A replication of Coulson & Williams (2005)

Joshua R. de Leeuw[1], Daniel P. Bliss[1], Martin Burstein[1], Nona Chen[1], Julissa Coplin[1], Duc Dang[1], Mira Genkovska[1], Chuqi Hu[1], Dora Law[1], Emma Leshock[1], Natasha Orellana[1], Shivani Pandey[1], Yaser Pena[1], Naima Saini[1], Raia Stern[1], Orcun Tasdemir[1], Yuchen Wang[1], Ava Waters[1], Zachary Watson[1], Lily Yan[1], & Yuchen Zhou[1]

[1] Department of Cognitive Science, Vassar College

Author Note

The authors made the following contributions. Joshua R. de Leeuw: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing; Daniel P. Bliss: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization; Martin Burstein: Formal analysis, Investigation, Methodology, Visualization; Nona Chen: Formal analysis, Investigation, Methodology, Visualization; Julissa Coplin: Formal analysis, Investigation, Methodology; Duc Dang: Formal analysis, Investigation, Methodology; Mira Genkovska: Formal analysis, Investigation, Methodology; Chuqi Hu: Formal analysis, Investigation, Methodology; Dora Law: Formal analysis, Investigation, Methodology, Software; Emma Leshock: Formal analysis, Investigation, Methodology; Natasha Orellana: Formal analysis, Investigation, Methodology; Shivani Pandey: Formal analysis, Investigation, Methodology; Yaser Pena: Formal analysis, Investigation, Methodology; Naima Saini: Formal analysis, Investigation, Methodology; Raia Stern: Formal analysis, Investigation, Methodology; Orcun Tasdemir: Formal analysis, Investigation, Methodology, Software; Yuchen Wang: Formal analysis, Investigation, Methodology; Ava Waters: Formal analysis, Investigation, Methodology, Software; Zachary Watson: Formal analysis, Investigation, Methodology; Lily Yan: Formal analysis, Investigation, Methodology, Software; Yuchen Zhou: Formal analysis, Investigation, Methodology.

Correspondence concerning this article should be addressed to Joshua R. de Leeuw, Enter postal address here. E-mail: jdeleeuw@vassar.edu

Abstract

ADD LATER

Keywords: keywords

Word count: X

A replication of Coulson & Williams (2005)

Introduction

The functional difference in language processing between the two cerebral hemispheres has been extensively studied, with each hemisphere suggested to be responsible for a separate yet complementary role (Beeman & Chiarello, 1998). Dating back to the influential findings of Broca's area and Wernicke's area, it has been shown that damage to the left hemisphere impairs speech production (Broca, 1861) and language comprehension (Wernicke, 1874), and later lesion studies further extend the functional relevance of the left hemisphere to speech perception (Blumstein, 1994) and object naming (Damasio, 1992). On the other hand, damage to the right hemisphere does not seem to interrupt the aforementioned basic language functions, but rather seems to elicit less obvious language-related impairments, such as the inability to comprehend familiar idiomatic phrases (Van Lancker & Kempler, 1987), metaphoric statements (Winner & Gardner, 1977; Brownell, 1988), discourse inferences (Beeman, 1993; Brownell et al., 1986), and indirect requests (Stemmer et al., 1994). This hemispheric difference in language processing has also been found in non-lesion, non-invasive studies: while Szymanski et al. (1999) magnetoencephalography (MEG) study finds that the left hemisphere is more responsive to vowels, Mashal et al. (2007) study using functional magnetic resonance imaging (fMRI) implies that the right hemisphere is more specialized in processing novel metaphors. From these results, it can be theorized that while the left hemisphere is responsible for fundamental processes of language production and comprehension, the right hemisphere is necessary for high-level semantics and pragmatics tasks that require integrating the explicit utterance with context and background knowledge (Joanette et al., 1990).

Understanding joke expressions involves the integration of explicit meaning and background information. For example, consider the joke "Statistics indicate that Americans spend 80 million a year on games of chance, mostly weddings" (Coulson & Kutas, 2001). Its structure sets readers up to have a background expectation of the game-of-chance category, then subverts the expectation by matching weddings to such category. Hence, comprehending the joke may require readers to recall prior knowledge on the game-of-chance category and weddings, reorganize such existing information into a new frame with weddings as a game of chance, then draw inferences on the whole utterance — Coulson (2001) coins these mental operations "frame-shifting." With frame-shifting and other mental processes included in joke comprehension being high-level language tasks that are often attributed to the right hemisphere, there is a question of whether the right hemisphere also has functional specializations in joke comprehension. Some studies have supported this hemispheric differences: an fMRI study shows that joke comprehension elicits higher right hemisphere activity (Goel & Dolan, 2001) and multiple lesion studies find correlations between right hemisphere damage and difficulty understanding humor, while left hemisphere damage is often correlated with more drastic and severe communicative impairments (Bihrle et al., 1986; Brownell et al., 1983, Gardner et al., 1975).

Contributing to the evidence of hemispheric difference in joke comprehension, Coulson & Williams (2005) electroencephalography (EEG) study finds that when presented to the left hemisphere via the right visual field (RVF-LH), sentences with unexpected joke endings elicit higher N400 amplitude than sentences with unexpected non-joke endings, yet when presented to the right hemisphere via the left visual field (LVF-RH), both jokes and non-jokes elicit equal N400 amplitude. N400 is a component of the event-related potential (ERP) with a negative-going deflection peaking around 400 ms, and is initially discovered to be elicited in response to semantically incongruous words in a sentence (Kutas & Hillyard, 1980). Later studies finds that within a given sentence context, a more unpredictable word, also known as a word of lower cloze probability, elicits a higher N400 amplitude (Kutas & Hillyard, 1984); hence, N400 can be viewed as an indicator of how difficult a word is for integration into a given sentence. The N400 pattern also extends to the domain of joke comprehension: in particular, Coulson & Kutas (2001) found the "N400 joke effect", in which sentences with a joke ending, such as "Statistics indicate that Americans spend 80 million a year on games of chance, mostly weddings" elicit a higher N400 amplitude than the same sentences with non-joke but equally unexpected ending, such as "Statistics indicate that Americans spend 80 million a year on games of chance, mostly dice." Such N400 joke effect is shown in Coulson & Williams (2005) to be higher in the left hemisphere than in right hemisphere, implying that the right hemisphere has less difficulty integrating unexpected joke endings into a sentence and has a stronger semantic activation for jokes, which facilitates overall joke comprehension; such findings are in favor of the right hemisphere functional dominance in joke comprehension.

Despite Coulson & Williams (2005) being an influential paper — 220 citations on Google Scholar — that provides important evidence for theorizing the right hemisphere's pragmatics functions (Blake, 2017), developing a neural model of humor (Marinkovic et al., 2011), and even hypothesizing a cerebral model of interhemispheric differences and intrahemispheric collaboration in language processing (Federmeier et al., 2008; Jung-Beeman, 2005), there has been no published direct replication of the study. While some experiments with similar methodology of divided visual field paradigm and N400 recording have supported the study's general conclusion (Marinkovic et al., 2011, Wlotko & Federmeier, 2013), there have also been experiments that do not find a hemispheric differences in pragmatic tasks such as comprehending metaphors (Coulson & Van Petten, 2007; Kacinik & Chiarello, 2007). We design our current study to be a direct replication of Coulson & Williams (2005), using the same methodology and more targeted statistical analyses. If the original study finds a robust result that there is a stronger semantic activation during joke comprehension in the right hemisphere, then our current study should find the same result of a stronger N400 joke effect in the left hemisphere. If such results are not replicated, we then discuss how the differences between our current study and the original study can account for the replication failure.

This experiment was part of an undergraduate research methods course in cognitive science, which 2 of us co-taught and 20 of us were enrolled in. A major focus of this course was exposure to and training in practices that have developed in response to the replication crisis, including an increased emphasis on direct replications (Zwaan et al., 2017), pre-registration of experiments (Wagenmakers et al., 2012), and transparency through public sharing of materials, data, and analysis scripts (Nosek et al., 2015). To gain hands-on experience with these practices, the class conducted this replication study. We chose to replicate Coulson & Williams (2005) given its theoretical significance in the field, lack of prior direct replications, and practical considerations like the complexity of the data analysis and study design. > The purpose of an introduction in a research article is to clearly convey the rationale for the empirical work. The introduction should explain why the study was done, usually by explaining one or more unresolved questions in existing research and/or theory and describing how the experiment will help to answer those questions. For this assignment, this is a short (approx. 3 paragraphs) description about the need for replications in general and the general findings and theoretical relevance of the original study.

Methods

All stimuli, experiment scripts, data, and analysis scripts are available on the Open Science Framework at https://osf.io/38rga. The study pre-registration is available at https://osf.io/3yqah.

Overview

In both the original experiment and our replication, participants viewed short sentences presented to them on a computer screen one word at a time, with every word displayed at the screen center, except only for the last word, which was displayed on either right or left of the screen. Afterwards, participants were instructed to speak out loud the last word of each sentence when a question mark was displayed, and to evaluate each sentence's comprehension statement as true or false. Moreover, similar to the original experiment, participants were also told to focus their eyes at the screen center at all times when reading the short sentences. The short sentences were categorized into three conditions, defined by the original experiment as jokes, non-jokes and expected fillers, and the number of sentences used for each condition was the same as the original experiment: 80 jokes, 80 non-jokes, and 80 fillers.

Participants

46 Vassar College students participated in the study. We pre-registered our target participant number at 40 participants, 2.5 times the original sample size of 16 participants, based on the heuristic provided by Simonsohn (2015) that set replications at sufficient power for small effects to be both plausibly detectable by the replication study and the original study. Participants right-handedness were assessed and selected via the Edinburgh inventory, a self-reported measurement of hand dominance in everyday activities (Oldfield, 1971). Out of the 46 participants we recruited, 9 participants did not complete the experiment due to problems such as not being right-handed, feeling nauseous during the experiment, and facing EEG recording difficulties. Another subject did not generate good segments for various rejection threshold, thus were also excluded from the data analysis. Consequently, although we recruited more participants than our target of 40, we ended up with 36 subjects that successfully finished the experiment. We had to end our data collection prior to having 40 usable recordings on our pre-registered cutoff date of April 3rd, 2022. All participants provided informed consent and this study was approved by the Vassar College Institutional Review Board.

Materials

Materials included 160 pairs of experimental sentences with unexpected endings and 80 filler sentences with expected endings. Each experimental sentence pair had two identical one-line sentences up until the last word, where one ended with an unexpected joke ending and the other ended with an equally unexpected non-joke ending. Every joke was chosen with the requirement that it was not humorous before the last word, hence the punchline could only be realized at the final word, thereby understanding the joke required readers to reinterpret or frame-shift the meaning established previously in the sentence (Coulson, 2001). Each joke had its corresponding non-joke, which was created by replacing the ending of the joke with a non-humorous but equally unexpected ending. In contrast to the unexpected ending of experimental sentence pairs, the ending of filler sentences were highly predictable, as these filler sentences were assembled from the 80 sentences with highest cloze probability ending provided by Blake (2010).

We initially collected about 200 jokes from from various sources, including anthologies of one-liners, materials of other studies on jokes, and the original Coulson & Williams (2005) joke list at https://github.com/mekline/Jokes-Analysis/blob/master/Materials/nonlit_joke/materials.csv, then excluded some jokes that did not match the aforementioned requirement, were potentially sexist, misogynistic or offensive, or were seemingly too difficult to understand. Afterwards, we conducted a cloze experiment with a sentence completion task using our experimental sentence pairs on 200 participants not within the main experiment's participant pool, which resulted in a mean cloze probability of about 5.5% for joke sentences and about 8.5% for non-joke sentences. Since these numbers were quite higher than the cloze probabilities of jokes and non-jokes in the original study, which were respectively 0.9% and 2.2%, we excluded some additional experimental pairs and modified some sentences' ending to match the original cloze probabilities. In the end, our materials consisted of 160 joke/non-joke pairs and 80 filler sentences, with cloze probabilities at 1.71% (SE = 0.22%) for jokes, 3.01% (SE = 0.42%) for non-jokes, and 94.75% (SE = 0.15%) for fillers. Each experimental sentence pair and filler sentence was accompanied by a comprehension question which could be answered yes or no; within each condition, there was an equal number of yes and no answers.

In accordance to the original methodology, we created four different lists of stimuli with a within-participants design where both Sentence Type (jokes/non-jokes/fillers) and

Visual Field (LVF-RH/RVF-LH) were counterbalanced. In particular, each participant saw 80 jokes and 80 non-jokes chosen from different pairs, so that no participant viewed both versions of one experimental pair; presentation at left or right visual field as well as jokes and non-jokes were swapped between stimuli lists for counterbalance purpose. In addition, while fillers were unchanged the same throughout lists, their visual field presentation were also swapped between lists. The 240 sentences in each lists were also randomly split into four blocks of trials, so that participants could have short breaks between blocks.

Procedure

Participants completed the experiment in a quiet room seated at a computer screen and keyboard. The experiment was built using the jsPsych library (de Leeuw, 2015). At least two experimenters were required for each session: One experimenter stayed in the same room as the participant to record the participant's responses to the delayed naming task, while the other experimenter monitored the EEG signal from a different room and informed the former experimenter of any EEG recording anomalies that needed troubleshooting — such as equipment disconnection, excessive noise, or drastic eye and muscle movement — during the breaks between trial blocks. Before the main experiment, every participant must complete at least ten trials of practice sentences not within the main experiment stimuli list. The participant could choose to repeat the practice block if they did not feel comfortable with the task, did not fully understand the instructions, or were told by the monitoring experimenter to be excessively moving their eyes or muscles. Afterwards, the participants could continue with the main experiment.

Each trial started with a fixation cross in the center of the screen, followed by sentences appearing one word at a time also in the center of the screen, up until the final word that displaced either left or right of the screen. Participants were instructed to fixate their eyes centrally at all times, and critically not to move their eyes to see the last word. A blue question mark appeared around 2000 ms after the last word, to which the participant were asked to say the last word aloud, or to say "No" aloud if they had been unable to read it. The experimenter in the same room recorded on a paper form whether or not the correct word was produced. After the blue question mark in each sentence, a comprehension sentence was presented for participants to evaluate whether or not it was consistent with the first sentence, using "Y" on the keyboard for yes and "N" on the keyboard for no. Participants were allowed to move, blink and rest their eyes during the presentation of comprehension questions. Half of the participants responded these questions with their right hand, and the other half responded with their left hand for counterbalancing purpose.

Similar to the original study, stimuli were presented in a black Helvetica font against a white background to maximize contrast. Each word of an experimental sentence was shown for a duration that varied as a function of word-length, 200 ms + 32 ms/character, with the only exception being the final word that was presented for 200 ms and proceeded with a blank screen for 2500 ms. Afterwards, the blue question mark that signaled the delayed naming task appeared for 2 s, continued with a comprehension question that remained on the screen until the participant input an answer, then finally followed by a final blank screen for 2 s until the next trial began.

EEG Recording

We recorded data using a CGX Quick-20r v2 EEG headset at 19 electrodes referenced to the left earlobe: FP1, FP2, F3, F4, F7, F8, FZ, T3, T4, CZ, C3, C4, P7, P8, P3, P4, PZ, O1, and O2. Note that this was difference from the EEG setup in Coulson & Williams (2005), which recorded data at 29 electrodes referenced to the left mastoid, which included the aforementioned electrode sites (T5 and T6 were respectively equivalent to P7 and P8) except for C3 and C4, but also included FPZ, FT7, FC3, FCZ, FC4, FT8, TP7, CP3, CPZ, CP4, TP8, and OZ that were not available in our CGX system. Electrical impedances were minimized before data collection, eye movements (blinks and lateral movement) were monitored via the FP1 and FP2 electrodes, and EEG data with event markers was recorded continuously at 500 Hz (250 Hz for the original study).

After data collection, which also recorded EEG at the right earlobe, EEG data was referenced to the average electrical potential of left and right earlobes and filtered through a signal band of 0.01 Hz through 40 Hz; the original study also did this pre-processing step, but used mastoids instead of earlobes. Baseline was corrected by 44 ms (100 ms for the original study) based on a event markers' latency test, segments with voltage range exceeding 200 μV were excluded, and trials containing excessive eye or muscle movements during presentation of sentences were omitted from analysis. Similar to the original study, ERPs were recorded during a time window extending from 100 ms before the onset of each stimulus to 920 ms after the stimulus. In trials with errors in some electrodes' connections, data from functional electrodes was still included in the analysis, but trials with the participants not succeeding in the delayed naming task were excluded from analysis.

Data Analysis

Similar to Coulson & Williams (2005), we first analyzed the behavioral data of delayed naming accuracy, where participants repeated aloud the last word presented in each sentence after a short delay, and of comprehension accuracy, where participants answered true or false to each sentence's comprehension question. Both delayed accuracy scores and comprehension accuracy scores were converted to numeric values (1 for correct and 0 for incorrect), averaged for scores per subject, and subjected to repeated measures ANOVA with factors Sentence Type (Joke/Non-joke/Filler) and Visual Field (LVF-RH/RVF-RH). Note that unlike the original study, we did not discard a participant's response to the comprehension probe if they were unable to name correct the sentence-final word. The original study's effect we looked to replicate was the effect of sentence type on delayed naming accuracy and comprehension accuracy.

We then analyzed the EEG data, firstly the N1 visual potential that had been suggested to be correlated with the level of participation and attention in visual processing (Hillyard & Anllo-Vento, 1998). We replicated the original study's analysis by measuring the ERPs elicited after stimuli onset from 75 ms to 175 ms at four electrode sites of P7 (T5), P8 (T6), O1, and O2, then conducting a repeated measures ANOVA with factor Hemisphere (LH/RH) and Visual Field (LVF-RH/RVF-LH). We looked to replicate the original study's interaction effect between hemisphere and visual field, in that N1 amplitude was larger over RH electrode sites with stimuli presented to LVF-RH, and larger over LH electrode sites with stimuli presented to RVF-LH.

Finally, we analyzed the N400 joke effect by measuring the ERPs elicited after stimuli onset from 300 ms to 500 ms at three electrode sites of P3, P4, and Pz, then conducting a repeated measures ANOVA with factor Sentence Type (Joke/Non-joke/Filler) and Visual Field (LVF-RH/RVF-LH). Note here that our analysis of the key N400 joke effect was different from the analysis done in Coulson & William (2005), which found a significant effect in a three-way ANOVA with factors of Sentence Type (Joke/Non-joke/Filler), Visual Field (LVF-RH/RVF-LH), and electrode site, then visually observed the largest hemispheric difference in N400 joke effect at the centro-parietal electrode sites of P3, P4, Pz. However, we reasoned that a three-way interaction effect between electrode site, sentence type, and visual field might not be indicative of hemispheric difference in joke comprehension, since the effect could have been elicited by just different N400 values between electrode sites, and electrodes with excessively high or excessively low N400 amplitudes could have caused the significant effect instead of actual hemispheric difference. Furthermore, if the N400 joke effect was robust, then a two-way ANOVA with factors of Sentence Type and Visual Field at the aforementioned electrode sites should be sufficient to find a significant effect. We also conducted a separate repeated measure ANOVA for the only N400 amplitudes in jokes and non-jokes, both referenced to the N400 amplitudes in filler sentences; this two-way ANOVA had factors of Sentence Type (Joke/Non-joke, both referenced to Filler) and Visual Field (LVF-RH/RVF-LH)

A complete methods section should provide sufficient detail that someone could conduct a replication of the experiment without seeking out additional information from the researchers. Note that "sufficient detail" is a subjective judgment about what aspects of the method are crucial to reproduce the study and which aspects are free to change. For example, we don't usually report the clothes that participants wore in the experiment, because we don't believe that the experimental results depend on this factor. A typical methods section has a Participants section, a Materials section, and a Procedure section. I sometimes omit the Materials section in my own work because I find it clearer to describe these details in the context of the procedure. You can choose what to do here. Because this study is a replication, your methods section can be shorter than usual by referring to the original study for details. You should provide enough information that a reader doesn't need to consult with the original study to understand the gist of the experiment, but you don't need to be super detailed. You should pay careful attention to and describe all deviations from the original protocol.

Results

The results section should describe the analysis in sufficient detail that someone could reproduce your analysis if given the raw data. Note that one advantage of an R Notebook is that the code to do the analysis is right there in the document, so this is a pretty easy thing to do in this context! While the focus of a results section is on the analytical work, a good results section will carefully guide the reader through the analysis, explaining why each critical statistical test was conducted (e.g., by connecting it back to the questions raised in the introduction) and doing a little bit of interpretative work to explain the outcomes of each step.

Behavioral

Delayed Naming Task.

analysis_files/figure-latex/Plot Delayed Naming Accuracy-1.

```
##                             Effect DFn DFd        F            p p<.05        ges
## 2                    sentence_type   2  72 71.48964 7.901183e-18     * 0.34168917
## 3                    left_or_right   1  36 38.73041 3.498795e-07     * 0.17676464
## 4 sentence_type:left_or_right   2  72 30.71996 2.257784e-10     * 0.09721801
```

Comprehension Task.

analysis_files/figure-latex/Plot group-level comprehension d

```
##                             Effect DFn DFd         F            p p<.05         ges
## 2                    sentence_type   2  70 96.538326 7.508456e-21     * 0.321185245
## 3                    left_or_right   1  35  2.589677 1.165464e-01       0.004165972
## 4 sentence_type:left_or_right   2  70  2.573716 8.345195e-02       0.008761250
```

EEG

N1.

ERP Figure.

analysis_files/figure-latex/Plot N1 ERPs-1.pdf

Good Segments. This table is the number of good segments for each subject in each cell of the ANOVA (visual_field x hemisphere).

```
## # A tibble: 140 x 4
## # Groups:   subject, visual_field [70]
```

```
##    subject visual_field hemisphere     n
##    <chr>   <chr>        <chr>      <int>
##  1 01      left         left          91
##  2 01      left         right         85
##  3 01      right        left         114
##  4 01      right        right        106
##  5 02      left         left          95
##  6 02      left         right         95
##  7 02      right        left         115
##  8 02      right        right        115
##  9 03      left         left          95
## 10 03      left         right         95
## # ... with 130 more rows
```

The smallest number is 56 and the largest is 120 out of 120 possible segments. Note that these numbers also factor in excluded segments for when a participant did give the correct answer in the delayed naming task.

ANOVA.

```
##                     Effect DFn DFd           F           p p<.05          ges
## 2             visual_field   1  34  0.005093701 0.943521321       9.203357e-06
## 3               hemisphere   1  34  0.576022832 0.453105279       1.739535e-03
## 4 visual_field:hemisphere   1  34 12.457723317 0.001217681     * 3.129732e-02
```

N400.

analysis_files/figure-latex/Plot N400 ERP-1.pdf

ERP Figure.

analysis_files/figure-latex/Plot Difference Waves N400-1.pdf

Difference Waves.

Good Segments. This table is the number of good segments for each subject in each cell of the ANOVA (visual_field x hemisphere).

```
## # A tibble: 210 x 4
## # Groups:   subject, ending [105]
##    subject ending  visual_field     n
##    <chr>   <chr>   <chr>        <int>
##  1 01      filler  left            40
##  2 01      filler  right           40
##  3 01      joke    left            24
##  4 01      joke    right           36
##  5 01      nonjoke left            27
##  6 01      nonjoke right           38
##  7 02      filler  left            38
##  8 02      filler  right           40
##  9 02      joke    left            24
## 10 02      joke    right           36
## # ... with 200 more rows
```

The smallest number is 20 and the largest is 40 out of 40 possible segments. Note that these numbers also factor in excluded segments for when a participant did give the correct answer in the delayed naming task.

ANOVA.

```
##                Effect DFn DFd          F            p p<.05         ges
```

```
## 2        visual_field   1  34  2.9555647 9.467585e-02       0.005774673
## 3              ending   2  68 12.8171963 1.890164e-05     * 0.034744582
## 4 visual_field:ending   2  68  0.9171844 4.045291e-01       0.002620953
```

Follow-up to this result: Is there a difference between non-joke and joke endings? We can use the difference waves to figure this out. Run another ANOVA on just the difference wave data to see if there is an effect of ending.

```
##                Effect DFn DFd         F         p p<.05         ges
## 2        visual_field   1  34 1.1892341 0.2831526       0.009559216
## 3              ending   1  34 0.4911500 0.4881836       0.001170757
## 4 visual_field:ending   1  34 0.4075856 0.5274748       0.001773028
```

Discussion

In summary, our current study successfully replicated all auxiliary findings of behavioral performances and N1 visual potential in Coulson & Williams (2005), but not the key finding of hemispheric difference in joke comprehension. Our behavioral analysis results matched all the patterns of the original study, which confirmed the robustness of multiple behavioral effects. First, expected filler sentences triggered a higher delayed naming accuracy rate than unexpected jokes or non-jokes; furthermore, non-jokes triggered a higher delayed naming accuracy rate than jokes. This suggested that the unexpectedness of a sentence's ending word influences the participants' ability to process, memorize and repeat the word, and humor might also affect such ability. Second, the comprehension accuracy rate of filler sentences was also higher than of either jokes and non-jokes, implying that highly expected sentences were more easily comprehended than unexpected sentences. Finally, there was a reliable effect of visual field in delayed naming accuracy rates, in which RVF stimuli presentation elicited higher accuracy than LVF stimuli presentation, hinting at a potential correlation between visual field of stimuli presentation and word perception or short-term memory.

Furthermore, we also successfully replicated the N1 visual potential effect, in which stimuli presentation in one visual field elicited higher N1 amplitude in the contra-lateral hemisphere. No individual effect but only the interaction effect was found, mirroring the original's study N1 analysis. N1 amplitude had been suggested to be modulated by visual attention (Hillyard & Anllo-Vento, 1998), hence our findings of N1 effect suggested that the divided visual field paradigm was effective its goal of selectively stimulating opposite hemispheres. Nevertheless, we failed to replicate the key finding in our N400 analysis. The only significant effect we found was of sentence type on N400 amplitude, yet the effect was not found in a separate ANOVA analysis where the sentence types were jokes and non-jokes referenced to the N400 amplitude of filler sentences. Therefore, this effect could be qualified as the well-known N400 effect where unexpected sentences elicited a higher N400 amplitude than expected fillers (Kutas & Hillyard, 1984), instead of the N400 joke effect (Coulson & Kutas, 2001). We discussed two possible reasons for the key effects not being replicated below.

One potential reason of replication failure was the differences in EEG equipment, EEG pre-processing, and stimuli. Instead of using original study's 29-electrode EEG system, we only had access to a 19-electrode EEG system, and we also pre-processed our EEG data using a different baseline correction time of 44 ms, not 100 ms in the original study. It was nonetheless unclear how these equipment differences might drastically change the result of the experiment, especially when most of the stimuli and procedure were unchanged. Still, even though we did not excessively change the original stimuli list, we did minimally modify the list by removing jokes that were deemed too offensive, too inappropriate, or too old to be understood by college students. Changing the stimuli, however, might come with the disadvantage of the new jokes not fulfilling the criteria of frame-shifting; in particular, some jokes were mostly based on puns, instead of an actual frame-shift that required participants to re-interpret what they had read previously. Since Coulson & Williams (2005) attributed joke comprehension to the process of frame-shifting, the replication failure of the current study might be accounted for by the fact that a few stimuli did not actually require frame-shifting as a comprehension mechanism.

Another potential source for the replication failure was the differences in statistical analysis: while the original study conducted a three-way ANOVA with factors of Electrode X Visual Field X Sentence Type, our current study only focused on the two-way effect with factors of Visual Field X Sentence Type. Our reasoning was that a two-way interaction effect of visual field and sentence type was more indicative of the hemispheric difference in N400 joke effect than a three-way interaction effect, which could possibly be accounted by excessively high or low values in outlier electrodes. To verify this, we conducted an additional three-way ANOVA similar to the original, but only restricted to centro-parietal electrodes of P3, P4 and Pz. The rationale for the smaller scale three-way analysis was based on how the key findings were inferred in the original study — after finding a significant three-way interaction effect, Coulson & williams (2005) visually observed a strong hemispheric difference in N400 amplitude within the centro-parietal electrodes, then concluded that the differences in N400 amplitude were related to hemispheric difference. Therefore, if the hemispheric difference in N400 joke effect was largest at the centro-parietal electrodes, then we would expect a significant three-way interaction effect there. However, our three-way analysis did not find any three-way interaction effect, but only an individual effect of electrode and a two-way interaction effect of Electrode X Visual Field. Moreover, our three-way analysis done separately on jokes or non-jokes (both still referenced to the N400 amplitude of filler sentences) did not find any significant effect, suggesting that the aforementioned Electrode X Visual interaction effect was mostly modulated by the N400 amplitude's difference in electrodes. Therefore, the original study's effect actually might not be caused by hemispheric difference and the N400 joke effect, but instead simply by random differences in the electrodes reading of the N400 component. Considering that there were new evidence that denied hemispheric difference in pragmatic tasks (Coulson & Van Petten, 2007; Kacinik & Chiarello, 2007), we believed it was likely that the original study's effect was not robust, and there might be no hemispheric difference in understanding jokes comprehension. Further investigations needed to be conducted in order to establish a concrete connection between brain laterality and joke comprehension.

The goal of a discussion section is to answer the question: what do we now know about our original questions that we didn't know before conducting the research? There are many different stylistic approaches to a discussion section, so you'll have to find what is comfortable for you. In this assignment, the discussion should focus on the ways in which our study did or did not replicate the original experiment.

References

Beeman, Mark. "Semantic processing in the right hemisphere may contribute to drawing inferences from discourse." Brain and language 44.1 (1993): 80-120.

Beeman, M. J., & Chiarello, C. (1998). Complementary right-and left-hemisphere language comprehension. Current directions in psychological science, 7(1), 2-8.

Bihrle, A. M., Brownell, H. H., Powelson, J. A., & Gardner, H. (1986). Comprehension of humorous and nonhumorous materials by left and right brain-damaged patients. Brain and cognition, 5(4), 399-411.

Blake, M. L. (2017). Right-hemisphere pragmatic disorders. In Research in clinical pragmatics (pp. 243-266). Springer, Cham.

Blumstein, S. E. (1994). Impairments of speech production and speech perception in aphasia. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 346(1315), 29-36.

Broca, P. (1861). Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech). Bulletin de la Société Anatomique, 6, 330-357.

Brownell, H. H. (1988). Appreciation of metaphoric and connotative word meaning by brain-damaged patients. In Right hemisphere contributions to lexical semantics (pp. 19-31). Springer, Berlin, Heidelberg.

Brownell, H. H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients. Brain and language, 18(1), 20-27.

Brownell, H. H., Potter, H. H., Bihrle, A. M., & Gardner, H. (1986). Inference deficits in right brain-damaged patients. Brain and language, 27(2), 310-321.

Coulson, S. (2001). Semantic leaps: Frame-shifting and conceptual blending in meaning construction. Cambridge University Press.

Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. Neuroscience letters, 316(2), 71-74.

Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension?: ERP evidence from hemifield presentation. Brain research, 1146, 128-145.

Coulson, S., & Williams, R. F. (2005). Hemispheric asymmetries and joke comprehension. Neuropsychologia, 43(1), 128-141.

Damasio, A. R. (1992). Aphasia. New England Journal of Medicine, 326(8), 531-539.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behavior research methods, 47(1), 1-12.

Federmeier, K. D., Wlotko, E. W., & Meyer, A. M. (2008). What's 'Right' in Language Comprehension: Event-Related Potentials Reveal Right Hemisphere Language Capabilities. Language and linguistics compass, 2(1), 1-17.

Gardner, H., Ling, P. K., Flamm, L., & Silverman, J. E. N. (1975). Comprehension and appreciation of humorous material following brain damage. Brain: a journal of neurology, 98(3), 399-412.

Goel, V., & Dolan, R. J. (2001). The functional anatomy of humor: segregating cognitive and affective components. Nature neuroscience, 4(3), 237-238.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. Proceedings of the National Academy of Sciences, 95(3), 781-787. Joanette, Y., Goulet, P., Hannequin, D., & Boeglin, J. (1990). Right hemisphere and verbal communication. Springer-Verlag Publishing.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. Trends in cognitive sciences, 9(11), 512-518.

Kacinik, N. A., & Chiarello, C. (2007). Understanding metaphors: Is the right hemisphere uniquely involved?. Brain and language, 100(2), 188-207.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. Nature, 307(5947), 161-163.

Marinkovic, K., Baldwin, S., Courtney, M. G., Witzel, T., Dale, A. M., & Halgren, E. (2011). Right hemisphere has the last laugh: neural dynamics of joke appreciation. Cognitive, Affective, & Behavioral Neuroscience, 11(1), 113-130.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. Brain and language, 100(2), 115-126.

Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck et al. "Promoting an open research culture." Science 348, no. 6242 (2015): 1422-1425. Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97-113.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. Psychological science, 26(5), 559-569.

Stemmer, B., Giroux, F., & Joanette, Y. (1994). Production and evaluation of requests by right hemisphere brain-damaged individuals. Brain and Language, 47(1), 1-31.

Szymanski, M. D., Rowley, H. A., & Roberts, T. P. (1999). A hemispherically asymmetrical MEG response to vowels. Neuroreport, 10(12), 2481-2486.

Van Lancker, D. R., & Kempler, D. (1987). Comprehension of familiar phrases by left-but not by right-hemisphere damaged patients. Brain and language, 32(2), 265-277.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. Perspectives on Psychological Science, 7(6), 632–638. https://doi.org/10.1177/1745691612463078 Wernicke, C. (1874). Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis. Cohn..

Winner, E., & Gardner, H. (1977). The comprehension of metaphor in brain-damaged patients. Brain, 100(4), 717-729.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2017). Making replication mainstream. The Behavioral and brain sciences, 41, e120. https://doi.org/10.1017/S0140525X17001972

A complete reference list in APA format.