half tested in different languages across test sessions, which were separated by a week. We expected greater conceptual divergence in responses made in different languages than in the same language across sessions. Of additional interest was whether differences among bilinguals in prior language brokering experience would affect bilinguals' degree of divergence. Specifically, would prolonged early experience of informal translation lead to a greater conceptual overlap in the category exemplars generated across languages by brokers relative to non-brokers?

Method

Participants

A total of 125 proficient Spanish–English bilinguals were recruited from the psychology participant pool at a large southwestern university in the U.S. They were administered a detailed language background and informal translation inventory (Vaid, 2012). This instrument contained several items pertaining to the frequency and context of informal translation. For all participants, information was obtained on how often they engaged in informal translation, starting at what age, for whom, in what settings (e.g., home, school, doctors' offices, law offices, church, etc.) and for what kinds of materials (e.g., job applications, school notes, restaurant menus, etc.). Based on this information a composite measure of extent of brokering was created and individuals were classified as “brokers” if they scored high on the composite measure and as “non-brokers” if they scored low on the measure. There were 67 brokers (including 44F) and 58 non-brokers (including 38F).

Demographic profile

The mean age was 19.49 years ($SD = 2.45$) for brokers and 20.26 ($SD = 4.75$) for non-brokers. Sixty percent of the brokers and 82.8% of the non-brokers were born in the U.S. The majority of those born outside the U.S. were from Mexico. Of those brokers who were not born in the U.S. ($n = 26$), the mean age of arrival in the U.S. was 7.45 years ($SD = 4.8$); information on age of arrival of two brokers was not available. For the 9 non-brokers who were born outside the U.S. the age of arrival showed more variability (largely due to the inclusion of a few non-brokers who had recently arrived), with the mean age at arrival being 14.72 ($SD = 9.6$). All participants were enrolled at a university at the time of testing. The mean years of parental education was 11.02 vs. 10.43 years for brokers' mothers and fathers,

respectively, and 13.97 and 14.52 years for non-brokers' mothers and fathers, respectively.

Language background and use

For the majority of brokers and non-brokers, the language of instruction from elementary school through college was English. Among brokers a majority reported using Spanish when speaking with their parents (74.2% with their mother, 68.7% with their father); among non-brokers the corresponding percentages of Spanish use were 33.3% (mother) and 32% (fathers). A majority of participants in each group reported using Spanish with their grandparents ($M=79.0%$ for brokers and 61.8% for non-brokers). Both groups reported using a combination of Spanish and English when speaking with their siblings.

Proficiency

Assessment of language proficiency was based on two self-report measures. These were considered sufficient since in other research self-report measures of proficiency have been found to correlate with objective measures of proficiency (see Dunn & Fox Tree, 2009; Flege, Mackey & Piske, 2002). Mean self-ratings of proficiency were obtained separately for speaking, reading, writing and comprehension in each language, on a 7-point scale (1=not at all proficient; 7= highly proficient). These ratings are summarized in Table 1. Averaging across these ratings, the composite self-rated proficiency was fairly high in both groups: Brokers: Spanish - 6.09, English - 6.56; non-brokers: Spanish - 5.64, English - 6.66.

Participants were also asked to judge if they felt equally at ease in their comprehension of English and Spanish, or better in one language than the other. On this item, 60.2% of brokers reported they were equally at ease with their two languages, whereas 31.7% of non-brokers reported this, with 56.1% of non-brokers reporting that they felt more at ease in English than Spanish.

Materials and procedure

Participants were tested in two sessions. In each session they were given a list of 10 common categories, using the English category labels for half and Spanish for the other half. Their task was to generate as many exemplars of each category as they could think of, using the same language as the category label. Of the ten categories used, five were drawn from the list of categories used by Ward et al. (2005) – animals, breakfast foods, sports, types of music, and vegetables, and 5 were new categories – beverages, colors, holidays, moral values, and weather conditions. This was done to have a broad range of categories. Two quasi random orders were used in presenting the list, with the stipulation that the two food-related categories (vegetables and breakfast foods) would not be presented back to back.

Table 1. Mean Self-Ratings of Proficiency in English and Spanish by Group (7 pt. scale)

Experiment 1				
	Brokers (N = 67)		Non-Broker (N = 58)	
	English (SD)	Spanish (SD)	English (SD)	Spanish (SD)
Speaking	6.48 (.73)	6.33 (.94)	6.59 (.82)	5.71 (1.2)
Reading	6.64 (.64)	5.93 (1.17)	6.74 (.61)	5.74 (1.32)
Writing	6.40 (.87)	5.46 (1.24)	6.57 (.90)	5.03 (1.61)
Comprehension	6.67 (.66)	6.61 (.85)	6.76 (.51)	6.09 (1.33)
Experiment 2				
	Brokers (N = 82)		Non-Brokers (N = 71)	
	English (SD)	Spanish (SD)	English (SD)	Spanish (SD)
Speaking	6.34 (.91)	6.49 (.83)	6.59 (.85)	5.32 (1.64)
Reading	6.51 (.79)	6.13 (1.02)	6.85 (.44)	5.07 (1.62)
Writing	6.31 (1.05)	5.57 (1.46)	6.76 (.55)	4.39 (1.85)
Comprehension	6.64 (.64)	6.60 (.77)	6.84 (.47)	5.97 (1.23)

Note: Standard deviations are provided in parentheses.

The specific categories assigned to a particular language, and language order were counterbalanced across participants.

In Session 1 participants were given answer sheet packets with 5 categories listed in one language (in either English or Spanish) on one side of a response sheet and 5 different categories listed in the other language on the other side of the response sheet. Below each category label there were 12 lines and participants were instructed to write down as many exemplars as they could think of for each of the categories, responding in the same language as the category label (thus, they responded in Spanish to half of the items and in English to the other half). As in Ward et al. (205), participants were allowed to respond at their own pace and were not timed, although pilot testing showed that each trial took no longer than a minute.

Participants were brought back to the laboratory a week later and were administered the same task again using the same stimuli. However, this time about half of the participants (37 of the brokers and 29 of the non-brokers) were given the category labels in the same language as before and the remainder (30 of the brokers and 29 of the non-brokers) were given category labels in the other language than that used for those items in the previous session. Once again, participants were to respond in the language of the category label. As a result, they now responded in their other language for half of the categories. For example, if a participant had initially responded in Spanish to the category “Beverages”, he/she would either have to respond in Spanish again (Same

Language condition) or in English (Different Language Condition).

Data coding

Responses were initially screened for duplicates or irrelevant responses, which were excluded. These constituted a negligible percent of the responses. Furthermore, following the practice followed by Peña et al. (2002), items that appeared to be borrowed or code-switched items (e.g., *rock, salsa, cumbia, punk, techno*, etc. for “types of music”) were counted as being in the target language of the condition. Two bilingual raters who were unaware of the hypothesis or of the language designation of the conditions coded the data. They were given guidelines on how to perform the coding. They were to flag any occurrences of the same exemplars (or translation equivalents) across test sessions. The raters were given the option to consult a dictionary if they were unsure whether a particular response was a translation equivalent but this turned out not to be necessary as there was little uncertainty about how to code the responses. The raters’ scorings were subsequently compared with an independent scoring by the first author of a randomly selected subset of 18 participants’ responses per group. Cohen’s kappa (Cohen, 1960) showed high agreement (.95).

Data analysis: Degree of conceptual overlap

To determine whether there were differences in the degree of category overlap across sessions the Common Element Correlation (CEC) was calculated. This measure was first

developed by Bellezza (1984) and subsequently used by Ward et al. (2005). It refers to the number of category exemplars that were recalled in both sessions divided by the square root of the product of the total number of category exemplars recalled in session 1 and the total number of exemplars recalled in session 2 (CEC = # of overlapping items / $\sqrt{[(\text{total \# of items in session 1} \times \text{total \# of items in session 2})]}$). The CEC produces a geometric mean and ranges from 0 to 1; a score of 1 would indicate that the participant generated the exact same exemplars across the two testing sessions.

The CEC scores were computed for each of the four conditions (the two same language ones –SS and EE - and the two different language ones – SE and ES). These were in turn averaged to produce a composite same language score and a composite different language score per participant. Mean CEC scores averaged across the 5 categories per condition were entered into a 2 x 2 analysis of variance to examine the effects of group (broker vs. non-broker) and condition (same language vs. different language across test sessions) on degree of conceptual overlap.

Results

Average CEC scores by response language condition and group

The results from the 2 Condition (same vs. different) X 2 Group (broker vs. non-broker) analysis of variance conducted on the average CEC scores indicated a significant main effect for condition, $F(1, 121) = 142.70, p < .0001, \eta^2 = .54$. CEC scores were significantly higher for the same language condition ($M = .67; SE = .01$) than the different language condition ($M = .50; SE = .01$). See Figure 1. The main effect for group was not significant, $F(1, 121) = .037, p > .05, \eta^2 = .000$. The interaction between condition and group was also not significant, $F(1, 121) = .75, p > .05, \eta^2 = .006$.

Discussion

Our first aim in this study was to determine if the pattern of greater same-language than cross-language category overlap that was previously found for Chinese–English speakers by Ward et al. (2005) would also obtain for languages that are more similar to each other. Our results showed that this was indeed the case. Participants showed a significantly higher overlap in category exemplars when they performed the task in the same language across sessions than when they performed it in different languages across the two sessions. Thus, our results show that, even when the two languages of bilinguals are structurally fairly similar, bilinguals still show significantly greater divergence in the exemplars produced when they switch languages than when they

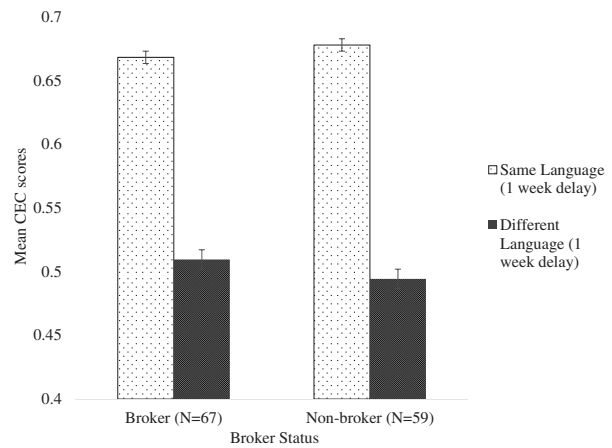


Figure 1. Mean degree of overlap in exemplars generated by brokers vs. non-brokers across test sessions for same language vs. different language response in Experiment 1 (Bars indicate standard error)

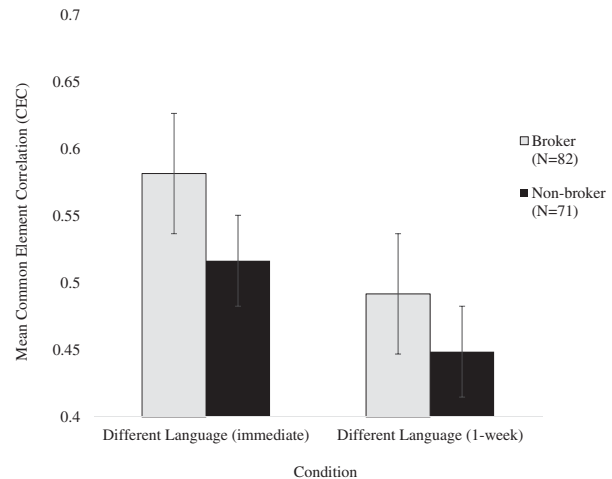


Figure 2. Mean degree of overlap in exemplars generated by brokers vs. non-brokers across test sessions for different language response in Experiment 2 (Bars indicate standard error)

stay within a given language. More generally, our findings, together with those of Ward et al. (2005) and Peña et al. (2002), allow us to conclude that the language in which category exemplars are elicited constrains the particular exemplars that come to mind.

A second question examined in the present study was whether prior experience in informal translation will be associated with a higher degree of convergence in the exemplars generated across languages. Our results do not provide support for this possibility as there was no significant effect of group nor a group by condition interaction.

It is possible that the lack of a group effect reflects aspects of our study design. In particular, the number

of data points in the critical different language condition were fairly low (based on 5 categories). The small number of categories in this condition and perhaps also the variation in response to the different categories could have reduced the sensitivity to detect any group differences. It is also possible that if the reason for a group difference is that brokers are more likely to seek out translation equivalents in the different language condition, such a strategy would not be effective when sessions are separated by a week.

To address these two issues, we conducted a second experiment. In this experiment only the different language condition was included (that is, the condition in which responses were to be in the same language across test sessions was eliminated). The same 10 categories that were used in Exp. 1 were kept in Exp. 2 but now all 10 were subjected to the response language switch, thereby increasing the potential sensitivity to detect a group difference. The other change in Exp. 2 was that participants were tested in two different intervals: one week apart (as in Exp. 1) or no delay between sessions.

Experiment 2. Category exemplar generation in immediate vs. delayed. Different language test sessions

Method

Participants

A total of one hundred and fifty-three bilingual participants (none of whom had participated in the previous experiment) were recruited from two universities in the southwestern region of the U.S., using the same criteria as in the previous experiment. They were subdivided into two groups based on their responses to questions about how often they engaged in translation, starting at what age, for whom, in what settings, and for what kinds of materials (Vaid, 2012). There were 82 brokers ($n = 62$ females) and 71 non-brokers ($n = 58$ females).

Demographic profile

The mean age of brokers was 21.74 years ($SD = 3.29$) and that of non-brokers was 23.31 ($SD = 6.66$). The vast majority of participants (76.9 % of the brokers and 90.0 % of the non-brokers) were born in the U.S. Of those not born in the U.S., the majority were from Mexico. Among brokers who were not born in the U.S. ($n = 15$), the reported mean age of arrival in the U.S. was 9.6 years ($SD = 8.39$); data on age of arrival of three brokers was not available. For the 7 non-brokers who were born outside the U.S. the age of arrival showed more variability (largely due to the inclusion of a few non-brokers who had recently arrived), with the mean age at arrival being 11.76 ($SD = 6.53$). Brokers and non-brokers were enrolled at university at the time of testing. The mean years of education for the

parents of brokers was 10.17 (mothers) and 10.85 years (fathers), and that for non-brokers was 12.72 (mothers) and 12.45 years (fathers).

Language background and use

For the majority of brokers and non-brokers, the language of instruction from elementary school through college was English. Among brokers a majority reported using Spanish when speaking with their parents (81.5% - mother, 73.4% - father); among non-brokers the corresponding percentages of Spanish use were 28.2 % (mother) and 35.7% (father). A majority of participants in each group reported using Spanish with their grandparents (93.8% of brokers and 70.0 % of non-brokers). Both groups reported using a combination of Spanish and English when speaking with their siblings.

Proficiency

Assessment of language proficiency was based on two self-report measures. As noted earlier, self-report measures of proficiency have been found to correlate with objective measures (see Dunn & Fox Tree, 2009; Flege et al., 2002). Mean self-ratings of proficiency in speaking, reading, writing and comprehension, on a 7 point scale (1=not at all proficient; 7= highly proficient) are summarized in Table 1. A composite score obtained by averaging across the ratings showed an average rating for brokers of 6.20 for Spanish and 6.44 for English; non-brokers' composite score was 6.77 for English and 5.23 for Spanish. With respect to self-reported ease of comprehending each language, 25% of the brokers and 20.3% of the non-brokers reported being equally at ease with each language. In addition, 40.7% of brokers, as compared to 62.3% of non-brokers reported being more at ease with English than Spanish.

Materials and procedure

Participants were recruited from two universities in the southwestern region of the U.S. and were assigned to one of two delay conditions: no delay ($n = 90$) and 1-week delay ($n = 63$). For the no delay condition, participants were tested in groups in a classroom setting. Half of the classroom was arbitrarily designated as the Spanish group (Spanish first, $n = 43$) and the other as the English group (i.e., English first, $n = 47$).³ Participants in the English first group were to perform the task in English and those in the Spanish first designated group were to respond in Spanish. Participants were given a list of 10 categories, presented with category labels in the designated language of the condition (these were the same 10 categories used in Exp. 1) and were to come up with as many exemplars as they could for each category in that language. They were allotted 30 seconds per category to write down their responses. After completing responses to all 10 categories, response sheets were collected. Participants were then

asked to do the task again, for the same 10 categories presented in the same order, but were to respond in the other language than before. Participants were unaware that they would have to repeat the task.

For those in the 1-week delay condition ($n = 63$), testing was done individually or in small groups in a laboratory setting. Participants were randomly assigned to a language order, English first ($n=35$) or Spanish first ($n=28$). A week later, participants were called back and were shown the same 10 categories in the same order as before, but presented in the other language, and participants were to generate category exemplars in that language. In both sessions, 30 seconds was allotted for writing responses per trial. Upon completion of the second session, participants completed the language brokering and background questionnaire.

Data coding

The same criteria as used in the previous experiment were used to code the data in this experiment. Two bilingual raters who were unaware as to the underlying hypothesis coded the responses across sessions to identify translation equivalents. They were allowed to consult a dictionary if they were unsure whether a particular response was a translation equivalent but in practice this was not necessary as translation status was easy to determine. As in the previous study, the raters' scorings were subsequently compared with an independent scoring by the first author of a randomly selected subset of 18 participants' responses per group. Cohen's kappa (Cohen, 1960) again showed high agreement (.98) indicating that the scoring procedure was very reliable.

Results

Average CEC scores by condition and group

To determine whether there were differences in the degree of category overlap across sessions, Common Element Correlation (CEC) scores were calculated (following Ward et al., 2005), averaged across the categories, and were entered into a 2(Condition: 1-week delay vs. immediate) X 2(Group: broker vs. non-broker) analysis of variance by-participants, and a separate one by-items.

The by-participant analysis showed a main effect of broker status, $F(1, 149) = 7.91, p = .008, \eta^2 = .95$. Brokers ($M = .55; SD = .11$) had significantly higher CEC scores than non-brokers ($M = .49; SD = .12$). The main effect for condition was also significant, $F(1, 149) = 15.26, p = .0001, \eta^2 = .09$, indicating that the no delay condition produced higher category overlap than the one week delay condition. The interaction between condition and group was not significant, $F(1, 149) = .30, p > .05, \eta^2 = .002$. The by-item analysis showed a near significant effect of broker status, $F(1,9) = 4.08, p = .074, \eta^2 =$

.312, no effect of condition ($p > .05$) and no interaction ($p > .05$).

Discussion

The present experiment used the same category stimuli as were used in Experiment 1, with a new set of participants, and looked only at the different language condition. The question of interest was whether brokers would show more overlap in category structure across languages as compared to non-brokers, and whether this effect would interact with the spacing of the two test sessions (immediate vs. one-week delayed).

Our results showed the hypothesized effect of brokering experience: brokers showed significantly greater overlap in their category exemplars produced in different languages than did non-brokers (although this group effect did not reach an acceptable level of significance in the by-items analysis). Furthermore, there was more overlap in responses when the test sessions were presented with no delay than when they were separated by a week. Importantly, the effect of group did not interact with the effect of test session timing: thus, brokers showed a larger overlap in category structure across languages than non-brokers, regardless of whether the language switch was within the same day (immediate succession) or spanned a week.

General discussion

The present research aimed at examining whether individual differences in early language brokering experience of Spanish–English bilinguals affect how conceptual features associated with words in their two languages may be accessed. To address this issue we compared the relative degree of convergence in category exemplars generated by bilinguals when performing the task in the same language across two test sessions than when performing it in different languages (Exp. 1). We also compared the relative degree of convergence in responses across languages by brokers and non-brokers under immediate vs. delayed presentation of the switched language condition (Exp. 2).

Our findings in Experiment 1 replicated the finding reported previously with Chinese–English bilinguals (Ward et al., 2005) of a higher overlap in category exemplars listed when the same language was used across two test sessions separated by a week than when different languages were used across sessions. That is, we found that when participants were tested twice in Spanish or twice in English, they produced more overlap in exemplars than when they were tested once in Spanish and once in English. The fact that this pattern was obtained not only in languages that are structurally dissimilar (Chinese vs. English), but also in languages as structurally close

to each other as Spanish and English suggests that language distance is not critical for this effect to emerge. Of further note was the finding that the same-language convergence advantage was equally robust in bilinguals with prior translation experience and those without such experience.

The novel aspect of our study, not previously examined in research on categorization, was the question of individual differences in category structure related to prior language brokering experience. We reasoned that language brokering experience would facilitate the activation of conceptual features that overlap for translation equivalents of category exemplars. As such, brokers were expected to generate a greater number of translation equivalents than non-brokers. Although our first experiment did not show a group difference, a significant group effect did emerge in the second experiment.

The lack of a group effect in our first experiment was due, we believe, to insufficient sensitivity in that experiment, given that there were half as many data points in the critical different language condition as there were for that condition in Experiment 2. That is, responses in the different language condition in Experiment 1 were based on data from only five categories whereas responses in Experiment 2 were based on 10 categories.

Importantly, the group effect obtained in Experiment 2 was independent of an effect of test session timing: brokers showed greater cross-language convergence in their generation of category exemplars not only when they were tested immediately in the other language, but also when the other language condition was delayed by one week. This result suggests that the mechanism underlying the observed effect of brokering on category structure may be at the level of representation rather than reflecting a translation strategy effect (e.g., brokers actively look for translation equivalents of their previously generated exemplars). That is, our findings suggest that prolonged experience with translation may effect a change in conceptual organization in the form of a closer linking of cross-language equivalents in the mental lexicon of bilingual brokers than non-brokers.

Our finding that brokers are more likely to show convergence in their selection of exemplars across languages, whereas non-brokers are less likely to do so is compatible with other recent studies that have examined effects associated with language brokering. For example, one study found that Spanish–English brokers were significantly faster than non-brokers at verifying translations of idiomatic phrases (Vaid & López, 2014). Another study found that Spanish–English brokers were equally fast at making semantic relatedness judgments for phrases when a target word was presented in the same language or different language than the language

of the phrase; by contrast, non-brokers showed a same-language preference in their response patterns (López & Vaid, 2016).

We offer the following interpretation of how brokering experience may lead to the effects observed in the present study. Extensive early informal experience in translation results in a reduced threshold for activation of semantic features that overlap in translation equivalents of words or phrases. Thus, not only are explicit translation skills facilitated by early experience in informal translation, making brokers more adept at making judgments of translation equivalence (as noted by Vaid & López, 2014; see Tzou, Vaid & Chen, 2016, for a similar effect noted in bilinguals with formal training in translation), but the extensive practice in verifying and generating translations may also lead brokers to be more attuned to meanings that are shared across languages, EVEN WHEN THE TASK DOES NOT EXPLICITLY REQUIRE THEM TO ESTABLISH EQUIVALENCE. In the category exemplar task, although translation equivalence is not an explicit task demand, brokers – being much more practiced than non-brokers at translating – are more attuned to retrieving category exemplars whose underlying semantic elements (or conceptual features) overlap across the category label equivalents in their two languages than retrieving exemplars that have less shared overlap. By contrast, non-brokers, who lack this accumulated practice, may have a more elevated threshold of activation of conceptual features associated with exemplars that overlap across translation equivalent category labels (Paradis, 2014b). Further work should be designed to test this proposed mechanism using other tasks and more on-line methods, such as the visual world paradigm.

Since our study was not designed with a priori hypotheses about differences due to the individual categories for which exemplars were elicited, we do not report the data for individual categories, although pilot analyses done for each category revealed a range of overlap across test items. This variability may have contributed to greater noise in the by-item analysis, and the failure for a group effect to emerge (although there was a trend in the expected direction).

In light of previous work suggesting that the range and frequency of associations elicited in response to a given category may vary depending on the category domain (Berney & Cooper, 1969), it will be important to consider category type effects in future work. For example, one might expect that certain categories (e.g., breakfast foods) may evoke more culture-specific (and perhaps also more language-specific) primary associations than other categories (e.g., weather conditions) (Pekkala, 2012). In addition, responses may well differ depending on how participants self-identify culturally. As Kolers (1963, p. 299) noted, “one who thinks of himself as a German living in America is likely to give different responses

from one who thinks of himself as an American born in Germany". It is possible that differences observed in the present study between brokers and non-brokers may in part reflect differences in cultural self-identification (see Vaid, 2006, for an effect of cultural self-identification on bilinguals' humor preferences). It would be interesting to relate measures of self-identification to the types of responses elicited to categories that are likely to be more culture-laden as compared to those that are more neutral.

Another manipulation that could be informative in future work is to consider possible group differences in the treatment of goal-based categories, such as "items to pack for an overseas trip". To the extent that such categories would need to be constructed "on the fly" as compared to categories for which there are pre-established, stored exemplars, one might expect that experience-based factors aligned with group differences in language use may affect responses to the former type more.

In summary, our research allows the following conclusions. First, bilingual speakers show language-specific associations to categories: they tend to produce a certain set of exemplars when operating in a given language and a somewhat different set of exemplars to the same concept when asked to respond in a different language. Second, our study showed that differences in a particular language practice among bilinguals (informal translation) differentially affects the pattern of category exemplars evoked by category labels. Bilinguals with extensive translation experience show more overlap in the exemplars they generate to a concept when responding in different languages than do bilinguals with less or no translation experience, who show greater divergence in their pattern of response across languages to a given concept. Moreover, this effect is not simply due to brokers actively looking for translation equivalents of the exemplars they generated in an earlier session, since the effect persists even when they perform the second task a week after the first one, and, thus, could not have remembered the particular exemplars they selected in the initial session. Thus, it would seem that brokering experience may alter the pattern of activation of semantic features that are shared by concepts, regardless of the particular language of the concept label. As such, our study provides the first empirical demonstration of an effect of language brokering in the domain of concept representation. Our study suggests that brokering experience may result in making translation equivalents more readily accessible even when the task at hand does not require invoking translation equivalents.

Although our study was not designed to provide a test of different models of bilingual lexical/semantic or conceptual representation, we believe that the theoretical framework proposed by van Hell and de Groot (1998) of a distributed feature model – rather than a hierarchical model – offers a more parsimonious account. The

conceptual feature model proposed by Paradis (2004, 2014a), which shares many properties with the distributed feature model, is also a relevant theoretical framework for our findings. In these accounts, extensive prior experience in informal translation may be said to facilitate retrieval of shared semantic elements (or shared conceptual features) associated with category exemplar translation equivalents (Paradis, 2004, 2014a; van Hell & de Groot, 1998). However, whereas van Hell and de Groot's (1998) study suggested that concrete words have more overlapping underlying semantic elements than abstract words, our study (which contained several categories that could be considered to have concrete exemplars) suggests that concrete words themselves may differ in their degree of shared features across languages. A stronger convergence of exemplars generated when the task is performed in the same language than in different languages would arise in these frameworks because same language exemplars presumably share a higher number of overlapping features.

Furthermore, our findings may also be theorized in terms of the notion of the activation threshold hypothesis discussed in the first language attrition literature (e.g., Paradis, 2007). In this literature it is noted that frequency of use affects the ease of access to a word; the less often a word is used the higher the activation threshold for retrieving conceptual features associated with the word's meaning. In our study, the finding of greater cross-language convergence obtained among brokers than non-brokers for the no-delay condition may be interpreted to suggest that skill in seeking out meaning equivalence (which is arguably the essence of brokering experience) leads brokers to think more readily of items that are translation equivalents, that is, items that contain more shared conceptual features across languages, thereby providing more chances of lowering the activation threshold of those features, leading to easier access to the translations (see Paradis, 1997, 2004, 2007). This facilitation would be enhanced in childhood among brokers as they have to constantly search for those words whose conceptual representations share the greatest number of features. Non-brokers would not be called upon to engage in this task and their performance thus remains within language, keeping the activation threshold of the non-selected language high (Paradis, 2014b). In conclusion, our findings suggest that extensive prior experience in informal translation leads to greater cross-language overlap in the category exemplars generated across languages and this most likely reflects faster access to similar feature groupings in the other language, as a result of the frequency of performing informal translation over a prolonged period of time.

The present study adds to the body of work that suggests that language brokering experience affects how words in the two languages may be accessed and represented. In further research it will be important

to examine such issues as the amount or length of brokering experience that is sufficient to produce an effect. A broader contribution of our study is to show that individual differences in bilingual language experiences have cognitive repercussions and, thus, that all bilinguals should not be expected to behave alike. Researchers in bilingualism are increasingly arguing for the need to consider the context of language use when studying bilinguals, rather than treating all bilinguals as an undifferentiated group to be compared to monolinguals (see Cook, 1991; Genesee, 2014; Green, 2014; Grosjean, 1989). In line with other calls in the field (e.g., Baum & Titone, 2014; Vaid, López & Martinez, 2015; Vaid & Meuter, 2016), our study demonstrates the importance of investigating variation in language experience WITHIN bilinguals, and argues for a shift away from monolingual vs. bilingual comparisons in favor of a more systematic examination of sources of variation among different bilingual subgroups.

References

- Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics* 35, 857–894.
- Bellezza, F. S. (1984). Reliability of retrieval from semantic memory: common categories. *Bulletin of Psychonomic Science*, 22(4), 324–326.
- Berney, T., & Cooper, R. (1969). Semantic independence and degree of bilingualism in two communities. *The Modern Language Journal*, 53 (3), 182–185.
- Cohen, J. (1960). A coefficient for agreement for nominal scales. *Educational and Psychological Measurement*, 20, 687–699.
- Cook, V. (1991). The poverty of the stimulus argument and multicompetence. *Second Language Research*, 7, 103–117.
- Dunn, A. L., & J. E. Fox Tree (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition*, 12(3), 273–289.
- Flege, J. E., Mackay, Ian R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, 23, 567–598.
- García, A., Ibanez, I., Huepe, D., Houck, A., Michon, M., Lezama, C. G., Chadha, S., & Rivera-Rei, A. (2014). Word reading and translation in bilinguals: the impact of formal and informal translation expertise. *Frontiers in Psychology*, 5, Article 1302, 1–14.
- Genesee, F. (2014). Commentary on “Moving toward a neuroplasticity view of bilingualism”: The early years. *Applied Psycholinguistics*, 35, 905–909.
- Green, D. W. (2014). Individual variability and neuroplastic changes. *Applied Psycholinguistics*, 35(5), 910–912.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1), 3–15.
- Guan, S-S A., Nash, A., & Orellana, M. F. (2016). Cultural and social processes of language brokering among Arab, Asian, and Latin immigrants. *Journal of Multilingual and Multicultural Development*, 37(2), 150–166.
- Hull, R., Tosun, S., & Vaid, J. (2016). What’s so funny? Modelling incongruity in humor production. *Cognition and Emotion*. <http://dx.doi.org/10.1080/02699931.2015.1129314>
- Kastenbaum, J. (2015). *The influence of proficiency and language combination on bilingual lexical access*. M.Sc. thesis, Boston University.
- Kolers, P. A. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior*, 2, 291–300.
- Lambert, W. E., & Moore, N. (1966). Word association responses: Comparisons of American and French monolinguals with Canadian monolinguals and bilinguals. *Journal of Personality and Social Psychology*, 3(3), 313–320.
- Lichtenberk, F., Vaid, J., & Chen, H-C. (2011). On the interpretation of alienable and inalienable possession: A psycholinguistic investigation. *Cognitive Linguistics*, 22(4), 659–689.
- López, B. G., & Vaid, J. (2016, under review). *Facil or A piece of cake*: Does variability in bilingual language brokering experience affect idiom comprehension? *Bilingualism: Language and Cognition*.
- Malt, B., & Majid, A. (2013). How thought is mapped onto words. *WIREs Cognitive Science*, 4(6), 583–597.
- Malt, B. C., Ping, L., Pavlenko, A., Zhu, H., & Ameer, E. (2015). Bidirectional lexical interaction in Mandarin-English bilinguals. *Journal of Memory and Language*, 82, 86–104.
- Morales, A., & Hanson, W. E. (2005). Language brokering: An integrative review of the literature. *Hispanic Journal of Behavioral Sciences*, 27 (4), 471–503.
- Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic approaches* (331–354). Mahwah, NJ: Erlbaum.
- Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.
- Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Köpcke and B. Schmid (eds.), *First language attrition* (pp. 121–134). Amsterdam: John Benjamins.
- Paradis, M. (2014a). The nature of concepts. Unpublished manuscript, McGill University.
- Paradis, M. (2014b). Personal communication, November 19, 2014.
- Pavlenko, A., & Malt, B. C. (2011). Kitchen Russian: Cross-linguistic differences and first- language object naming by Russian-English bilinguals. *Bilingualism: Language and Cognition*, 14(01), 19–45.
- Pekala, S. (2012). Verbal fluency tasks and the neuropsychology of language. In M. Faust (Ed.), *The handbook of the neuropsychology of language, Volume 1&2* (pp. 619–634). Oxford, U.K.: Wiley-Blackwell.
- Peña, E., Bedore, L., & Zlatic-Giunta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language, and Hearing Research*, 45, 938–947.

- Roberts, P. M., & LeDorze, G. L. (1997). Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency. *Brain and Language*, 59(3), 412–449.
- Taylor, I. (1976). Similarity between French and English words—a factor to be considered in bilingual language behavior? *Journal of Psycholinguistic Research*, 5(1), 85–94.
- Tse, L. (1995). Language brokering among Latino adolescents: Prevalence, attitudes, and school performance. *Hispanic Journal of Behavioral Sciences*, 17(2), 180–193.
- Tzou, Y-Z., Eslami, Z., Chen, H-C., & Vaid, J. (2012). Effect of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: Evidence from Mandarin-English speakers. *International Journal of Bilingualism*, 16, 213–227.
- Tzou, Y-Z., Vaid, J., & Chen, H-C. (2016). Does formal training in translation/interpreting affect translation strategy? Evidence from idiom translation. *Bilingualism: Language and Cognition*, 1–10. DOI: [10.1017/S1366728915000929](https://doi.org/10.1017/S1366728915000929)
- Vaid, J. (2006). Joking across languages: Perspectives on humor, emotion, and bilingualism. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 152–182). Clevedon, U.K.: Multilingual Matters.
- Vaid, J. (2012). *Language proficiency and brokering questionnaire*. Texas A&M University.
- Vaid, J., & López, B. (2014, Nov.). *Informal translation experience facilitates translation verification of idiomatic phrases*. Poster, Psychonomics Society, Long Beach, CA.
- Vaid, J., López, B., & Martinez, F. (2015). Linking the figurative to the creative: Bilinguals' comprehension of metaphors, jokes, and remote associates. In R. R. Heredia and A. Cieslicka (Eds.), *Bilingual figurative language processing* (pp. 53–86) New York: Cambridge University Press.
- Vaid, J., & Meuter, R. (2016). Not through a glass darkly: Refocusing the psycholinguistic study of bilingualism through a 'bivocal' lens. In Cook, V., & Wei, L. (Eds.), *Cambridge handbook of linguistic multicompetence* (pp. 50–76). Cambridge, U.K.: Cambridge University Press.
- Van Hell, J., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3), 193–211.
- Villanueva, C. M., & Buriel, R. (2010). Speaking on behalf of Others: A qualitative study of the perceptions and feelings of adolescent Latina language brokers. *Journal of Social Issues*, 66(1), 197–210.
- Ward, T. B., Chu, A., Vaid, J., & Heredia, R. H. (2005). Divergence and overlap in bilingual concept representations. In B. Bara, M. Bucciarelli, & I. Barsalou (Eds.), *Proceedings of the 27th annual meeting of the Cognitive Science Society* (pp. 2342–2346). Mahwah, NJ: Lawrence Erlbaum.
- Ward, T. B., Kolomyts, Y., Chu, A. H., Vaid, J., & Heredia, R. H. (2009). Graded category structure in Chinese–English bilinguals. *The International Journal of Creativity & Problem Solving*, 19 (2), 47–59.
- Ward, T. B., Patterson, M., Sifonis, C., Dodds, R., & Saunders, K. (2002). Graded category structure and imaginative thought. *Memory & Cognition*, 30(2), 199–216.
- Zinszer, B., Malt, B., Ameel, E., & Li, P. (2014). Native-likeness in second language lexical categorization reflects individual language history and linguistic community norms. *Frontiers in Psychology*, 5, 82–96.