Testing pragmatic inferences: The impact of language and culture

Paula Carneiro[*1], Ana Lapa[1], Júlia Reis[2], Tânia Ramos[2,3]

[1]CICPSI, Faculdade de Psicologia, Universidade de Lisboa, PORTUGAL

[2]Faculdade de Psicologia, Universidade de Lisboa, PORTUGAL

[3]New York University, NY, USA

Pragmatic inferences are one way to study false memories in real-world situations. We aimed to investigate variances in responses to pragmatic implication sentences between Portuguese and American data, presenting, for the first time, normative data of cued recall and recognition for pragmatic implication sentences in Portuguese. In Study 1 we analyzed cued-recall data for Portuguese pragmatic sentences. The proportions of cued- recall for correct and inference responses of each sentence did not significantly correlate with the values of American normative data. In Studies 2a and 2b we analysed forced-recognition data for pragmatic sentences, one with American participants and English sentences (Study 2a) and the other with Portuguese participants and Portuguese sentences (Study 2b). Moreover, two conditions of sentences presentations were applied to eventually detect an influence of the sentences' format, which was not find in both studies. The levels of recognition for correct and inferences were very similar between those two studies but the correlation, sentence by sentence, was low. Together, these results suggest an important recommendation for further studies - normed material for a specific language/ culture is a crucial factor to be considered when conducting research on pragmatic inferences.

"Why did you say to John that I fell off from the skateboard? I just told you that I have lost my balance." In everyday conversations people very often infer consequences of behaviours or actions that do not correspond exactly to what was told by others. Because of these non-intentional inference processes sometimes individuals remember the information in a distorted way.

This article focuses on this specific form of inducing everyday false memories, that is, on pragmatic inferences. One aim of this study was to create normative data of cued-recall and recognition for pragmatic implication sentences in Portuguese and compare them with data already normed for English. Another aim was to analyse the influence of structure presentation of pragmatic sentences on false recognition. We intend to know whether participants could detect the structure or the format of the implication sentences and then could respond accordingly, instead of relying simply on their memory trace.

In the introduction, we will start by presenting a brief review of the relevant literature on the reconstructive nature of memory using verbal material.

The Reconstructive Nature of Memory. Remembering is reconstructive: this means that people do not always remember information exactly as it was experienced. Reconstructions can be distorted or filled with errors. Information can be non-intentionally distorted for many reasons - to be consistent with our current knowledge, to achieve coherence, to align it with our expectations, to obtain the overall meaning of the information, and sometimes simply because our associative memory betrays us.

Bartlett (1932) was one of the first authors to demonstrate that memory is reconstructive. This author presented participants with stories for later reproduction and noticed that the general meaning of the text was usually retained, but not the sentences' literal wording. Participants omitted parts of the text that were inconsistent with the gist of the story and tended to include information that was more consistent and in accordance with their own understanding. Bartlett argued that remembering is guided by our mental constructions of one's understanding of the events, which he called schemas. For example, if an individual listens to news about a bank robbery, he/she later might refer to having heard the word "gun" because gun has a high probability of being part of the "bank assault" schema. Other authors such as Sulin and Dooling (1974) extended Bartlett's idea that our general knowledge of the world can distort memory. These authors asked a group of participants to read paragraphs of biographical texts about a famous person (e.g., Helen Keller, a deaf-blind person) and presented the exact same paragraphs to another group but changed the name of the character to Carol Harris.

Approximately 60% of the Helen Keller group falsely recognized that the unpresented sentence "She was deaf, dumb and blind" was in fact presented, whereas the Carol Harris group rarely made this error. Thus, sentences consistent with knowledge about the character tended to be falsely recognized. During memory reconstruction of the sentences, they appear to have become distorted to accommodate common knowledge.

In addition to the distortion, some authors found that the errors made in the initial recall of text passages tend to persist despite the re-presentation of the original text passages in subsequent sessions, which means that the initial recall of the errors makes them resistant to change (e.g., Fritz, Morris, Bjork, Gelman, & Wickens, 2000; Howe, 1970; Kay, 1955).

In the same sense, a specific fact can also become distorted from being integrated into a broader, consistent, schema: People sometimes integrate information from individual sentences to construct larger ideas. Bransford and Franks (1971) constructed an experiment with four-fact sentences (e.g., the ants in the kitchen ate the sweet jelly that was on the table) and deconstructed them into smaller sentences (e.g., the ants in the kitchen ate the sweet jelly – a three-fact sentence; the ants in the kitchen ate the jelly - a two-fact sentence; the jelly was sweet - a one-fact sentence). Participants heard the 1-, 2- or 3- fact sentences and then had to choose the sentences they previously heard from an answer set comprised of correct sentences, nonpresented consistent sentences, and nonpresented nonconsistent sentences. Unsurprisingly, almost all the nonconsistent sentences were correctly rejected as being previously presented. Notably, regarding correct and consistent sentences, the higher the number of facts the original sentences had, the more likely participants were to judge them as previously presented and the more confident they were in their responses. Additionally, except for the sentences that expressed only one fact, participants were equally confident when recognizing correct sentences and nonpresented consistent sentences. As Bartlett had already pointed out, people do not retain the literal form of the sentences, but their gist. This phenomenon emphasizes the reconstructive nature of our cognitive processes.

Deese (1959), introduced another methodology in order to explore the reconstructive nature of memory, which consisted of using lists of associative words. Interested in studying the associative memory, this author found that participants falsely remembered words that were not presented (critical words) but were highly associated to the presented words. For example, the presentation of a list composed of words such as "bed, rest, awake, tired, dream, wake, snooze, and blanket" frequently produced high false recalls and recognitions of the critical word "sleep" (Roediger & McDermott, 1995). This methodology was the origin of the DRM paradigm (Deese, Roediger

and McDermott), which is one of the paradigms most frequently used to study false memories.

Among other methodologies proposed to study false memories, the pragmatic inference paradigm is an approach to studying false memories of everyday events. A pragmatic implication occurs when information in a sentence leads the listener to expect something neither explicitly stated nor necessarily implied in the sentence. For example, the sentence "the new baby *stayed awake* all night" pragmatically implies that "the new baby *cried* all night." A *but not* conjunction is used to evaluate if a sentence implied a pragmatic inference. A sentence could be classified as having a pragmatic implication if one implies another and the two sentences could be joined with the *but not* conjunction to produce an acceptable sentence. In the previous example, it is acceptable that "the new baby *stayed awake* all night *but did not* cry." This *but not* test is very specific for pragmatic implications and excludes semantically synonymous sentences, logically implied sentences, or sentence pairs with no relation to each other.

Brewer (1977) showed that participants very often falsely recall pragmatic implications of sentences. This author demonstrated that the sentences containing pragmatic implications yielded more pragmatic inferences in a cued recall task (26%) than correct responses (19%). According to this author, the high level of false recall was due to the interaction of episodic information provided by the sentence with the relevant knowledge in long-term memory, showing that participants actively processed the material they were exposed to. This finding was also replicated with a recognition task (Chan & McDermott, 2006). Participants were presented with half of the sentences in the implication condition (e.g., "the new baby *stayed awake* all night") and the other half in the target condition (e.g., "the new baby *cried* all night"). During the recognition task, participants had to indicate whether the presented sentence was identical to the one presented at study. The sentences presented in the recognition test could be implication sentences (e.g., "the new baby *stayed awake* all night"), target sentences (e.g., "the new baby *cried* all night"), or inconsistent sentences (e.g., "the new baby *slept* all night"). The authors found that when sentences were presented in their implication form during study, the proportions of correct and false recognitions were very high (respectively, .73 and .65), suggesting that pragmatic inferences are recognized to a similar degree as the exact correct sentences.

Using a two-alternative forced-choice recognition task, McDermott and Chan (2006) in Experiment 1 studied the effects of repetition on memory for pragmatic inferences. They found that, even when presented with the original sentence and pragmatic inference, participants chose the inference sentence

almost as often as the literal correct sentence. Moreover, even after three encoding opportunities, participants still recognized the inference answer in almost one-third of the final test trials, demonstrating the difficulty of correcting this type of false memory (see Mullet & Marsh, 2016 study about the factors most important to correct this kind of false memory). In Experiment 2, McDermott and Chan (2006) used a cued-recall test and found similar proportions of correct and inference responses (respectively, .30 and .34) with only one study phase. This type of memory error is usually accompanied by high confidence (Mullet & Marsh, 2016). However, Sampaio and Brewer (2009) found that pragmatic inferences were held with higher confidence in a simple recognition task than in a two-alternative forced-choice recognition task, probably because in this last task participants were more aware of the deceptive nature of the schema-inference items.

In conclusion, pragmatic inferences seem to be a robust way to study false memories in an authentic environment, that is, in sentences that are part of everyday conversations. Considering the value of pragmatic inferences as a rich material for studying the reconstructive nature of memory and discourse processing, the current study presents the normative data of cued-recall (Study 1) and forced-choice recognition (Studies 2a and 2b) for pragmatic implication sentences in Portuguese. These data will facilitate the development of further investigations in the areas of inferences and false memories with Portuguese-speaking populations. One of our aims was also to establish a comparison between the Portuguese and American/ English data to answer the following question: Do linguistic or cultural factors matter for the reconstruction of memory? – and, more specifically, are the pragmatic implication sentences that produce the highest levels of false cued recall or recognition in English the same as those in Portuguese? Another aim of Studies 2a and 2b, besides creating recognition norms for Portuguese sentences, was also to examine the role of the structure of the pragmatic implication sentences. Can participants detect the format of these sentences and use this knowledge strategically to respond to the memory test? The pragmatic sentences usually present information that does not correspond to the final consequence of the action whereas de inference contains the final result (e.g., pragmatic sentence: "John *had lost his balance*"; inference: "John *fell out*"). Other times, pragmatic sentences contain the less expected information and the inference corresponds to the more likely response (e.g., pragmatic sentence: "The hypnotist *put his fingers together* and awakened his client; inference: "The hypnotist *snapped his fingers* and awakened his client". If participants notice the different format between the pragmatic sentence and its inference they could easily identify the presented words (correct) and its inference in a forced-recognition test, where both alternatives

are present. By comparing different presentation formats of pragmatic sentences (studies 2) we can analyse whether participants are responding according to the application of this strategy or according to memory.

STUDY 1: Cued Recall of Pragmatic Implication Sentences

METHOD

Participants. Seventy university students from the Faculty of Psychology at the University of Lisbon participated in this study ($M_{age} = 21.5$; 54 female) and were rewarded with course credits. Participants provided consent for their participation.

Materials. Forty-seven pragmatic sentences were selected from those used by McDermott and Chan (2006) and adapted[1] to Portuguese (one sentence was excluded because no meaningful Portuguese translation existed). McDermott and Chan had chosen these 48 sentences from a larger pool of 70 sentences, including some from Brewer's study and others constructed by Chan, as the sentences most effective at eliciting the recall of pragmatic inferences. The authors of the present study constructed additional 13 pragmatic sentences and added them to the pragmatic sentence pool. To construct these new pragmatic sentences, we created different sentences and respective expected pragmatic inferences and joined them using a *but not* conjunction. These constructed sentences lead the listener to expect something neither explicitly stated nor necessarily implied in the sentence. Six other sentences without pragmatic implications were also created for practice purposes. In total, 66 sentences were presented to participants: six practice sentences and 60 pragmatic implication sentences (see Appendix A).

Procedure. Participants performed the task individually on computers at the laboratory. They were instructed to read the sentences presented on the computer and to memorize them for later recall. Before starting the task, participants completed a practice trial with six sentences (without pragmatic implications). The practice sentences were presented on the computer screen and followed by a distractor task for 30 seconds. Afterwards, participants performed a practice cued-recall test, in which a fragment of each of the six practice sentences was presented on the screen and participants were asked to complete the missing words. After concluding the practice trials, the

---

[1] Some of the English sentences were not literally translated because they were not good examples of pragmatic sentences in Portuguese (e.g., "the bully put his fist to the geek"). This happened for five sentences from McDermott and Chan (2006) study.

participants were reminded to read and memorize the sentences that would be presented. In the same manner as in the practice trials, the pragmatic implication sentences were presented on the computer for 4.5 seconds each, in a random order, with a 1 second interval between them. After studying all 60 sentences, participants performed a distractor task (spot the differences, in which participants had to indicate the differences between two similar but not equal images) for 1 minute. Immediately afterwards, participants were asked to complete a cued recall task. Each sentence was randomly presented without its critical words (e.g., "the new baby ___all night") and they had to complete them by writing the missing words on the computer. The task had no time limit, but no participant took more than 20 minutes (see the scheme of the procedure for the three studies in Figure 1).
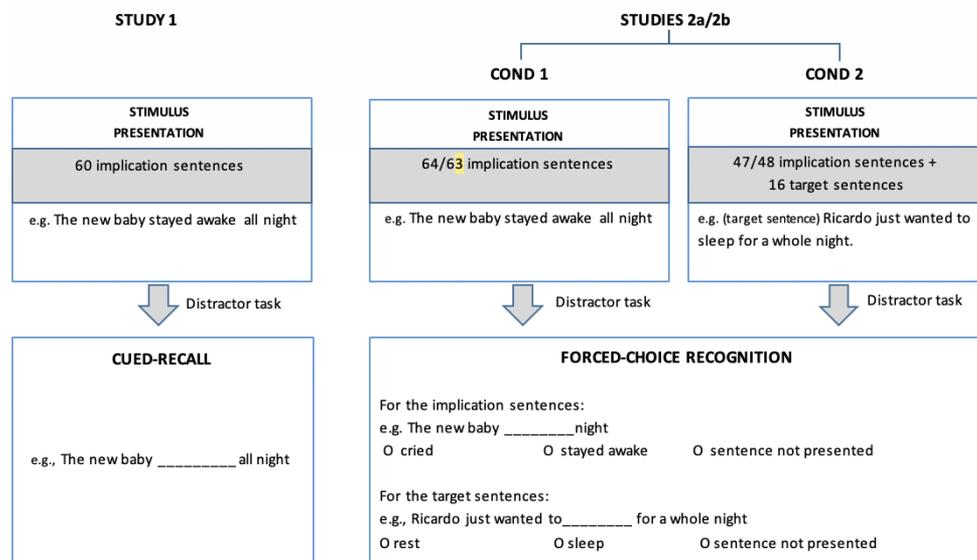

Fig. 1. Scheme of the three studies

RESULTS AND DISCUSSION

All the meaningful and legible responses were registered in the database. Spelling errors were corrected, but words were kept in their original forms as much as possible. The criteria to group words together included the following: singulars and plurals, feminine and masculine, and diminutives. Responses from each participant were classified in one of the following four categories (e.g., for the sentence "the new baby *stayed awake* all night"): correct response (C) "*stayed awake*", pragmatic inference (PI) "*cried*", omission (O) (*no response, blank*), or intrusion (I) (e.g. *slept, laughed, ate*).

Three criteria were used to classify the responses, as in Brewer (1977) study. With a strict criteria, C consisted of the responses written exactly as the original sentences; PI consisted of the responses that corresponded exactly to the ultimate consequence expected and meet the *but not* test; O consisted of the cases in which participants did not write anything; and I consisted of all the other responses different from the correct or pragmatic implications. With the standard criteria, the classification was almost the same as the strict, except that synonyms were accepted as correct or pragmatic implications responses, as long the meaning of the sentence was preserved. With the liberal criteria, every response that meet the *but not* test was considered a pragmatic inference and thus, scored as an implication.

Appendix A displays the C and PI scores of each sentence according to the three criteria. Notably, there is a substantial range of proportion for the pragmatic inferences elicited by all 60 sentences (from 0 to .60 with the strict and standard criteria and from 0 to .69 with the liberal criteria).

In Table 1 we show the mean proportions for all the types of responses in the three criteria. The omission of responses was high (.36) but did not differ much from the proportion of omissions found in Brewer's study (.34). Moreover, the proportions obtained in PI were quite similar to Brewer's results (.21, .26, and .31 in Brewer's study; and .22, .24, and .31 in the current study, respectively in the strict, standard, and liberal criteria). The only evident difference regards the correct responses that were much lower in Brewer's study than in this study (.12, .19, and .19, in Brewer's study; and 27, .31, .31 in this study, respectively, in the strict, standard, and liberal criteria). This could be because in Brewer's study, participants were only provided with the first substantive and asked to recall the remaining sentence, probably increasing the difficulty of the task. Moreover, instructions did not include a reference to memorizing the exact wording of the sentences, as ours did.

Table 1.Mean proportions and standard deviations of cued recall for Correct (C), Pragmatic implication (IP), Omission (O) and Intrusion (I) responses (Study 1) across different classification criteria.

	strict		standard		liberal	
	M	*SD*	*M*	*SD*	*M*	*SD*
C	.27	.12	.31	.13	.31	.13
PI	.22	.10	.24	.11	.31	.13
O	.36	.16	.36	.16	.36	.16
I	.14	.08	.09	.06	.02	.02

The comparison among the proportions of the three types of responses (C, PI, and I) with a strict criteria using a within ANOVA was significant, $F(2,138) = 28.26$, $MSE = .01$, $p < .001$, $\eta_p^2 = .29$. This significant effect showed that the proportion of C responses was higher than the proportion of PI responses ($M = .27$, $SD = .12$ vs. $M = .22$, $SD = .10$; $p = .03$, Bonferroni post-hoc comparisons) and the proportion of I responses ($M = .27$, $SD = .12$ vs. $M = .14$, $SD = .08$; $p < .001$). The same analysis with the standard criteria also showed a significant effect, $F(2,138) = 76.82$, $MSE = .01$, $p < .001$, $\eta_p^2 = .53$, with the proportion of C responses higher than the proportion of PI responses ($M = .31$, $SD = .13$ vs. $M = .24$, $SD = .11$; $p = .02$), and with PI responses higher than the proportion of I responses ($M = .24$, $SD = .11$ vs. $M = .09$, $SD = .06$; $p < .001$). With the liberal criteria, a significant effect also emerged, $F(2,138) = 146.90$, $MSE = .01$, $p < .001$, $\eta_p^2 = .68$, with equal proportions between C and PI responses ($M = .31$, $SD = .13$ for both response types; $p = .1$) and lower proportions of intrusions ($M = .02$, $SD = .02$; $p < .001$).

To make PI comparisons between our data and the pilot work of English norms (presented in the Appendix of McDermott and Chan's, 2006 study) we compared the same 47 pragmatic sentences used in both studies and found that the probability of false recall for these sentences using the standard criteria was on average much lower in our study ($M =. 24$, $SD = .11$, ranging from 0 to .57) than the probability of false recall presented for the same sentences in that pilot work ($M = .47$, $SD = .19$; ranging from .21 to .91).

Moreover, we performed a correlation between the English and the Portuguese data of PI for the similar 47 sentences. The result showed that the correlation was not significant, independently of the adopted criteria (strict: $r = .24$, $p = .11$; standard: $r = .28$, $p = .05$; liberal: $r = .20$, $p = .18$). Only with

the standard criteria was the correlation marginally significant. Moreover, we did a correlation between the two data of PI, excluding the five sentences that were not literally translated from English, and we also found a nonsignificant correlation ($r = .19$, $p = .24$, $n = 42$). Thus, probably because of linguistic or cultural differences, the proportion of inferences for each sentence differs among distinct languages or countries.

In the current study, the 13 pragmatic sentences created by the authors of this study produced similar proportions of C, PI, O, and I compared with the other 47 sentences. This result suggests that the 13 created sentences might also be appropriate for testing pragmatic inferences. Only one sentence from the McDermott and Chan study did not produce an inference "The flimsy shelf *weakened* under the weight of the books". Notably, this sentence produced one of the highest levels of false recall in English (*PI* = .73). Therefore, we suggest that this sentence might not be appropriate to be used in further recall studies of pragmatic inferences in Portuguese. This highlights the need to consider cultural and linguistic differences when using this type of stimuli. Four other sentences produced one persistent intrusion at a higher level of recall than the actual pragmatic inference defined before the study (e.g., for the sentence "King Kong *stood on* the Empire State Building", based on the English norms, we expected the pragmatic inference to be "*stood on top of*" but instead we found "*climbed*" to be more common). Because these intrusions pass the *but not* test, they can better function as PIs. Hence, in the following studies pragmatic inferences for these sentences were substituted with the persistent intrusions found in this study.

STUDY 2a/2b: Forced-Choice Recognition of Pragmatic Sentences

With these studies we aimed to extend the findings of Study 1 from a cued recall task to a three-alternative forced choice recognition task. Overall, Studies 2a and b followed the same procedure as Study 1 except that in the test phase, when participants were presented with sentence fragments, instead of being asked to recall and write the missing critical information, they were given the choice of three responses: the information presented in the study phase (C) the inference for that sentence (PI) or the option that indicated that the sentence had not been presented before (NP). The methodology applied to Study 2a and Study 2b was identical. The only difference was that Study 2a was run in the USA with American participants and tested with the PI sentences in English, while Study 2b was run in Portugal with Portuguese participants and tested with the same sentences, but in Portuguese.

The second aim of these studies was to analyse whether there are differences in the recognition means of C, PI and NP responses between American and Portuguese samples and whether data for each sentence positively correlate in these two populations. That is, we sought to know if there are similarities or dissimilarities between the two types of cultures/ languages in participants' responses to pragmatic inferences.

The last aim of these studies was to understand how structural differences in information's presentation affect the reconstructive aspect of memory. Higham and Brooks (1997) showed that even when participants could not describe the selection rules of categories, they could discriminate unpresented words consistent with the selected rules, from inconsistent new words. This structural effect indicates that participants are somehow tacitly sensitive to the structure of memory lists. This is why some studies that explore the occurrence of other types of inferences, for example, spontaneous trait inferences include sentences that are the target of the experiment, but also other sentences just to distract participants from identifying the structure of trait-inference sentences. For example, in the false recognition paradigm, one of the most used paradigms to study trait inferences (Todorov & Uleman, 2002, 2003), experimental trait-implying sentences of personality are presented during encoding, such as "He stepped on his girlfriend's feet during the foxtrot", which leads participants to falsely recognize the word "clumsy" as having been presented in the sentence in a later memory test. However, in order to prevent participants from realizing the sentences' structure, other sentences are also included in which the personality trait word is actually presented, for example, "He was so aggressive that he threatened to hit her unless she took back what she said". The inclusion of these trials in which the to-be-inferred trait is included is essential, otherwise participants could realize that the correct answer would be always to say that the trait was not previously presented in the sentence.

The ability to learn from the structure or format of the presentation made us question if participants might grasp some aspects about the format of the sentences prone to pragmatic inferences, and therefore answering without relying only on memory traces. In other words, we intend to clarify in the present study, if other strategies, different from the mnesic ones, are taking place when recognizing sentences, and how this eventually affects the level of false memories (pragmatic inferences) obtained. Thus, in the following studies, we manipulated the sentences' format to analyse whether it can influence the level of false inferences produced.

In a typical pragmatic inference paradigm, participants are presented with sentences in their implication form, where the more unusual, or pre-consequent alternative is always presented in the study phase (e.g., "The

unskilful skateboarder *lost his balance on* the skateboard"), and the inference alternative is usually the more usual, precise or the ultimate consequence alternative in that context (e.g. "*fell off*"). To test the hypothesis that, instead of using memory, participants could consciously, or unconsciously, detect the format of the sentences and respond accordingly to this strategy, they were randomly assigned to one of two conditions. In the first condition (Cond 1) participants studied all the sentences in their implication form (e.g., "On the day of her interview, she *chose* her lucky shirt"), as they would in a typical pragmatic inference paradigm. In the second condition (Cond 2), participants also studied the sentences from McDermott and Chan (2006) in this same form but the new sentences created by the authors were presented in the target form (e.g., "On the day of her interview, she *wore* her lucky shirt," instead of "chose"). Hence, to break the eventual detection of the sentences format, in the study phase, the critical information of these new sentences was the actual pragmatic inference expected. In both conditions, the same sentences from McDermott and Chan had no modifications. In the test phase, for all the sentences, the same response alternatives were presented, namely the information presented in the study phase, the pragmatic inference for that sentence, or that the sentence had not been presented before (e.g. *chose* vs. *wore* vs. *Not presented*). If the results show significant differences between the recognition proportions of inferences for the constant sentences from McDermott and Chan between Cond 1 and 2, this could mean that participants' response to these sentences in Cond 1 was contaminated by using the strategy of learning the sentences' format, not relying only on memory.

METHOD

Participants. One hundred American volunteers participated in study 2a (M_{age} = 30.25, 55 female) and seventy-eight Portuguese volunteers participated in study 2b (M_{age} = 25.18, 48 female). All of them provided consent and were rewarded for their participation.

Materials. Study 2a included 64 sentences: all the 48 English sentences used in the study by McDermott and Chan (2006) and 16 pragmatic sentences created by the authors of the present study (13 already used in Study 1 and more 3 new sentences, all of them translated into English). Study 2b included 63 sentences: 47 sentences used in the study by McDermott and Chan (2006) translated or adapted to Portuguese (as described in Study 1, one sentence was excluded because of difficulties in translation) and the same 16 pragmatic sentences created by the authors presented in Portuguese.

The 47/48 sentences were the same in Cond 1 and Cond 2, whereas the 16 new pragmatic sentences were only presented in the implication form in Cond 1. In Cond 2 these 16 sentences were presented in the target form (e.g. "On the day of her interview, she *wore* her lucky shirt") instead of in the implication form[2] (e.g., "On the day of her interview, she *chose* her lucky shirt") (see Appendix B). For both conditions (Cond 1 and 2), participants performed a three-alternative forced-choice recognition task. A fragment of each sentence was presented with three response alternatives: the correct response, matching the actual sentence previously presented; the pragmatic inference expected for each fragment, (according to the results from Study1[3]); and the *not presented* alternative, if participants did not remember seeing the sentence in the study phase (cf. McDermott & Chan, 2006).

Procedure. The procedure for the presented study is depicted in Figure 1. Both studies followed the same procedure. Participants were informed that they had to memorize sentences presented on the computer for a later memory task and that during the experiment they had to answer to some questions only to confirm that they were paying attention to the task (five attention checks).

In both studies, participants were randomly assigned to Cond 1 and 2. Before starting the task, all participants completed a practice trial with five sentences (without pragmatic implications). The practice sentences were presented on the computer screen and were followed by a distractor task (*spot the differences*) for 30 seconds, after which participants were asked to do a practice forced recognition task for these sentences. Sentences were presented on the computer for 4.5 seconds each, in a random order, with a 500 ms interval between them. Amid presentation of the sentences, participants were asked twice to respond to attention checks: they were asked to press an *arrow* button in order to resume the presentation of the sentences and were given a 10 seconds window to do so (a timer was displayed). After studying all the sentences (64 in Study2a and 63 in Study2b), participants performed the same distractor phase as in Study 1 for 1 minute. Immediately afterwards, they were asked to complete a forced –recognition task. Each fragmented sentence was presented on the computer screen, one at a time, with the critical information missing. Participants were, at the same time,

---

[2] Even though these 16 pragmatic sentences could be presented in their implication or target form, the critical fragments that changed between the two forms did not differ in word frequency in the Portuguese version, $t(15) = 0.11$, $p = .91$ (Study 2B, $M_{targets} = 14.50$ vs. $M_{inferences} = 13.24$). Norms regarding word frequency in English were not collected).

[3] As already aforementioned in the Results section of Study 1, the persistent intrusions for 4 sentences were classified as PIs in Studies 2a and 2b because they produced higher levels of false cued recall than the expected PIs of each sentence.

presented with the three response alternatives –correct response, inference or not presented - and were instructed to press the alternative that they thought had been previously presented in the study phase. Intermixed with the fragmented sentences, three additional attention checks were presented: a sentence asking them explicitly to choose a given response option and three response options (the one mentioned, a decoy, and the *Not presented* option). The presentation order for the fragmented sentences and attention checks was randomized and so was the position in which the correct and the pragmatic inference responses appeared. The *Not present* response option was always the third option.

RESULTS AND DISCUSSION (Studies 2a/2b)

Data from participants who failed the attention checks prompted during the procedure were excluded from data analyses. Hence, 24 American participants were excluded from the results in study 2a, with the remaining seventy-six American participants being considered for data analysis for this study ($M_{age} = 30.3$, 49 female); and 21 Portuguese participants were excluded from the results in study 2b, with the remaining fifty-seven Portuguese participants being considered for data analysis for this study ($M_{age} = 24.9$, 36 female).

Appendix B displays the proportions of recognition for the correct responses (C), pragmatic inferences (PI) and not presented option (NP) for each sentence, separately for each condition and study. Together the two studies - 2a and 2b - elicited a substantial range of proportion of recognition of pragmatic inference responses in the condition where all sentences were presented in its implication form, that is, in Cond 1 (ranging from .0 to .81). The correlations of responses between Study2a and Study2b in that condition, computed sentence by sentence, were all positive. However, the correlation for C responses was nonsignificant, $r = .25$, $p = .05$, whereas for PI and NP responses were significant, although not high (respectively, $r = .35$, $p = .004$; $r = .36$, $p = .004$). Even without the five sentences that were not literally translated (see Note 1) the correlation for C and PI responses were still nonsignificant or low (respectively, $r = .26$, $p = .05$, $n = 58$; $r = .36$, $p = .006$, $n = 58$). These results suggest that the same sentences did not elicit the same level of C responses and that were not highly equivalent for PI and NP responses between participants of different cultures. It seems that when the methodology remained constant, but participants were from different countries (Study 2a and 2b) the results obtained sentence by sentence were not similar. This result corroborated the lack of significant correlation found in Study 1 for cued-recall between our data and that of McDermott and Chan

(2006). It seems that the interpretation of each sentence is dependent of linguistic and cultural specificities, and hence, the proportion of inferences and, mainly of correct responses, for each sentence differed among distinct languages.

Table 2 presents the mean proportions for the three types of responses of the recognition test (C, PI, and NP) for all 63 sentences (the sentence not presented in Portuguese was omitted from this Table and subsequent analyses) in Cond 1 for each study, and also the mean proportions for the 47 sentences from McDermott and Chan, and mean proportions for the 16 sentences created by the authors in each condition for each study.

Table 2. Mean proportions and standard deviations of recognition for Correct (C), Pragmatic Implication (IP), and Not Presented (NP) responses (Studies 2a and 2b) presented in the implication form (Cond 1) and separated by Condition (1 and 2).

	Study 2a					**Study 2b**			
	Cond 1					Cond 1			
For the 63 sentences – implication form					For the 63 sentences – implication form				
	M	*SD*					*M*	*SD*		
C	.52	.16				C	.51	.13		
PI	.36	.13				PI	.30	.09		
NP	.12	.12				NP	.18	.11		
	Cond 1		Cond 2			Cond 1		Cond 2	
For the 47 sentences (McDermott & Chan, 2006)					For the 47 sentences (adapted from McDermott & Chan, 2006)				
	M	*SD*	*M*	*SD*			*M*	*SD*	*M*	*SD*
C	.52	.16	.52	.16		C	.50	.14	.47	.14
PI	.38	.13	.34	.12		PI	.32	.09	.30	.12
NP	.10	.11	.14	.12		NP	.17	.11	.22	.15
For the 16 sentences (translated from Portuguese)					For the 16 sentences (presented in Portuguese)				
C	.54	.17	.61	.16		C	.53	.16	.63	.18
PI	.29	.17	.18	.13		PI	.28	.11	.14	.11
NP	.16	.17	.20	.18		NP	.19	.14	.23	.16

As we can see from Table 2 the mean proportions for each type of alternative in both Studies were very similar.

In Study 2a, the comparison of the 16 implication sentences (Cond 1) created by the authors of this study with the same sentences in their target form (Cond 2) showed that C responses were marginally lower in Cond 1 than in Cond 2, $t(74) = 1.84$, $p = .07$, $d = .004$ (respectively, $M = .54$, $SD = .17$ vs. $M = .61$, $SD = .16$) and PI responses were significantly higher in Cond 1 than in Cond 2, $t(74) = 2.96$, $p = .004$, $d = 0.73$ (respectively, $M = .29$, $SD = .17$ vs. $M = .18$, $SD = .13$). In Study 2b, similar results were obtained for the same 16 sentences. C responses were lower in Cond 1 than in Cond 2, $t(55) = 2.34$, $p = .02$, $d = 0.59$ (respectively, $M = .53$, $SD = .16$ vs. $M = .63$, $SD = .18$) and PI responses were significantly higher in Cond 1 than in Cond 2, $t(55) = 4.84$, $p < .001$, $d = 1.27$ (respectively, $M = .28$, $SD = .11$ vs. $M =$

.14, SD = .11). This similar result obtained in both studies suggested that when the sentences were presented in their target form they were more correctly recognized and, mainly, less vulnerable to errors, that is, to false inferences.

To analyse the sentences' format effect, the McDermott and Chan pragmatic sentences were compared in Cond 1 and Cond 2. They were analysed by ANOVAs 2 (study condition: Cond 1 vs. Cond 2) manipulated between-participants x 3 (response type: C vs. PI vs. NP) manipulated within-participants, separated for each study. Study 2a showed only a main effect for response type, $F(2, 148) = 108.56$, *MSE* = .03, $p < .001$, $\eta_p^2 = .59$, with the proportion of C option (*M* = .52, *SD* = .16) being higher than the PI option (*M* = .36, *SD* = .13) and NP option (*M* = .12, *SD* = .12), and the proportion of PI being higher than NP (for both comparisons, Bonferroni tests, $p < .001$). No significant interaction was found between study condition and response type. Study 2b also showed only a main factor of response type, $F(2, 110) = 49.48$, *MSE* = .02, $p < .001$, $\eta_p^2 = .47$, with the proportion of C option (*M* = .48, *SD* = .14) higher than PI option (*M* = .31, *SD* = .10) and NP option (*M* = .20, *SD* = .13) (for both comparisons, Bonferroni test, $p < .001$). The proportion of PI was also significantly higher than NP (Bonferroni test, $p < .001$). No significant interaction was found between study condition and response type.

The similar mean results found for the 47 sentences in both conditions and conducted with populations from different countries showed that participants' responses were not affected by the format of the sentences. This means that, contrary to what was predicted, we did not observe an effect of sentence format learning.

GENERAL DISCUSSION

The importance of this study can be summarized in the following points: a) It presented for the first-time normative data for cued recall and recognition of respectively 60/63 pragmatic implication sentences in Portuguese; b). Pragmatic inferences sentences seem to be language specific since cued-recall and recognition results, analysed sentence by sentence, showed low correlations between American and Portuguese participants; and, finally, c) the effect of sentence format was not found for pragmatic implication sentences.

Cued Recall Comparative Data. In the present study, the cued-recall data for the 47 sentences translated from McDermott and Chan study showed

similar levels of correct recall and inferences when the data were analysed by a liberal criterion. When the strict and standard criteria were used, the mean proportion of cued recall for correct responses was higher than for the pragmatic inferences. The other 13 implication sentences, created by the authors of this study, had levels of cued-recall inferences similar to the correct 47 sentences, which could mean that they are appropriate to study false memories.

By comparing the present results of cued-recall with that of Brewer's study, we found similar levels of cued-recall for the inferences, regardless of which criteria were used. Obviously the more liberal criteria usually produced higher levels of inference responses because it includes the expected inference and also all other words that follow the *but not* criteria. However, in a comparison between the levels of false recall for implication sentences presented in the Appendix of the study of McDermott and Chan, our data demonstrated much lower levels of false recall (mean proportions = .22 strict/ .24 standard/ .31liberal vs. .47 standard in McDermott & Chan). Furthermore, the amount of inferences obtained for the same sentences in each language did not significantly correlate with each other: the sentences that produced the highest and lowest levels of false recall in Portuguese and English were different. Thus, it seems that each culture has specific situations and language terms that are more appropriate to study false memories with pragmatic implication sentences. The consequence is that the ranking order from the highest to lowest levels of false recall in Portuguese and English language is different. We recommend that instead of using pragmatic sentences translated from English, researchers should create or use normative studies in the language of the participants to obtain the specific ranking of sentences for that population. In the future, it might be interesting to explore what differences in culture or in language might explain the low correspondence between the two samples. Not only it can shed some light regarding how culture and language may shape our processes but it can also inform future researchers when creating or adapting research materials.

Recognition Comparative Data. Study 2b also presents, for the first time, normative data for forced-choice recognition of 63 pragmatic implication sentences in Portuguese. Studies 2a and 2b provided recognition data for correct responses and inferences, one done with American participants and lists presented in English and the other with Portuguese participants with the same lists in Portuguese. In both studies the correct responses (C) were higher than the inference responses (PI) and these responses higher than the *not presented* responses (NP). This pattern of recognition results is similar to other studies in this area (e.g., Chan & McDermott, 2006). It is not surprising that the levels of inferences were lower

than the correct responses because in the forced recognition task both alternatives (C and PI) were simultaneously presented, which makes the comparison between the two easier.

The mean values for correct and inference responses were very similar between Study 2a and Study 2b. This means that when participants did the task with a similar methodology, although from different nationalities, the general results for the levels of correct recognition and inferences were equivalent. While the general results were very alike, the responses for each sentence were different, such that the correlation, sentence by sentence, between the two studies was low. This poor correlation found for recognition sentence by sentence between Study 2a and Study 2b suggested that the pragmatic implication sentences that produce the highest levels of false memories in English and in Portuguese were not the same. This reinforced the conclusion of cued-recall data. Before testing participants of different countries this type of material should be normed for their own language.

The Pragmatic Inference Structure. By manipulating the format or structure of sentences at presentation, we were able to analyse if other strategies, different from relying on memory, were at play. Studies conducted with American participants (Study 2a) and with Portuguese participants (Study 2b) did not show that individuals responded according to the format of pragmatic sentences.

It is important to take into account that identifying the sentence format may be easier and more evident in other inferential study paradigms, such as in the case of spontaneous trait inferences (e.g., Todorov & Uleman, 2002). In this case, all sentences presented in the study phase are all behaviours that strongly imply a personality trait which makes it easy for participants to perceive its underlying common format and to use this format as a response strategy. However, in the case of pragmatic inferences, the sentences have extremely varied forms and contents, which make it more difficult to extract a common format. However, we believe that this aspect should be explored in greater detail in future studies.

CONCLUSIONS

This study adds to the literature by contributing to the body of research on false memories conducted with pragmatic implication sentences. This type of false memories (inferences) seems to occur every day in our lives and in all types of cultures, having practical and important implications.

For the first time, data of cued-recall and forced recognition of pragmatic implication sentences were presented in Portuguese, contributing

to further research in the areas of false memories, reasoning, and language done in Portuguese speaking countries. More important, the results have generalizability for all languages because our two main conclusions are pertinent to pragmatic implication research in general. Firstly, we found that the amount of false recall and false recognition produced by each implication sentence was strictly dependent on the language or specific culture of participants, which lead us to propose the use of specific norms for the language or culture of participants in further studies. Secondly, the presentation format of pragmatic sentences did not influence the amount of false recognitions obtained. We believe that participants were not easily detecting the format of the pragmatic sentences and were not responding accordingly. Instead, it seems that they were using their memory and the inferential processes that are inherent to it.

REFERENCES

Bartlett, F.C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press, Cambridge, UK.

Bransford, J.D., & Franks, J.J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2*, 331–350. https://doi.org/10.1016/0010-0285(71)90019-3

Brewer, W.F. (1977). Memory for the pragmatic implications of sentences. *Memory and Cognition, 5*, 673–678. https://doi.org/10.3758/bf03197414

Chan, J. C. K., & McDermott, K. B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology, 20*, 633–639. https://doi.org/10.1002/acp.1215

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22. https://doi.org/10.1037/h0046671

Fritz, C. O., Morris, P. E., Bjork, R. A., Gelman, R., & Wickens, T. D. (2000). When further learning fails: Stability and change following repeated presentation of text. *British Journal of Psychology, 91*, 493–511. https://doi.org/10.1348/000712600161952

Higham, P.A., & Brooks, L.R. (1997). Learning the experimenter's design tacit: Sensitivity to the structure of memory lists. *The Quarterly Journal of Experimental Psychology, 50*, 199-215. https://doi.org/10.1080/027249897392297

Howe, M. J. A. (1970). Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology, 61*, 214-219.

Kay, H. (1955). Learning and retaining verbal material. *British Journal of Psychology, 46*, 81–100. https://doi.org/10.1111/j.2044-8295.1955.tb00527.x

McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition, 34*, 1273-1284. https://doi.org/10.3758/bf03193271

Mullet, H. G., & Marsh, E. J. (2016). Correcting false memories: Errors must be noticed and replaced. *Memory & Cognition, 44*, 403-412. http://dx.doi.org/10.3758/s13421-015-0571-x

Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814. https://doi.org/10.1037//0278-7393.21.4.803

Sampaio, C. & Brewer, W.F. (2009). The role of unconscious memory errors in judgments of confidence for sentence recognition. *Memory & Cognition, 37*, 158-163. https://doi.org/10.3758/mc.37.2.158

Sulin, R.A., & Dooling, D.J. (1974). Intrusion of a thematic idea in the retention of prose. *Journal of Experimental Psychology, 103*, 255–262. https://doi.org/10.1037/h0036846

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of personality and social psychology*, *83*, 1051-1065. https://doi: 10.1037//0022-3514.83.5.1051.

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, *39*, 549-562. https://doi.org/10.1016/S0022-1031(03)00059-3.