Emotion–memory effects in bilingual speakers: A levels-of-processing approach

Ayse Ayçiçegi–Dinn
Istanbul University

Catherine L. Caldwell–Harris
Boston University

Address for Correspondence

Catherine L. Caldwell–Harris, Ph.D.
Psychology Department
Boston University
64 Cummington St.
Boston, MA 02215
USA Email: charris@bu.edu

RUNNING HEAD: Emotion–memory effects

Abstract

Emotion-memory effects occur when emotion words are more frequently recalled than neutral words. Bilingual speakers report that taboo terms and emotional phrases generate a stronger emotional response when heard or spoken in their first language. This suggests that the basic emotion-memory will be stronger for words presented in a first language. Turkish-English bilinguals performed a deep processing task (emotion intensity rating) or shallow processing task (counting letter features) and two additional deep processing tasks (translation and word association) on five categories of words (taboo words, reprimands, positive words, negative words, and neutral words), followed by a surprise recall task. Reprimands had the highest recall in English (L2), which may be a novelty effect. If reprimands are set aside, then overall emotion-memory effects were similar in the two languages, with taboo words having the highest recall, followed by positive words. Negative words had no recall advantage over neutral words, an unexpected finding. Results indicate that emotional attributes of words are equally available to boost memory in a first and second language in both shallow and deep processing tasks, athough some task-specific effects did occur. (183 words)

Emotion–memory effects in bilingual speakers: A levels-of-processing approach

Ayse Ayçiçegi–Dinn
Istanbul University

Catherine L. Caldwell–Harris
Boston University

The current paper unites two phenomena. The first is that bilingual speakers report experiencing greater emotion when using their first language compared to their second, especially for topics of personal significance (Pavlenko, 2005; Marian and Kaushanskaya, 2004). For example, in psychotherapy, the first language is frequently preferred for emotional topics and the second language for maintaining an aura of emotional distance (Altarriba and Santiago-Rivera, 1994; Schrauf, 2000). Anxiety-arousing stimuli such as taboo and sexual references are more easily expressed in the second language (Dewaele, 2004a; Dewaele, 2004b; Gonzalez-Regiosa, 1976; Harris, Ayçiçegi and Gleason, 2003). The second phenomenon is emotion words' superior recall to neutral words (Rappaport, 1942; Reber, Perrig, Flammer, and Walter, 1994; Rubin and Friendly, 1986). Referred to as the emotion memory effect, it is observed when participants perform a task on a list of words, and are later asked to recall as many words from the prior task as possible. Emotion words are proposed to have superior recall to neutral words because they automatically elicit a deeper level of processing than neutral words due to their inherent interest or because emotional stimuli recruit attention (Jay, Caldwell-Harris, and King, 2008 ; MacKay and Ahmetzanov, 2005).

A combination of these two phenomena (L1 frequently experienced as more emotional than L2; emotion words recalled better than neutral words) suggests that emotion-memory effects will be stronger in a first language than in a second language. Anooshian and Hertel (1994) pursued this exact question and found the predicted effect: emotional memory effects were present in L1 but absent in L2. In their study, balanced Spanish-English bilingual participants rated neutral and emotion words for emotional intensity and then performed a surprise recall task. Emotional-memory effects were obtained by subtracting percent recall for emotion words from recall of the neutral words.

Conflicting findings have been obtained on emotion memory effects in bilingual speakers. Ayçiçegi and Harris (2004) attempted to extend Anooshian and Hertel's findings to individuals who had initially learned English as a foreign language. Participants were Turkish-English

bilingual speakers who had arrived in the U.S. after age 17, judged Turkish to be their superior language, and rated their English proficiency as fair to very good. They had lived in the US for an average of 5 years. Participants heard or read on a computer screen five types of emotion stimuli: positive, negative, and neutral words, mild-to-moderate taboo words (e.g., breast, whore), and childhood reprimands (short phrases such as "Shame on you!" and "Go to your room!"). Strong emotion-memory effects were obtained in both a first and a second language. Indeed, emotion-memory effects were stronger in English (L2) than in Turkish, particularly for the taboo words and reprimands.

Finding stronger emotion-memory effects in L2 was surprising, because the same taboo words and reprimands had elicited larger skin conductance responses in Turkish-L1 compared to English-L2 (in a prior study, also using Turkish-English bilingual speakers who were residing in the U.S. (Harris, Ayçiçegi and Gleason, 2003). Two additional studies that included more diverse types of emotional phrases confirmed that emotional phrases presented in a first language elicited higher skin conductance responses than emotional phrases in a second language, for Turkish-English bilinguals residing in Istanbul (Caldwell-Harris and Ayçiçegi -Dinn, under review) and Spanish-speaking immigrants residing in the U.S. (Harris, 2004). The physiological data is thus in line with the view that greater emotion is elicited by the first language. Why do the recall studies of Anooshian and Hertel (1994) and Ayçiçegi and Harris (2004) report conflicting findings? The following possibilities are explored in the current paper.

English-L2 more highly activated due to immersion context. Strong emotion-memory effects for English-L2 may have occurred because English was generally activated to a higher level due to participants' immersion in an English-speaking environment at the time of testing (see Green, 1986). Many were permanent or long-term residents of North America. Would different results be obtained if bilingual speakers were tested while residing in their country of origin?

Some English-L2 items were processed more deeply during the initial emotional-intensity rating phase. Deeper processing leads to enhanced recall, as demonstrated by research on depth-of-processing effects (Craik and Lockhart, 1972). English emotional stimuli could have attracted deeper processing than neutral stimuli in the study by Ayçiçegi and Harris (2004) for a number of reasons. For example, the English taboo words and reprimands may have seemed more novel or amusing than their Turkish equivalents, and thus they could have been processed more deeply.

Alternatively, the participants may have been silently translating English emotion words into Turkish to decide on a precise rating. They may have found it unnecessary to translate English neutral words, recognizing that they had low emotionality. Since translation is effortful, the English emotional words would have benefited from the additional depth of processing that accompanies translation. This would have selectively boosted recall for English emotional stimuli. Or, more generally, the need to judge emotional intensity for English emotional words could have been more effortful than for other stimuli, thus increasing processing effort and boosting recall. The need to translate or indeed to perform any deep processing could be avoided by asking participants to perform a shallow processing task, such as processing visual features of the display.

Varying the incidental study task is a method for controlling how deeply participants process the stimuli. In depth-of-processing studies (e.g., Craik and Lockhart, 1972; Jay et al., 2008), deep processing is elicited by asking participants to perform tasks which require processing words for their meanings, such as verifying whether the word is a congruent completion to a sentence template (*The _____ is blue)* or making judgments about categorization (*Is this a body part*?). Tasks that require only shallow processing include verifying whether a word is in upper or lower case, or answering questions involving the letters in the word (e.g., Fletcher, Stephenson, and Carpenter, 2003; Vilkki, 1987).

No prior research has investigated whether depth of processing effects vary for a first vs. a second language. We thus investigated two additional tasks, translation and word association, which have been used in bilingual research (Berny, Cooper, and Fishman, 1968; de Groot, 1992), but not in a depth-of-processing/recall framework. These tasks will likely facilitate deeper processing in the L1 than in the L2, as explained below. If superior recall is nonetheless obtained for L2 words, this superior recall can be interpreted as an emotion-memory effect or a novelty/interest effect.

Our goals in the current study were thus to investigate emotional memory effects using a broader range of incidental-study tasks than has been used previously, and to utilize college students who have acquired English as a foreign language while residing in the country of their first language. Participants were randomly assigned to one of the following four study conditions.

1. Rate the word for emotional intensity, which is the task used in Anooshian and Hertel, (1994), and Ayçiçegi and Harris (2004).

2. Shallow processing task: How many letters contain a closed circle? (p, q, o, a, b, d, g).

3. Translation task: If the word appears in Turkish, translate and report it to the experimenter in English; If in English, translate and report it in Turkish.

4. Word association task. Provide as many word associates as come to mind in 10 seconds.

The motivation for the four tasks is outlined below.

Ratings for emotional intensity. This is the basic task used in prior studies, and thus will serve to replicate Ayçiçegi and Harris (2004), with the difference that the participants are immersed in the country of the first language. If emotion memory effects are found to be stronger in L1 (Turkish), then we conclude that obtaining stronger emotion-memory effects for L1 may require testing in the country of L1.

Letter counting task. Word recall following a shallow task such as the letter-counting task is expected to be poorer than for a task that requires semantic processing. We thus expect lower recall in this task for both L1 and L2, compared to the emotional intensity task. Since participants can perform the letter-counting task without attending to the meaning of the stimuli, we expect covert translation to be absent, and any obtained emotion-memory will reflect automatic or incidental processing. At present, the main task that has been used to compare automatic processing in a first vs. a second language is the Stroop task, where the meaning of words is generally found to be processed automatically, even if the task is to attend words printed color. In some studies with bilingual speakers, Stroop interference is reduced when color words appear in participants' second language (Zied, Philippe and Karine, 2004). However, some studies find equivalent Stroop interference in both languages (Eilola, Havelka, and Dinkar, 2007; Sutton, Altarriba, Gianico, and Basnight-Brown, 2007). Consciously controlling Stroop interference may be more difficult in a second language (Tzelgov, Henik, and Leiser, 1990). It is thus currently unclear whether equivalent encoding of words will occur when instructions do not require attention to word meaning.

Translation task. Translation is a "deep processing" task in that it requires attending to meaning. It is probably more effortful than emotion-intensity rating and we thus expect higher recall rates for this task than the emotion-rating and letter-counting tasks. However, because translation may be more difficult for some words than others, recall may be enhanced if deeper

processing results for some words compared to other words, depending on translation difficulty. Bilingual researchers generally assume that translating from L2 to L1 is the easier translation direction, because during initial learning, L2 words may be learned as direct translations of L1 words (Kroll and Stewart, 1994). Lexical access may be more efficient in L1 and slower or more arduous in the L2 (being the less proficient language). Translating from L2 means that one doesn't have to perform lexical access in the less proficient language. Greater translation difficulty for L1-> L2 could therefore result in deeper processing for the L1 words, leading to higher recall. Therefore, if the translation-task condition yields a finding of greater emotion-memory effects in L2 (as occurred in Ayçiçegi and Harris, 2004), this may represent a genuine example of L2 memorability, rather than being a by-product of deep processing following translation.

Word association task. A fourth task was chosen which, like the translation task, is expected to require deeper processing than the emotion-rating and letter-count tasks, and may also stimulate deeper processing in the L1. Bilingual speakers can generate a greater number of lexical associations in their more proficient language (Kroll and Stewart, 1994). (Indeed, as described below, the current participants generated more words when asked to perform a word fluency task in their L1, Turkish, than in their L2, English). We thus expected the word association task to generate more words for Turkish items, and that this would enhance recall for the Turkish items. Consequently, a recall advantage for L2 over L1 would need to be attributed to factors other than depth of processing engendered by task demands during the study period.

Method

Participants

Native speakers of Turkish were recruited from among students at Istanbul University and the surrounding community (N=59, 88% female). Participants were recruited by a teacher who knew them to have high levels of English ability, in order to make the participants comparable in English ability to those in our prior study, who had been residing in the U.S. for an average of 5 years.

Demographic variables appear in Table 1. All participants identified themselves as native speakers of Turkish and reported Turkish to be their dominant language. Participants began regular English study (defined as at least 5 hours per week) when they entered middle school at

age 12 or 13. Participants rated their English fluency in conversation, reading, understanding and writing on a 1 to 5 scale corresponding to some, fair, good, very good, native-like ability. All participants rated their English as either good or very good for understanding, and most rated themselves as good or very good for speaking, reading and writing. As a more objective measure of Turkish and English abilities, we administered a word fluency task in both languages (Delis, Kaplan and Kramer, et al., 2001; Rosselli and Ardila, 2002). These required participants to generate as many words as they could beginning with the letters F, A and S (FAS task). The mean number of words generated in English was 29 (range 18-49), and in Turkish, 36 (range 22-53), a statistically significant difference, $F(1,54)=35$, $p < .0001$. Self-ratings of language proficiency have been shown to correlate well with objective measures and thus are argued to be a reasonable method for bilingualism research (Delgado, Guerrero, Goggin, and Ellis, 1999). To understand the associations in our different measures of language learning, we obtained correlations between age of regular exposure to English, self-ratings, and the two word fluency tasks. Age of exposure correlated with English FAS, r= -.29, and with self-reported speaking level, r= .29, both ps < .03, with English FAS and speaking level correlated at r=.27, $p < .05$. These were the strongest correlations.

Participants were randomly assigned to the four task conditions (Emotional intensity rating, N=15; Letter Counting, N=16; Translation, N=14 and Word association generation, N=14). Because of random assignment, groups should be equivalent on demographic and proficiency measures. One-way analysis of variance (ANOVA) conducted on each of these measures revealed no differences among groups, except for self-rated reading proficiency, $F(3,55)=4.1$, $p < .02$. Participants assigned to the Emotional intensity rating task had a mean reading proficiency of 3.2, while the means of the other three tasks ranged from 3.7 to 4.0. Because none of the other ratings differed between groups, we will tentatively assume that the groups have similar English proficiency.

Design and Materials

Stimuli were the same as those in Ayçiçegi and Harris (2004) and Harris, Ayçiçegi and Gleason (2003) and are listed in the Appendix. Positive, negative and neutral items (16 in each category) were selected using the pleasantness scale from the Handbook of Semantic Word Norms (Toglia and Battig, 1978); words in these three categories did not differ overall in

subjective familiarity, as measured by Toglia and Battig's ratings. The nine taboo words were modified from the list used by Gonzalez-Regiosa (1976) and did not contain the classic English four-letter taboo words. The seven reprimands include some that parents would frequently say to children and some used in adult contexts. Exact translation equivalents did not exist for all the reprimands or taboo words. The Turkish items in this category were selected to have similar frequency of use and emotional connotation. For example, for the English "Go to your room!" we used a Turkish phrase, used in similar reprimanding situation, that can be glossed as "I don't want to see you!" Turkish and English stimuli were presented intermixed, in random order. Two stimulus lists were constructed so that any given item was presented to different participants in English vs. Turkish. The same stimulus lists were used in all four tasks, with only instructions differing.

Procedure

The protocol was administered individually to participants. Participants were told that they would be seeing items displayed on a computer screen in both Turkish and English. They received their instructions via computer in Turkish and questions were answered in Turkish. Turkish was used as the sole language of instruction because it was the language of the immersion environment and it was the dominant language for all participants. Below we have italicized the material which was presented as part of the instructions.

Shallow processing task. In each word or phrase, count the number of letters containing a closed circle. Note that these are the following letters: p, q, o, a, b, d, g. Enter the number using the computer keyboard. Pilot testing revealed that this task took approximately the same length of time as the emotional rating and translation tasks.

Emotional rating task. Rate the word for emotional intensity on a 1-to-7 scale, by typing the number on the computer keyboard.

Translation task. If the word or phrase appears in Turkish, translate it and report it to the experimenter in English; if in English, translate and report it to the experimenter in Turkish.

Word association task. Report as many words as come to mind that are associated with the displayed words, in 10 seconds, in the same language as the displayed word. Participants reported words out loud verbally to the experimenter who wrote down the words immediately.

Stimuli were presented on an Apple Macintosh laptop using PsyScope, experimental control software developed by Cohen, MacWhinney, Flatt, and Provost (1993). Participants responded to

stimuli by typing on the computer's keyboard. The rating/study phase of the experiment, including instructions, took less than 15 minutes. Immediately afterwards all participants were given a surprise recall test, with instructions to write down as many words as could be remembered from the computerized presentation, in the language in which they had appeared on the computer screen, within a 10 minute time frame. Following the instructions in Anooshian and Hertel (1994), if participants stopped their recall before 10 minutes, they were reminded by the experimenter that they had more time and could write down further words if more came to mind.

Only stimuli recalled in the target language were scored as correct. Percentage correct was calculated over participants and over items, in each task. Item percentages, averaged over tasks, are listed in the Appendix.

Results

Our primary dependent measure was percent of items correctly recalled in each category, with only words recalled in the presented language scored as correct. (Words recalled in the opposite language of presentation are analyzed separately below.) Recall of items is plotted for each task in Figures 1 and 2. Error bars indicate standard error of the mean and can be used as a clue to likely statistically reliable differences (i.e., means that are within each other's error bars are most plausibly seen as differing only by chance).

To validate our stimulus selection and translation procedures, emotional intensity ratings (collected during the emotional intensity task) were analyzed. Ratings were similar across the two languages and did not differ by more than one standard error of the mean. The mean ratings (7 point scale, 7 indicating high intensity) for L1-Turkish and L2-English were, respectively, 3.6 and 3.5 for neutral words, 5.6 and 5.5 for positive words, 5.6 and 5.8 for negative words, and 4.8 and 5.3 for reprimands. When data was subjected to ANOVA, no main effect of language and no language X stimulus category interaction occurred (all Fs<1).

Recall characteristics common to all tasks

Scrutinizing the four graphs in Figures 1-2 reveals a number of commonalities which should be noted before investigating task differences. The basic emotion-memory effect of greater recall for emotional items over neutral items is present in each task. The exception to this is striking: in each task, the negative words were recalled at the same rate as the neutral. The emotional connotations of negative words thus did not boost recall, or some other property of the

negative words may have inhibited recall (see further discussion below). The positive words were generally higher than the neutral. The two highly emotional categories, reprimands and taboo words, lead to high recall in every task.

An initial analysis of variance (ANOVA) conducted on the full three-factor design (language X emotion category X task) revealed a significant three-way interaction for analyses conducted by participants, $F_1(12, 212)=2.5$, and items, $F_2(12, 177)=2.6$, both ps $< .01$. To verify the statistical significance of stimulus conditions noted above, we compared negative versus neutral stimuli, positive vs. neutral, and taboo/reprimand vs. positive. Negative words had no recall advantage beyond neutral words (22% vs. 24% respectively), $p > .18$, and word-type did not interact with language or task. Positive words elicited higher recall than neutral, $F_1(1,55)=11.6$, $p < .001$, $F_2(1,30)=3.3$, $p=.08$, and taboo items had higher recall than positive words, $F_1(1,55)=14.1$, $p < .001$, $F_2(1,23)=3.2$, $p=.08$. Reprimands had higher recall than positive words, but this was reliable only in the analysis by subjects, $F_1(1,55)=7.8$, $p < .01$, $F_2 < 1$. Taboo items and reprimands did not differ from each other, F_1 and $F_2 < 1$.

The significant three-way omnibus ANOVA licensed conducting separate language X emotion categories ANOVAS on each task, but reporting these separate ANOVAS does not seem useful, since a language X emotion interaction would be almost guaranteed, at least in analysis by subjects, by the higher recall of reprimands in English than Turkish, a finding which occurred in every task. The higher recall of English reprimands is visible in Figures 1 and 2 (a listing of percent recall of each item, averaged across tasks, appears in the Appendix). The greater recall of reprimands in English than Turkish in each task was statistically verified by conducting a 2-way ANOVA on just the reprimands, with task and language as independent factors. This was highly significant in an analysis by subjects, $F(1,55)=45.4$, $p < .001$, but given that only seven reprimands were used, analysis by items was not statistically reliable, $F(1,6)=1.6$, $p=.24$. Of these seven reprimands, all showed high recall (recalled by 1/3 or more of participants), except the English reprimand *Go to your room*, and Turkish *Senden nefret ediyorum* (I hate you) and *Dur* (Stop that). There was particularly high recall of two English reprimands, *Shut up* (recalled by 73% of participants) and *I hate you* (recalled by 70%; see Appendix). High recall of the reprimands replicates Ayçiçegi and Harris (2004). In that paper, we proposed that novelty and unusualness of English reprimands influenced recall, meaning that our study was tapping a novelty effect, and not purely an emotion-memory effect (McDaniel,

Einstein, DeLosh, May and Brady, 1995). The reprimands in a second language could have been interpreted as more unusual and less expected than reprimands in a first language. Several participants laughed when they encountered the English reprimands and during debriefing commented on being surprised by them.

If we accept that the recall of the reprimands represents a novelty effect and not an emotion-memory effect, then it is worth setting aside the reprimands in our statistical analysis of the tasks and focusing on the remaining categories in order to draw general conclusions about processing emotional language in a first and second language.

Task-specific findings

We thus eliminated reprimands and obtained emotion-memory effects by subtracting each participants' neutral score from each of the other scores. The task X language X emotion category ANOVA conducted on these differences scores revealed no main of language, F_1 and $F_2 < 1$. The three-way interaction was significant, $F_1(6,110)=3.4$, $F_2(9,132)=3.0$, both ps $< .01$, indicating that emotion-memory effects for the two languages varied across tasks.

Emotion-intensity rating task. The emotion-rating task is plotted in Figure 3 (left-hand panel). A two-way ANOVA revealed a significant interaction of language and stimulus category, $F_1(2,28)=3.4$, $F_2(2,38)=4.7$, both ps $< .05$. Emotion-memory effects were confined to the first language, Turkish. Here we see a pattern consistent with the classic view that L1 stimuli have greater emotional connotations than L2 stimuli.

Letter counting (shallow) task. As shown in Figure 3 (right-hand panel), both languages showed recall advantages for taboo and positive words, but not for negative words. No language differences in recall occurred: in the 2 X 3 ANOVA, there was no main effect of language and no language X stimulus type interaction. The main effect of stimulus type was significant, $F_1(2,30)=18$, $F_2(2,38)=9.4$, both ps $< .001$, and the pattern of Taboo > positive > negative follows the cross-task analysis reported above.

Translation task. A main effect of emotion category was revealed, following the general trend of taboo > positive > negative, $F_1(2,26)=18.3$, $F_2(2,38)=7.9$, both ps $< .001$ There was also an interaction of language and stimulus category, $F_1(2,26)=6.0$, $p < .01$, $F_2(2,38)=3.2$, $p < .05$. In contrast to the pattern observed for the emotion-intensity rating task, English had stronger emotion-memory effects (for taboo items and positive words, not negative words) than did Turkish.

Word association task. There was a strong main effect of emotion category, resulting from the standard recall pattern of taboo > positive > negative, $F_1(2,26)=5.9$, $F_2(2,38)=5.5$, both ps < .01. However, there were no language differences or interaction of language X emotion category, indicating that emotion memory effects were similar across languages (Fs < 1).

To visualize how task influenced emotion-memory effects, Figure 4 graphs recall difference scores (with recall of neutral words subtracted, as in the above analysis), averaging over emotion-word category (again omitting the reprimands). Here we can see what was discussed above: the greater emotion-memory effect for L1-Turkish in the emotion rating task, the greater emotion-memory effect for L2-English in the translation task, and the similarity between English and Turkish for the letter counting task and association tasks.

Incorrect recall: analysis of translation equivalents

During the recall period, fewer than half the participants wrote down even one translation equivalent of a studied word. This is consistent with prior reports that language-specific recall of previously presented words is performed well by bilinguals (Keatley, 1992). Despite the small number of recall errors, the pattern is worth examining. Recall of translation equivalents did not differ for items studied in Turkish and English. However, there were task and emotional category differences. Collapsing over categories, 14.5% of participants wrote down a translation equivalent for the emotion rating task, 17% for the translation task, and 19% for the word association task. In contrast, only 6% of participants wrote down a translation equivalent in the letter counting task. This lower rate of generation of translation-equivalents suggests that the shallow task was indeed successful in limiting processing of the meaning of words. Reprimands, neutral items, and taboo items were least frequently recalled in the alternative language (10-14%), while aversive and positive words were more frequently recalled (17%-24%). Low generation of translation equivalents for neutral items likely occurred because there was overall low recall of these items in the studied language. Reprimand and taboo items had the highest correct recall of any items. Thus, the lack of recall of translation equivalents suggests that participants had good episodic encoding of these items. That is, they genuinely remembered the item (which includes its language), rather than recalling its meaning.

No emotion-memory effect for negative words?

As mentioned above, the negative words had no advantage in recall over the neutral words in any of the tasks (see Figures 1 and 2), in either language. We scrutinized the negative and positive words, to determine if they differed along any dimensions other than valence (see listing in the Appendix). The positive words included basic concepts learned in early childhood, such as *mother, love, home*, and *happy*. Indeed, the higher-than-neutral recall for positive words was primarily carried by the very high recall (more than 50% of participants) for *mother, father* and *love* (in English) and for *mother, father* and *home* (in Turkish). No negative words approached these high rates of recall. The negative words included concepts that are typically acquired after early childhood (after age 4), such as *cruel, disease, poison,* and *crime*. Consulting age-of-acquisition (AoA) norms for English words (Iyer, Saccuman, Bates, and Wulfeck, 2001) revealed that positive words in our sample had mean AoA of 4.0 years, neutral had 4.2, and negative, 5.2; standard deviations about 1.4 years for these categories; difference between positive and negative, $t(13)=1.5$, $p=.15$. This difference could well be statistically significant with a larger sample. Of course, these are norms for the acquisition of words in a first language. Words that are acquired late in a first language, such as "passport," can be acquired early in a second language (Izura and Ellis, 2002). However, our stimuli were not selected to disentangle acquisition-age of words from acquisition-age of concepts, although this could be explored in future research. Whether larger emotion-memory effects hold for early acquired negative words than for late acquired negative words is also a topic worthy of future investigation.

Discussion

The primary goal of the current paper was to investigate whether emotion memory effects, meaning the recall advantage of emotional stimuli over neutral stimuli, are stronger in a first language than in a second language, using study tasks that differed in degree of elaborative processing.

- Averaging over the four study tasks, emotion memory effects were equally strong in a second language (English) as in a first language (Turkish), except for reprimands, which elicited higher recall in English. English reprimands may have benefited from novelty (McDaniel et al., 1995).

- Across tasks and languages, recall of emotion stimuli varied in this rank order: {taboo, reprimand} > positive > {negative, neutral}
- Negative words (war, fight kill) showed no recall advantage over neutral words in any task or language, which may be due to their slightly later age-of-acquisition.
- Emotion-memory effects and recall differences in L1 vs. L2 depended on the task.

The finding of equivalent emotion-memory effects in a first and second language replicates our prior finding (Ayçiçegi and Harris, 2004). The prior findings could have occurred because the second language was activated to a higher degree due to L2 immersion environment. The current study ruled this out since our bilingual participants were tested while residing in Istanbul, the country of their L1.

The current study used multiple tasks in order to rule out a second explanation for the prior finding of strong L2 emotion-memory effects: these effects occurred because the emotion-rating task of the prior study inadvertently encouraged participants to more deeply encode L2-English emotion words, in order to make subtle rating discriminations. Study-tasks were selected which would have required either similar processing in an L1 vs. an L2, or would be biased toward a deeper level of processing in L1. Therefore, if strong emotion-memory effects were nonetheless obtained in L2-English, these would indicate true emotion-memory effects, rather than being a by-product of deeper processing in the L2 induced by task demands. Task-specific finding were the following.

Emotion-intensity rating task. This was the basic task used in the two prior studies (Anooshian and Hertel, 1994; Ayçiçegi and Harris, 2004). In this task, emotion-memory effects were confined to L1-Turkish (Figure 3). The other tasks showed a different pattern and thus the results of the emotion-intensity rating task may be task or situation-specific. Evaluating the intensity of emotional stimuli may lead to more elaborative encoding (and thus better recall) for stimuli in the language which is most highly activated in the testing environment (L1-Turkish in the current study, L2-English in Ayçiçegi and Harris, 2004).

Letter-counting task (shallow processing). In the letter-counting task, participants counted the number of letters containing a closed circle. This task does not require any attention to the meaning of the stimuli, and is thus considered a shallow task in the levels-of-processing framework (Craik and Lockhart, 1972; Jacoby and Dallas, 1981). We had hypothesized that greater emotion-memory effects would emerge for L1-Turkish than for L2-English because the

meaning of words in L1 would be automatically processed even if shallowly processed. Instead, similar emotion-memory effects were obtained for both languages (Figure 4). Our finding of equivalent emotion-memory effects in L1 and L2 suggests that the shallow processing task allows equivalent automatic processing in both languages, similar to several Stroop task findings of equivalent color-word naming interference in an L1 and L2 (Eilola, Havelka, and Dinkar, 2007; Sutton, Altarriba, Gianico, and Basnight-Brown, 2007).

Note that the letter-counting task did succeed in its designation as a shallow-processing task, as overall recall was poorest in this task. However, stimuli were still processed for meaning in the letter-counting task, and were processed in a manner that resulted in similar recall for L1 and L2. Indeed, the letter-counting task showed the strongest overall emotion-memory effects of any task (12% advantage over the neutral stimuli, compared to %4 for the word association task, Figure 4). Shallow processing tasks frequently yield stronger emotion-memory effects than tasks requiring more elaborative processing. In recall tasks, neutral stimuli benefit more from elaborative processing than do emotional stimuli, because emotional stimuli are already processed more deeply than neutral stimuli even in a shallow task, due to their inherent emotionality (Jay et al, 2008).

Translation task. This task was included because a key proposal in the bilingualism literature is that translating from the L2 to the L1 is the easier translation direction (Kroll and Stewart, 1994). Consistent with this, some participants remarked during debriefing that translating from L1-Turkish was more difficult, because their Turkish taboo vocabulary was more extensive than their English taboo vocabulary. The unexpected result was that L2-English emotional stimuli showed greater emotion-memory effects than L1-Turkish. We infer that a second language is capable of demonstrating emotional-memory effects which are as strong as L1 effects, or stronger, at least for certain tasks such as translation.

Word association task. Because word associations can be more easily generated in a first language than in a second, this task was proposed to have at least some bias towards inducing more elaborative encoding in L1 than in L2. Other than the high recall of reprimands in English (present in all tasks), both recall and emotionality effects were similar in the two languages. This suggests that the word-association task induces a similar level of processing in both languages. Generating multiple associations for L1 words confers memorability, which leads to good L1

recall. But, for the L2 word association task, the difficulty of finding association for an L2 can also mean that it is processed deeply.

Future research

Emotion-memory effects need to be studied using different tasks rather than only using an emotion-intensity rating task or assuming that study-task is irrelevant. However, note that we were not able to predict which tasks would lead to language-specific effects. Future work needs to be grounded in models of bilingual memory and a theory of task difficulty in order to make more precise predictions.

A fruitful avenue may be to select tasks which lie on a continuum from shallow to deep processing. In the current study, the association task led to the highest overall recall rate (46% of items recalled), but the smallest emotion-memory effects (i.e., only a 4% advantage of emotional stimuli over neutral stimuli), because of the high recall of neutral stimuli. The high recall of neutral stimuli plausibly occurred because the association task conferred a great deal of elaborative encoding on the neutral words. In contrast, the letter counting task had the lowest overall recall rate (17%) but the highest emotion-memory effects (12% advantage over neutral), because emotional stimuli attract attention due to their inherent interest (see discussion in Jay et al, 2008). It could be useful in future research to employ a shallow task which is less time-consuming than this one, which had been chosen to equate decision time with the other tasks. The classic upper/lower case decision would be a good test of automatic processing, similar in time-requirements to the Stroop task. Inspection of individual items suggests that words referring to early-acquired concepts had higher recall than later-acquired concepts, in both languages. Future research could determine if age-of-acquisition is more important for recall within a language than is age-of-acquisition of the lexical item, following the logic described by Izura and Ellis (2002).

The methodology of study-task followed by surprise-recall test is both simple and powerful. It does not require use of a computer and lends itself to a classroom setting or student-conducted projects. It would thus be easy to pursue variations in tasks and L1/L2 attributes to assemble a database of findings which can test models of emotionality and bilingual memory.

Conclusions

The thesis that the first language is experienced as more emotional is powerful, because it fits both with contemporary neuroscience ideas about the primacy of events from early childhood, and with many anecdotal impressions obtained in surveys, interviews and in the work of bilingual writers (Pavlenko, 2005). However, greater emotionality in a first language is not invariably found (e.g., Dewaele, 2006; Schrauf and Rubin, 2004; Schrauf and Durazo-Arvizu, 2006). Bilingual speakers may dispute the claim that the first language is more emotional, noting the appeal, novelty or romantic difference of a newly learned language, as when an English speaker remarks on the cuteness of diminutives in German or Russian. The current finding of high recall for L2 reprimands may be a laboratory manifestation of this type of novelty effect. Aside from the reprimands, the overall finding of similar emotion-memory effects for a first and second language indicates that words' emotional attributes are processed to a similar extent in both a first and second language.

References

Altarriba, J. (2003). Does cariño equal "liking"? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism, 7*, 305–322.

Altarriba, J. & Santiago–Rivera, A.L. (1994). Current perspectives on using linguistic and cultural factors in counseling the hispanic client. *Professional Psychology: Research and Practice, 25,* 388–397.

Anooshian, J. L. & Hertel, T. P. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition and Emotion, 18,* 503–514.

Ayçiçegi, A. & Harris, C.L. (2004). Bilinguals' recall and recognition of emotion words. *Cognition and Emotion*, *18*, 977–987.

Berny, T.D., Cooper, RL. & Fishman, J.A. (1968). Semantic independence and degree of bilingualism in two Puerto Rican communities. *Revista Interamericana de Psicologia, 2*, 289–294.

Caldwell-Harris, C.L. & Ayçiçegi-Dinn, A. (under review). Emotion and lying in a non-native language. *International Journal of Psychophysiology*.

Cohen, J.D., MacWhinney, B., Flatt, M.R. & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25*, 257–271.

Craik, F.I. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

de Groot, A.M.B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1001-1018.

Delgado, P., Guerrero, G., Goggin, J.P. & Ellis, B.B. (1999). Self–assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioral Sciences*, *21*, 31–46.

Delis, D., Kaplan, E. & Kramer, J. (2001). *Delis-Kaplan executive function system (D-KEFS).* San Antonio, TX: The Psychological Corporation.

Dewaele, J.-M. (2004a). Blistering barnacles! What language do multilinguals swear in?! *Estudios de Sociolinguistica, 5*, 83–106.

Dewaele, J-M. (2004b). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development, 25*, 204–222.

Dewaele, J.-M. (2005). The effect of type of acquisition context on perception and self–reported use of swearwords in the L2, L3, L4 and L5. In A. Housen and M. Pierrard, (eds.), *Investigations in instructed second language acquisition,* pp. 531–559. Berlin: Mouton De Gruyter.

Dewaele, J.-M. (2006). Expressing anger in multiple languages. In A. Pavlenko (ed.), *Bilingual minds: Emotional experience, expression, and representation,* pp. 118–151. Clevedon: Multilingual Matters.

Eilola, T.M., Havelka, J. & Dinkar. S. (2007). Emotional activation in the first and second language. *Cognition and Emotion, 21,* 1064-1076.

Fletcher, P.C., Stephenson, C.M.E. & Carpenter, T.A. (2003). Regional brain activations predicting subsequent memory success: An event–related fMRI study of the influence of encoding tasks. *Cortex, 39*, 1009–1026.

Gonzalez–Reigosa, F. (1976). The anxiety arousing effect of taboo words in bilinguals. In C.D. Spielberger and R. Diaz–Guerrero (eds.), *Cross–cultural anxiety*, pp. 89–105. Washington, DC: Hemisphere.

Green, D. W. (1986). Control, activation and resource: A framework and a. model for the control of speech in bilinguals. *Brain and Language, 27*, 210–223.

Harris, C.L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 25*, 223–247.

Harris, C.L., Ayçiçegi, A. & Gleason, J.B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first than in a second language. *Applied Psycholinguistics, 4,* 561–578.

Harris, C.L., Gleason, J.B. & Ayçiçegi, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Languages and emotions of bilingual people,* pp. 257–283. Clevedon, UK: Multilingual Matters.

Iyer, G., Saccuman, C., Bates, E. & Wulfeck, B. (2001). A study of age–of–acquisition (AoA) ratings in adults. *Center for Research in Language Newsletter, Vol. 13*, No. 2. La Jolla: University of California, San Diego. Available from: http://crl.ucsd.edu/newsletter/13–2/article.html

Izura, C. & Ellis, A.W. (2002). Age of acquistion effects in word recognition and production in first and second languages. *Psicologia, 23,* 245-281.

Jay, T. (1999). *Why we curse: A neuro–psycho–social theory of speech.* Amsterdam: John Benjamins.

Jay, T., Caldwell–Harris, C.L. & King, K. (2008). Recalling taboo and nontaboo words. *American Journal of Psychology*, *121*, 83–103.

Jacoby, L.L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Keatley, C.W. (1992) History of bilingualism research in cognitive psychology. In R.J. Harris (ed.), *Cognitive processing in bilinguals*. Amsterdam: Elsevier.

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

LaBar, K.S. & Phelps, E.A. (1998). Arousal–mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science*, *9,* 490–493.

MacKay, D. G. & Ahmetzanov, M. V. (2005). Emotion, memory, and attention in the taboo Stroop paradigm: An experimental analog of flashbulb memories. *Psychological Science, 16*, 25-32.

Marian, V. & Margarita K. (2004). Self–construal and emotion in bicultural bilinguals. *Journal of Memory and Language, 51,* 190–201.

Manning, S.K. & Julian, L. (1975). Recall of emotional words. *Journal of General Psychology*, *92,* 237–244.

McDaniel, M.A., Einstein, G.O., DeLosh, E.L., May, C.P. & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 422–435.

Pavlenko, A. (2005). *Emotions and multilingualism*. Cambridge, MA: Cambridge University Press.

Rappaport, D. (1942). *Emotions and memory*. New York: International Universities Press.

Reber, R., Perrig, W., Flammer, A., & Walter, D. (1994). Levels of processing and memory for emotional words. *Schweizerische Zeitschrift für Psychologie, 53,* 78–85.

Rosselli, M. & Ardila, A. (2002). A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience*, *112*, 759-776.

Rubin, D.C. & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory and Cognition, 14,* 79–94.

Schrauf, R.W. (2000). Bilingual autobiographical memory: Experimental studies and clinical cases. *Culture and Psychology, 6,* 387–417.

Schrauf, R.W. & Rubin, D.C. (2004). The 'language' and 'feel' of bilingual memory: Mnemonic traces. *Estudios de Sociolinguistica, 5,* 41–58.

Schrauf, R.W. & Durazo-Arvizu, R. (2006). Bilingual autobiographical memory and emotion: theory and methods. In: A. Pavlenko (ed.), *Bilingual minds: Emotional experience, expression, and representation,* pp. 284–311. Clevedon: Multilingual Matters.

Sutton, T.M., Altarriba, J., Gianico, J.L. & Basnight-Brown, D.M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish-English bilingual speakers. *Cognition and Emotion, 21,* 1077-1090.

Toglia, M. P. & Battig, W.F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Tzelgov, J, Henik, A. & Leiser, D. (1990). Controlling Stroop interference: Evidence from a *bilingual task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 760–771.

Vilkki, J.U. (1987). Incidental and deliberate memory for words and faces after focal cerebral lesions. *Neuropsychologia, 25*, 221–230.

Zied, K.M, Philippe, A. & Karine, P. (2004) Bilingualism and adult differences in inhibitory mechanisms: Evidence from a bilingual stroop task. *Brain and Cognition*, *54*, 254–256.

Table 1: Learning History and Fluency for Turkish–English Bilingual Participants

	Mean	SD	Minimum	Maximum	
Age	21	3.9	13	41	
Years Education (12=high school grad)	15	1.9	13	20	
Age Regular English Exposure	13	2.3	7	19	
English Word Fluency (FAS Task[b])		28	6.2`	18	49
Turkish Word Fluency (FAS Task[b])		36	7.0	22	53
Self–reported ability in English[c]					
Speaking	3.1	0.9	1	5	
Understanding	3.8	0.6	2	5	
Writing	3.2	0.8	1	5	
Reading	3.7	0.7	1	5	

Table Notes. [a]Age at which participants took English classes for at least one hour a day

[b]Number of words generated beginning with letters F, A, S.

[c]1–5 scale (3=good, 4=very good, 5=near–native)

Figure Captions

Figure 1. Percent of trials recalled for the emotional–intensity rating (deep) task) and the letter counting (shallow) task.

Figure 2. Percent of trials recalled for the translation and word association tasks.

Figure 3. Percent of trials recalled after subtracting neutral–word recall, for the emotion rating and translation tasks, with reprimands omitted.

Figure 4. Percent of trials recalled after subtracting neutral–word recall for all tasks, averaging over stimulus type, with reprimands omitted from analysis.

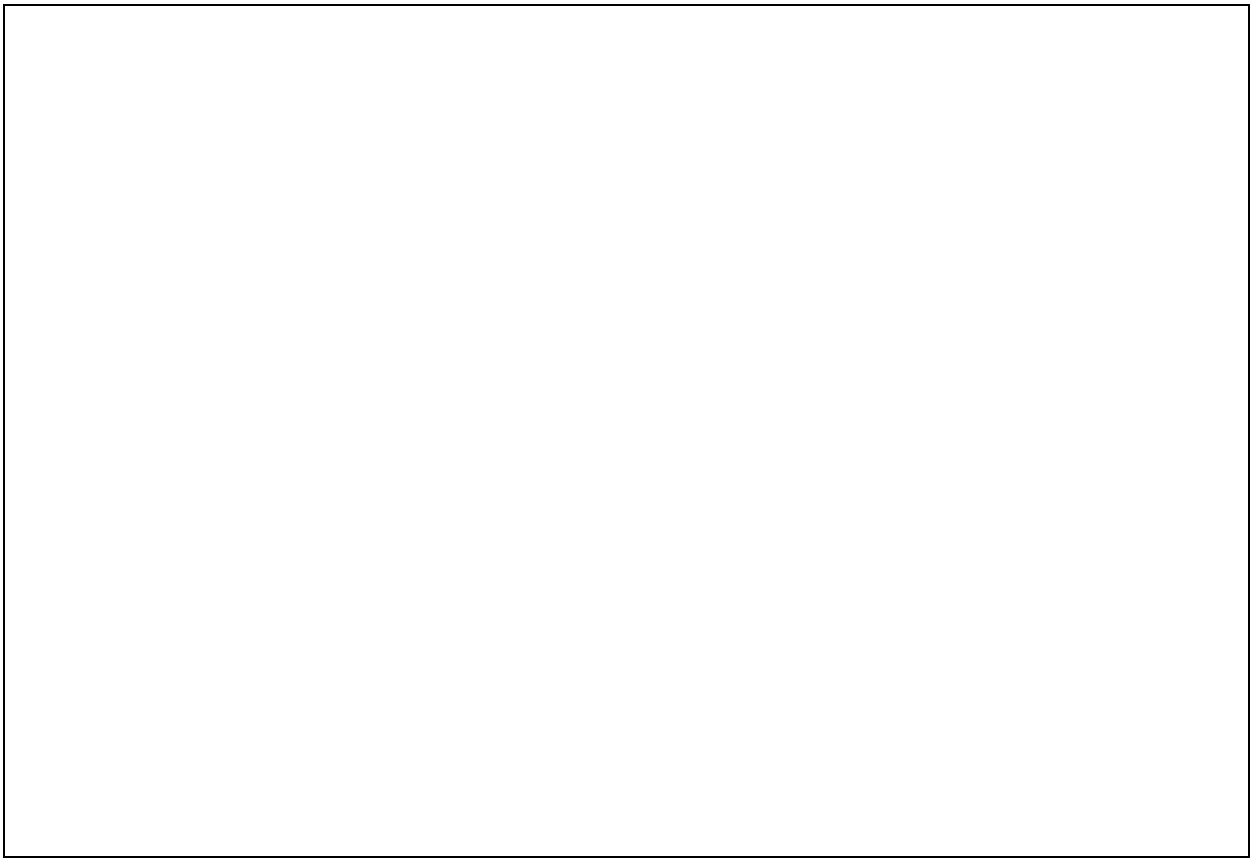

Figure 1. Percent of trials recalled for the emotional–intensity rating (deep) task) and the letter counting (shallow) task.

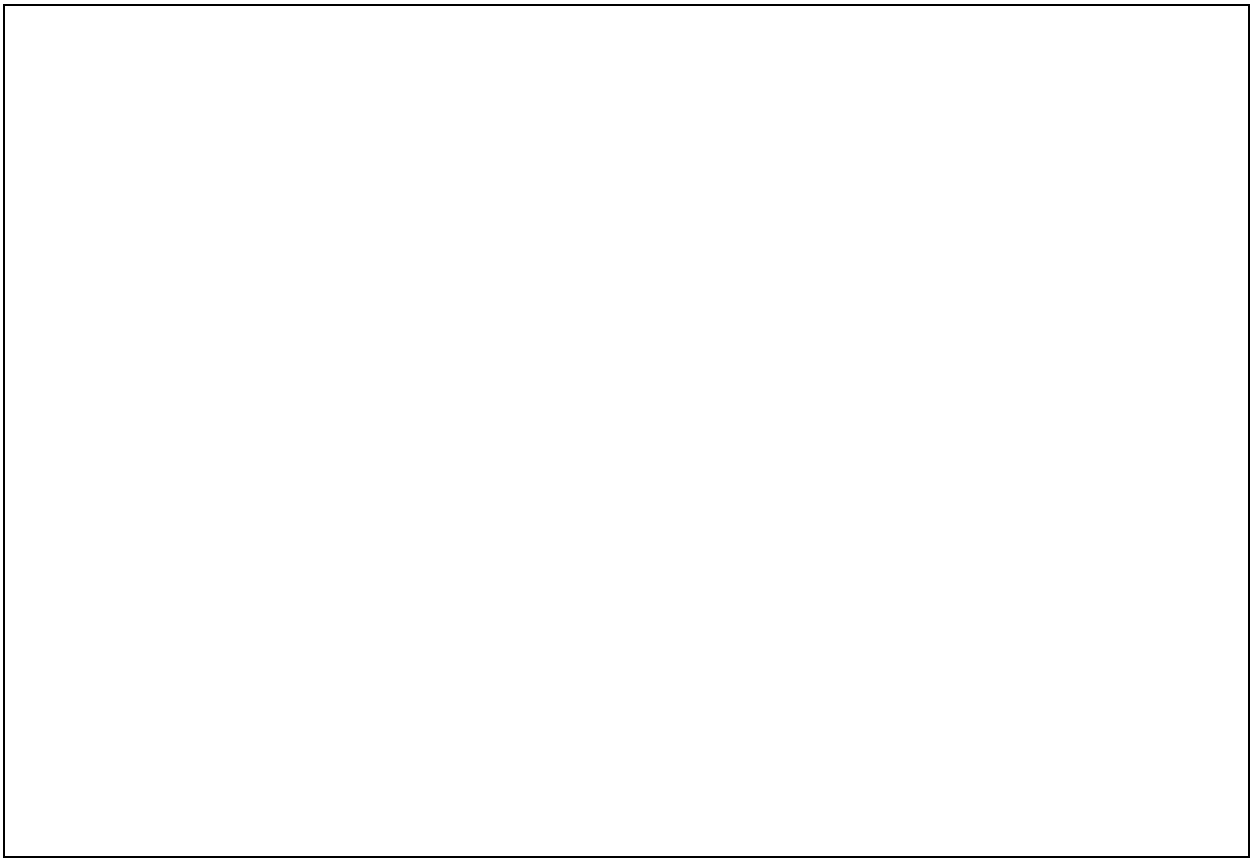

Figure 2. Percent of trials recalled for the translation and word association tasks.

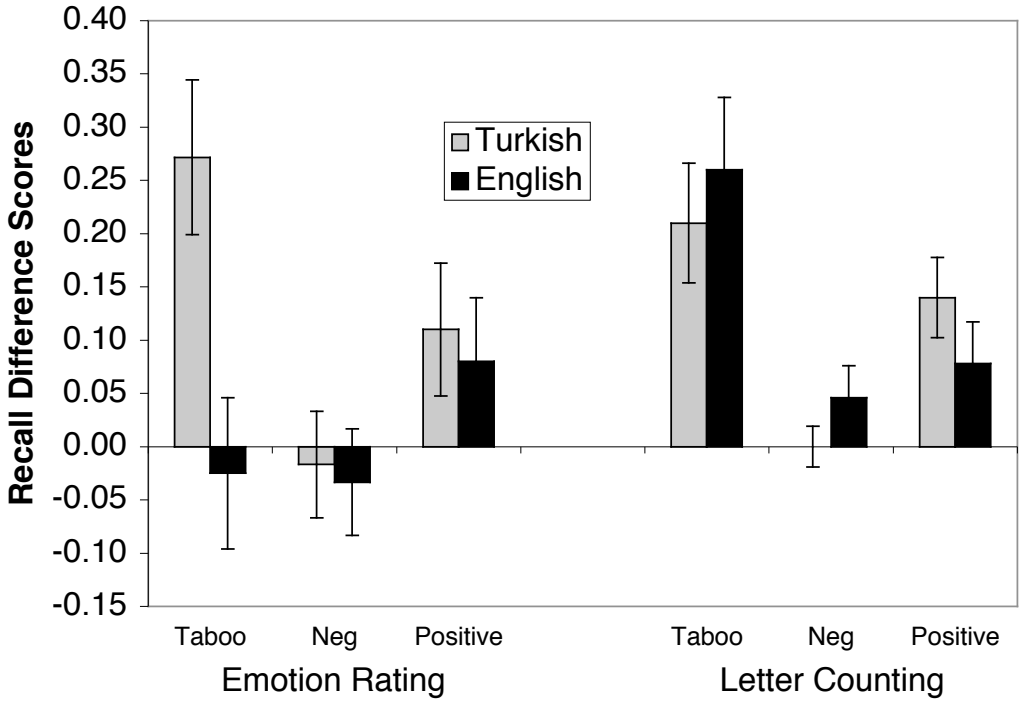

Figure 3. Percent of trials recalled after subtracting neutral–word recall, for the emotion rating and translation tasks, with reprimands omitted.

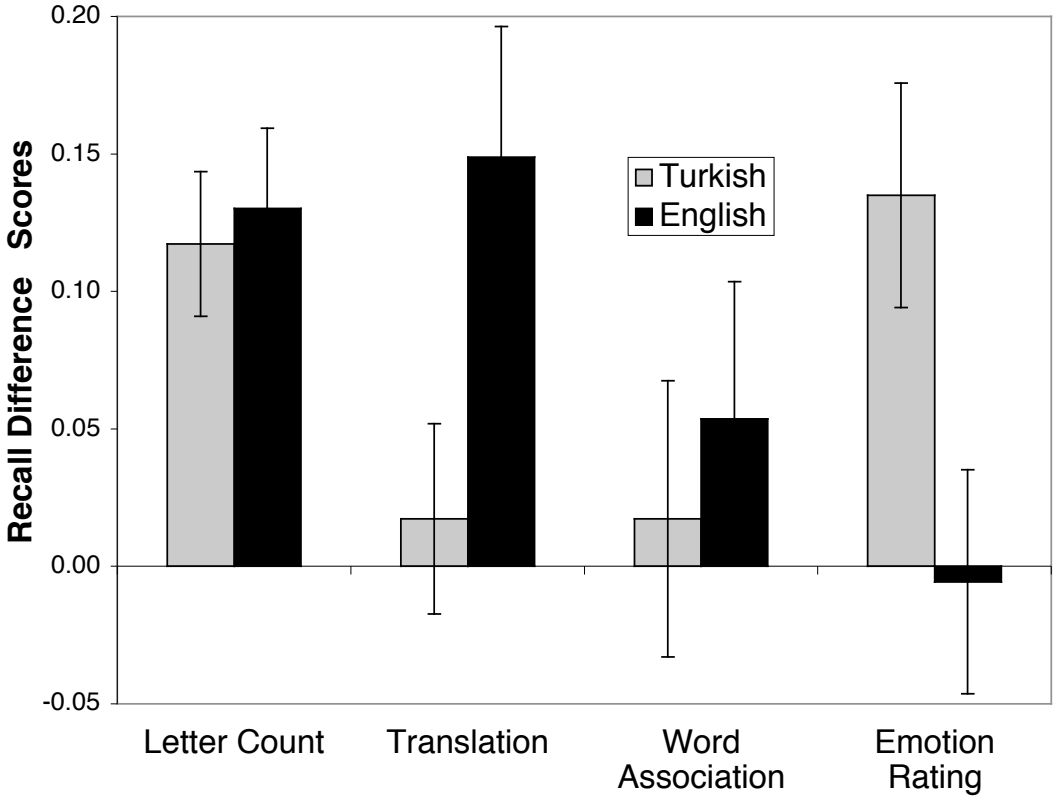

Figure 4. Percent of trials recalled after subtracting neutral–word recall for all tasks, averaging over stimulus type, with reprimands omitted from analysis.