Comparing Methods for Analyzing Music-Evoked Autobiographical Memories

Amy M. Belfi, Elena Bai, & Ava Stroud
Missouri University of Science and Technology

THE STUDY OF MUSIC-EVOKED AUTOBIOGRAPHICAL memories (MEAMs) has grown substantially in recent years. Prior work has used various methods to compare MEAMs to memories evoked by other cues (e.g., images, words). Here, we sought to identify which methods could distinguish between MEAMs and picture-evoked memories. Participants ($N = 18$) listened to popular music and viewed pictures of famous persons, and described any autobiographical memories evoked by the stimuli. Memories were scored using the Autobiographical Interview (AI; Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002), Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015), and Evaluative Lexicon (EL; Rocklage & Fazio, 2018). We trained three logistic regression models (one for each scoring method) to differentiate between memories evoked by music and faces. Models trained on LIWC and AI data exhibited significantly above chance accuracy when classifying whether a memory was evoked by a face or a song. The EL, which focuses on the affective nature of a text, failed to predict whether memories were evoked by music or faces. This demonstrates that various memory scoring techniques provide complementary information about cued autobiographical memories, and suggests that MEAMs differ from memories evoked by pictures in some aspects (e.g., perceptual and episodic content) but not others (e.g., emotional content).

ONE OF THE MOST SALIENT EFFECTS OF MUSIC is its ability to evoke rich, vivid, and emotional autobiographical memories. For example, hearing the song played during the first dance at your wedding may take you back to that moment, recalling the sights, sounds, and feelings you experienced during the original event. True to this experience, research indicates that music is an effective cue for autobiographical memory retrieval, such that autobiographical memories evoked by music tend to be highly emotional and vivid (Belfi, Karlan, & Tranel, 2016, 2018; Ford, Addis, & Giovanello, 2011; Janata, 2009; Janata, Tomic, & Rakowski, 2007; Sheldon & Donahue, 2017). Early work on this topic evaluated the characteristics of music-evoked autobiographical memories (MEAMs) in isolation (Cuddy, Sikka, Silveira, Bai, & Vanstone, 2017a; Janata, 2009; Janata et al., 2007), or compared MEAMs to autobiographical memories recalled during silence (El Haj, Fasotti, & Allain, 2012; El Haj, Postal, & Allain, 2012). While this provided foundational support indicating that music is an effective autobiographical memory cue, it did not identify whether music evokes autobiographical memories that are different from memories evoked by other sensory cues.

More recently, investigators have begun to explore this question by comparing MEAMs to autobiographical memories evoked by other cues. In our prior work, we investigated differences between autobiographical memories evoked by pictures of famous people and MEAMs evoked by pop-songs from the Billboard Hot 100 year-end charts (Belfi et al., 2016, 2018). Autobiographical memories evoked by faces contained more semantic content (i.e., general facts about the world or oneself), whereas MEAMs contained a greater proportion of episodic content (i.e., details about the time and place, emotions, and other aspects of an event). Similarly, other work has compared MEAMs evoked by famous songs (e.g., The Beatles, "Hey Jude") to memories evoked by pictures of famous world events (e.g., the assassination of John F. Kennedy; Baird, Brancatisano, Gelding, & Thompson, 2018). In this case, patients with Alzheimer's disease reported significantly fewer memories evoked by pictures than healthy comparison participants, but their frequency of MEAMs did not differ. Other work indicates that MEAMs evoked by Billboard songs contain more motor-perceptual details than memories evoked by verbal cues of a lifetime period (i.e., "10 years old") or a specific event (e.g., "Olympics, Sydney, Australia"; Zator & Katz, 2017). Taken together, this work suggests that autobiographical memories may differ based on how they were cued.

To specifically address this question, one prior study sought to investigate differences between autobiographical memories reported while listening to music, reading lyrics, or viewing an image of a musical artist (Cady, Harris, & Knappenberger, 2007). This work found no difference in the self-reported emotional content or vividness of the memories based on the cue. However, in this case, participants pre-selected the cue that was most likely to trigger a memory, which may have influenced the results. Other work found that autobiographical memories for musical events (e.g., attending a concert) did not differ from memories of other lifetime events, such as holidays (Halpern, Talarico, Gouda, & Williamson, 2018). Therefore, while some prior work indicates differences between memory cues, other work does not. This seeming conflict in results may be due to 1) the cue used as a comparison (e.g., images of faces, images of events, words), and 2) the techniques used to score the autobiographical memory data.

AUTOBIOGRAPHICAL MEMORY SCORING METHODS

One approach to scoring autobiographical memory data requires manual coding. Our prior work (Belfi et al., 2016, 2018) used the Autobiographical Interview coding protocol (AI; Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002), which characterizes the episodic nature of autobiographical memories. In this coding scheme, the researcher identifies the types of specific details (i.e., the episodic and semantic content) in the memories. Similarly, other work comparing MEAMs to picture-evoked memories has used a manual coding scheme based on the TEMPau (Test Episodique de Mémoire du Passé autobiographique; Piolino, Desgranges, & Eustache, 2009), which involves classifying memories based on their level of specificity (Baird et al., 2018). Such manual coding methods typically require substantial training of independent raters, in addition to the time required for the raters to read and score each memory.

Other work on MEAMs has used automated text analysis software (Cuddy et al., 2017; Janata et al., 2007; Zator & Katz, 2017). There are many such tools for text analysis, one of the most prominent being the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Boyd, Jordan, & Blackburn, 2015; Tausczik & Pennebaker, 2010). The LIWC analyzes text data using a dictionary method to classify the words in a text. Words can be categorized based on their function, such as pronouns, articles, or prepositions; as well as by their content (e.g., positive or negative emotional words). A similar tool to the LIWC is the Evaluative Lexicon (EL), which was recently developed to analyze texts with a particular focus on the affective content of evaluative language (Rocklage & Fazio, 2015; Rocklage, Rucker, & Nordgren, 2017).

One benefit of using tools like the LIWC and EL when studying MEAMs is that such programs are automated and can quickly generate metrics to quantify various components of a text. Compared to manual coding methods like the AI, which requires hours of training and coding, automated analyses confer a substantial benefit in terms of time and effort. However, manual coding methods may allow for more specific and nuanced analysis, and may uncover insights about a text that cannot be inferred using an automated system. Dictionary-based text analysis methods, like the LIWC and EL, only identify single words; for example, they would treat the sentences "I have never been happy" and "I have never been this happy" as relatively similar, since they both include the word "happy" (Iliev, Dehghani, & Sagi, 2015). Another key strength of the AI is that it was purposefully designed for autobiographical memory analysis. In contrast, the LIWC can be used for any textual data, while the EL was specifically designed for evaluative language (i.e., reviews, judgments, evaluations). In sum, while there may be a benefit to using automated software for studying MEAMs, this benefit might be outweighed by a possible cost in terms of the specificity of detail provided by these methods.

THE PRESENT STUDY

As prior work has suggested that MEAMs may differ from autobiographical memories evoked by other cues, the present study aimed to answer the following question: Based on a memory description alone, is it possible to accurately identify which cue evoked the memory? In answering this question, we sought to compare different scoring methods, with the goal of identifying which methods could successfully differentiate between music- and face-evoked memories. We compared three of such techniques: one manual coding method (the Autobiographical Interview, AI), and two automated methods: the Linguistic Inquiry and Word Count (LIWC) and Evaluative Lexicon (EL). While there are many other possible scoring methods (Gardner, Vogel, Mainetti, & Ascoli, 2012; Iliev et al., 2015; Mehl, 2007), we chose these three for the following reasons: We chose the AI as our manual scoring method because it has been frequently used in studies of autobiographical memory, including prior work on MEAMs (Belfi et al., 2016, 2018). For our automated methods, we chose the LIWC because it has also been previously used to characterize differences between MEAMs and memories evoked by other cues (Zator & Katz, 2017). While the EL was not

designed for autobiographical memory analysis, we used this method because of its focus on the affective content of a text. Prior research on MEAMs has indicated that they often contain emotional content (Janata et al., 2007), and the EL provides more nuanced measures of emotion than the LIWC.

We analyzed autobiographical memories evoked by music and pictures of famous faces with each of these three methods, and then used the resulting output to predict whether the memory was evoked by a face or a song. To do this, we trained a ridge logistic regression classifier to distinguish between memories evoked by music and faces. We hypothesized that all three scoring techniques would successfully differentiate between music- and picture-evoked memories. In addition to calculating the accuracy of each method, we also sought to characterize which features were most important in differentiating between music- and picture-evoked autobiographical memories. Based on previous research, we hypothesized that MEAMs would contain a greater proportion of episodic content and a greater number of emotional and perceptual words, while face-evoked memories would contain a greater amount of semantic content and fewer emotional and perceptual words.

Method

PARTICIPANTS

Participants were healthy adults ($N = 18$; 12 male, 6 female) aged 37–73 years old ($M = 57.2$, $SD = 12.3$), collected as a part of a larger study (Belfi et al., 2016, 2018). Participants were recruited through advertisements in the local community and a registry of healthy research participants. Inclusion criteria for the present study required that all raw data (memory transcriptions) were available for each participant. Participants in the present study had an average of 16.68 years of education ($SD = 1.37$) and an average Full-Scale IQ of 113.52 ($SD = 8.24$). Full-Scale IQ was approximated using the Wechsler Test of Adult Reading (Wechsler, 2001). This study was approved by the Institutional Review Board and all participants gave informed consent in accordance with the requirements of the Human Subjects Committee.

MATERIALS

Stimuli were pictures of famous faces and popular songs (Belfi et al., 2016, 2018). Songs were chosen from the Billboard Hot 100 year-end charts: First, a song database was created with the top 20 songs from each year between 1950–2012. Songs were randomly selected from this database for each participant based on their

age. That is, stimuli were selected to fall within the "reminiscence bump" age period for each participant, between the ages of 15 and 30 (Rubin & Schulkind, 1997). For example, a participant born in 1970 would hear songs that were on the Billboard charts between 1985–2000. The lower bound of this range was selected to roughly correspond to the age at which individuals develop musical preferences (Holbrook & Schindler, 1989; North & Hargreaves, 1999). Also, individuals report high familiarity for songs that were popular during their youth (Schulkind, Hennis, & Rubin, 1999). Additionally, prior work has indicated that participants tend to form their preferences for media (i.e., their favorite books, movies, and albums) in their mid-20s (Janssen, Chessa, & Murre, 2007). The goal was to choose stimuli that were highly likely to be familiar and evoke autobiographical memories for each participant. Each song clip was 15 seconds long and corresponded to the chorus or other highly recognizable parts of the song.

Faces were chosen from the Iowa Famous Faces test (Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996). The faces from this test include individuals of varying occupations, including athletes, politicians, and actors. In a sample of 90 healthy adults, these faces were correctly named 85% of the time, indicating high familiarity (Tranel, 2006). Each famous face was assigned to the years in which they were most popular (for example, when they were actively playing professional sports or holding political office). Faces were selected in the same way as songs: Participants were randomly presented with faces that fell during the reminiscence bump years. Each face was presented on the screen for 5 seconds.

PROCEDURE

After completing informed consent, participants began the experiment. Participants were seated in front of a computer with an experimenter present. The experimenter conducted the task using the MATLAB Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Participants listened to thirty songs and viewed thirty pictures of faces in a counterbalanced design. After each stimulus, if the stimulus evoked an autobiographical memory, participants verbally described this memory in as much detail as possible to the experimenter. Participants were given as much time as necessary to describe their memory. After each memory description, the experimenter provided a general probe (Levine et al., 2002). This general probe served to provide the participant with additional time, if needed, to think of more details to add to their memory description. As in Levine et al. (2002), data from the initial retrieval and from the

general probe were combined. To preserve the involuntary nature of these memories, participants were not given specific probes. All memory descriptions were audio recorded. After completing the task, participants were debriefed on the purpose and goals of the experiment.

DATA QUANTIFICATION

Autobiographical Interview. Recordings of memory descriptions were transcribed and coded (as in Levine et al., 2002). Each memory was segmented into details (single pieces of information) that were coded as either internal or external. Internal details pertain to the central memory and reflect episodic reexperiencing. External details do not directly pertain to the memory and primarily reflect semantic content. Memories were coded by three trained raters and each memory was coded by only one rater. Memories from two pilot participants were coded by all three raters and intraclass correlation was performed (using a two-way mixed model) to assess interrater reliability. The ICC value for internal composite was .93 and for external composite was .88. These values reflect high agreement among the three raters. After coding each detail, internal and external composite scores were created by counting the total number of internal and external details for each condition (music and faces) within each participant. The internal and external composite scores were then used to calculate a ratio of internal/total details. This ratio provides a measure of episodic detail that is unbiased by the total number of details (Levine et al., 2002). We included both the composite internal and external scores, as well as the ratio of internal/total details, in our subsequent statistical analyses.

Linguistic Inquiry and Word Count. The LIWC categorizes words into various groups, ranging from parts of speech to affective dimensions of a text. We selected a subset of these categories based on a priori predictions about which might be the most salient aspects of MEAMs. First, we included word count (WC) to identify whether face- or music-evoked memories contained more words overall. Given that prior research suggests that MEAMs are particularly emotional (Janata et al., 2007), we included the summary variables of "emotional tone" (e.g., overall positive or negative tone) and "authenticity." These summary variables are composites calculated within the LIWC from a combination of features. Texts rated with higher purported "authenticity" are those that can be predicted to be more truthful. That is, the measure of authenticity was developed by comparing texts containing truthful information to texts that contained untruthful information: Texts with higher

authenticity ratings were more cognitively complex, contain more self-references and other-references, and use fewer negative emotional words (Newman, Pennebaker, Berry, & Richards, 2003). We also included the variables of affective processes (e.g., "happy," "cried"), positive emotion (e.g., "love," "nice"), and negative emotion (e.g., "hurt," "ugly"). Since autobiographical memories are related to the self, and MEAMs have been shown to be associated with activity in brain regions important for self-referential processes (Ford et al., 2011; Janata, 2009) we included the variable of personal pronouns (e.g., "I," "me"). We also included the variable of social processes (e.g., "talk," "they"), since such words were found to be particularly prevalent in MEAMs (Janata et al., 2007). Another variable was "certainty" (e.g., "always," "never," "definitely") as we expected participants to be more confident of the content of the MEAMs. Additionally, some of the words included in the "certainty" variable could be reflective of repeated or general autobiographical events (e.g., "We *always* sang this song in choir rehearsals"). Since MEAMs have also been shown to have a greater number of sensory details (Belfi et al., 2016), we included the perceptual variables of seeing (e.g., "view," "saw"), hearing (e.g., "listen," "hearing"), and feeling (e.g., "feels," "touch"). Finally, since autobiographical memory is related to future thinking, we included three time-orientation variables: focus on the past (e.g., "ago," "did"), focus on the present (e.g., "today," "now"), and focus on the future (e.g., "may," "will"). In total, we included the following 15 variables in our LIWC analysis: word count, authenticity, emotional tone, personal pronouns, overall affect, positive emotion, negative emotion, social words, certainty, seeing, hearing, feeling, past focus, present focus, future focus.

Evaluative Lexicon. The EL produces substantially fewer variables than the LIWC, since it is specifically focused on the affective nature of adjectives within a text. Given this, we included all output variables from the EL, including an overall word count variable, the overall ratings of valence (i.e., positive or negative), extremity, and emotionality, as well as the total count for positive and negative words. The overall rating of "extremity" represents how extreme the word used is, for example, the word "magnificent" has a higher extremity rating than the word "commendable" (Rocklage & Fazio, 2015). The overall rating of "emotionality" captures the degree of emotionality of a text. For example, the adjective "amazing" has a higher emotionality rating than the adjective "flawless" (Rocklage et al., 2017). Higher emotionality ratings are associated with more "feeling" words (e.g., "I feel," "emotional"), while lower emotionality

ratings are associated with more "cognitive" words (e.g., "I believe," "I think").

ANALYSIS

For each of the three coding schemes (AI, LIWC, EL), we performed two complementary statistical analyses. The first analysis evaluated whether each coding scheme could successfully differentiate between memories evoked by music and faces. That is, to what degree of accuracy can the coding scheme distinguish between whether a memory was evoked by a song or a face? In the first analysis, we trained a ridge logistic regression classifier (using the sklearn package in Python; Pedregosa et al., 2011) to distinguish between memories evoked by music and faces on a subset of the data and evaluated performance on the remaining data. The first analysis consisted of two parts (A and B), which differed in how we divided the data into training and validation (i.e., test) subsets. Analysis 1A used ten-fold cross validation: We first randomly assigned each memory to one of ten subsets. Each subset served as a validation set one time. When a subset was not a validation set it was used for training the model. In analysis 1B, each participant's memories were the validation set in one train-test iteration. When a participant's memories were not part of the validation set, they were part of the training set. For example, the model would train on participants 1 to 17, and test on participant 18. This was repeated for all participants. For analysis 1A and 1B, we conducted a one-sample *t*-test to evaluate whether validation set model performance was better than expected by chance. We assumed a chance of 50% given the two categories (faces vs. music). The second analysis focused on what features helped differentiate memories evoked by faces and music. In this second analysis, we pooled all data into a single training set and trained a logistic regression (using the statsmodels package in Python; Seabold & Perktold, 2010). We then examined the feature weights to understand which features were predictive when differentiating between memories evoked by music and faces.

The first and second analyses both used logistic regression, but the intent of each analysis differed. The first analysis focuses on *whether* we can differentiate between memories evoked by faces and music. To serve this purpose, we evaluated model accuracy on validation sets, which prevents models from performing better due to overfitting. We also used ridge logistic regression, which helps models avoid overfitting and performs better on validation sets. While ridge logistic regression helps prevent overfitting, it also biases all feature weights towards 0, which makes interpreting feature weights more tenuous. We designed the second analysis to focus on *how* the models differentiate between memories evoked by faces and music. In this second analysis we used all our data to train logistic regression models (i.e., no validation set). This opens the possibility that the model has overfit the data (making inference using model accuracy tenuous), but the feature weights become more meaningful. In sum, the goal of these two analyses were to identify: 1) how accurate our coding schemes are at predicting whether memories are evoked by faces or music, and 2) which features of the coding schemes are most important when predicting memory cue type.

Results

AUTOBIOGRAPHICAL INTERVIEW

In analysis 1A (classic ten-fold cross validation), the classifiers averaged 63% accuracy, which is significantly better than chance, $t(9) = 3.32$, $p = .05$. In analysis 1B (each participant-wise validation set) the classifiers averaged 61% accuracy, which is also significantly better than chance, $t(17) = 2.58$, $p = .003$. In order to validate these analyses, we performed each analysis a second time, but with the cue type (faces vs. music) randomly assigned to each memory. For analysis 1A with randomly assigned memory types, accuracy was 45%, which was not different than chance, $t(9) = -1.46$, $p = .17$. For analysis 1B with randomly assigned memory types, accuracy was 55% which was also not different from chance, $t(17) = 0.89$, $p = .38$.

In the final analysis, we found that the number of internal details ($Z = 1.33$, $p = .18$), and the ratio of internal/total details ($Z = 0.16$, $p = .86$), were not predictive of cue type, but that the number of external details was predictive: More external details were associated with face-evoked memories ($Z = -4.10$, $p < .001$). See Figure 1 for a graphical depiction of these results.

LINGUISTIC INQUIRY AND WORD COUNT

In analysis 1A (classic ten-fold cross validation), the classifiers averaged 86% accuracy on validation sets which is significantly better than chance, $t(9) = 12.33$, $p < .001$. In analysis 1B (each participant-wise validation sets), the classifiers averaged 83% accuracy on validation sets which is also significantly better than chance, $t(17) = 9.56$, $p < .001$. In order to validate these analyses, we performed each analysis a second time, but with the cue type (faces vs. music) randomly assigned to each memory. In analysis 1A with randomly assigned memory types, validation accuracy was 55% which was not different than chance, $t(9) = 1.42$, $p = .19$. In analysis
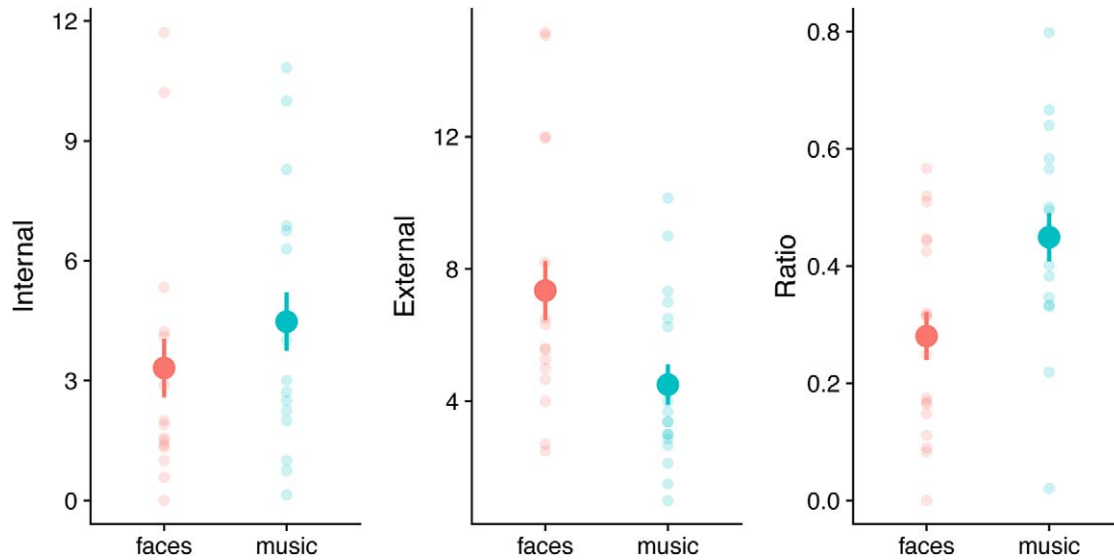
FIGURE 1. Autobiographical Interview (AI) data. From left to right: Internal depicts the average number of internal details per memory, external depicts the average number of external details per memory, and ratio depicts the average ratio of internal/total details per memory. Individual subjects' data is plotted in transparent circles; average data is plotted in opaque circles; error bars depict standard error of the mean.

1B with randomly assigned memory types, validation accuracy was 55% which was also not different from chance, $t(17) = 1.28$, $p = .22$.

For analysis 2, we found that memories evoked by music had greater authenticity ($Z = 4.06$, $p < .001$) and a larger number of auditory perceptual details ($Z = 7.64$, $p < .001$) and physical perceptual details (e.g., "feel", $Z = 2.31$, $p = .02$). Memories evoked by faces contained a greater number of visual perceptual details ($Z = -3.66$, $p < .001$). There were no significant differences between MEAMs and face-evoked memories for the following variables: word count (faces $M = 107.76$, $SD = 63.76$; music $M = 82.68$, $SD = 56.95$; $Z = -0.26$, $p = .78$), emotional tone ($Z = 0.06$, $p = .94$), overall affect ($Z = 0.12$, $p = .89$), positive ($Z = -0.14$, $p = .88$) and negative affect ($Z = -0.49$, $p = .61$), personal pronoun use ($Z = -1.40$, $p = .16$), social words ($Z = 1.02$, $p = .30$), certainty ($Z = -0.62$, $p = .52$), or time focus [past ($Z = -0.26$, $p = .78$), present ($Z = 0.36$, $p = .71$), or future ($Z = -0.85$, $p = .31$)]. See Figure 2 for a graphical depiction of these results.

EVALUATIVE LEXICON
In analysis 1A (classic ten-fold cross validation), the classifiers averaged 54% accuracy on validation sets which is not significantly better than chance, $t(9) = 1.45$, $p = .18$. In analysis 1B (each participant-wise validation sets), the classifiers averaged 52% accuracy on validation sets which is also not significantly better than

chance, $t(17) = 0.81$, $p = .40$. In order to validate these analyses, we performed each analysis a second time, but with the cue type (faces vs. music) randomly assigned to each memory. In analysis 1A with randomly assigned memory types, validation accuracy was 45% which was not different than chance, $t(9) = -2.10$, $p = .06$. In analysis 1B with randomly assigned memory types, validation accuracy was 55% which was also not different from chance, $t(17) = 1.41$, $p > .17$. We were unable to do analysis 2 for this dataset due to high multicollinearity between the variables. Note that while this does not impact model performance, it does prevent reliable feature weight estimation. See Figure 3 for a graphical depiction of these results.

NUMBER OF MEMORIES
While our main analysis of interest was the content of the memories, we lastly sought to investigate differences in the number of memories evoked by faces and music. A paired-samples *t*-test indicated that faces evoked significantly more memories ($M = 11.17$, $SD = 6.99$) than music ($M = 7.06$, $SD = 3.78$; $t(17) = -2.58$, $p = .01$, 95% CI: -7.46, -0.75).

Discussion

The present study had two goals: First, we sought to evaluate three text scoring methods and to identify which could successfully differentiate between autobiographical
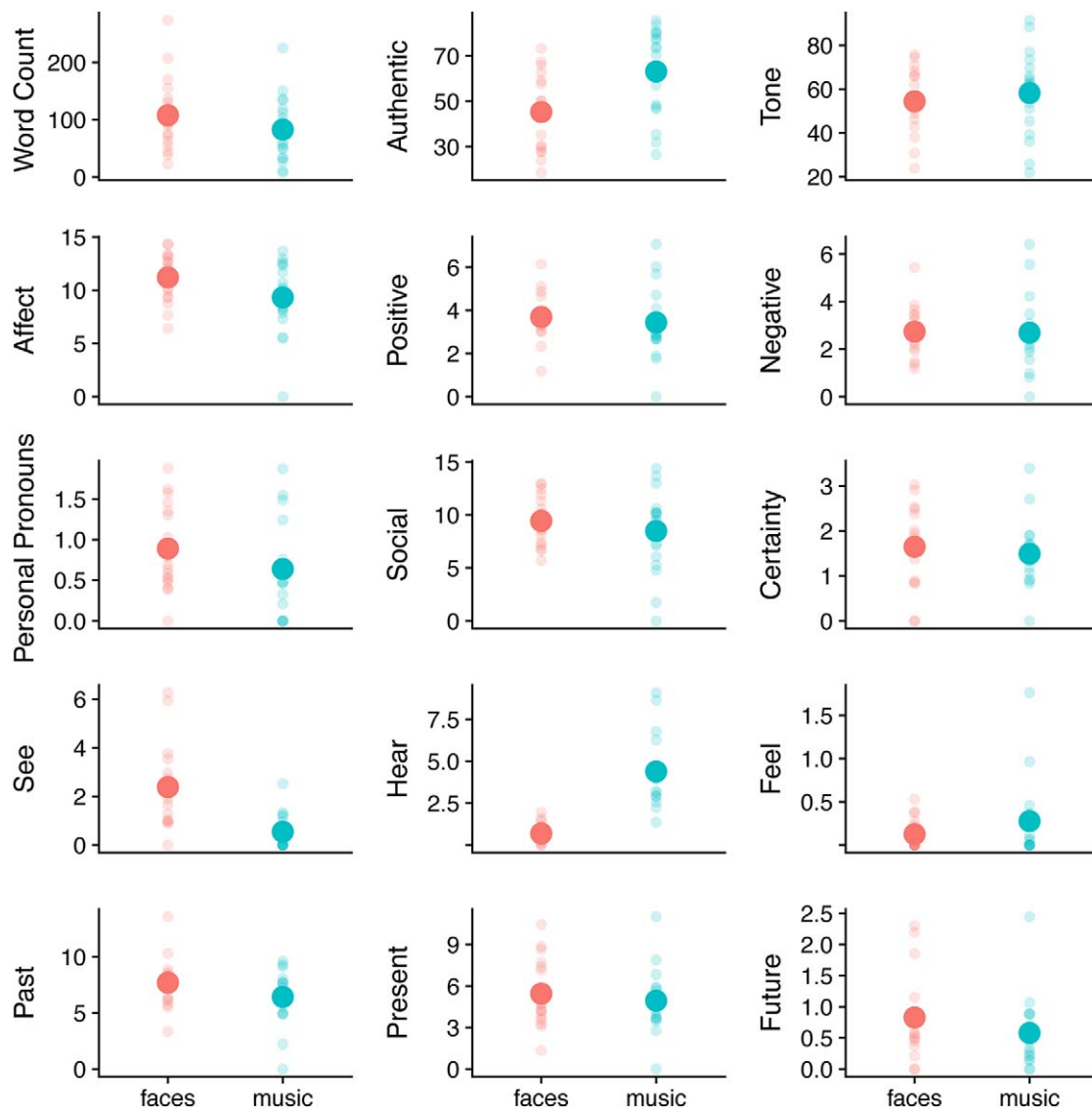
FIGURE 2. Linguistic Inquiry and Word Count (LIWC) data. Word Count depicts the average total number of words per memory; Authentic and tone depict the percentiles on these variables as determined by the LIWC; all other variables depict the percentage of total words in each category per memory. Individual subjects' data is plotted in transparent circles; average data is plotted in opaque circles; error bars depict standard error of the mean.

memories evoked by faces and music. Second, we aimed to identify which specific features of these coding schemes were most influential in distinguishing between memories evoked by the two cue types. With regard to the first goal, our results indicate that both the AI and LIWC are able to successfully distinguish between MEAMs and face-evoked memories with above-chance accuracy, while the EL is not. While not directly comparing the accuracy between models, the models using the LIWC data had greater accuracy than those trained on the AI data, suggesting that the

LIWC variables are better able to predict whether a memory was evoked by music or a face. When considering this difference in accuracy between the two coding types, it is important to note that the cross-validation models used here control for the number of features, as using validation data prevents coding schemes with more features from necessarily performing better. Despite the fact that the AI only used three features while the LIWC used 15, this should not have an effect on the validation accuracy. Therefore, the greater validation accuracy of the LIWC is due to the
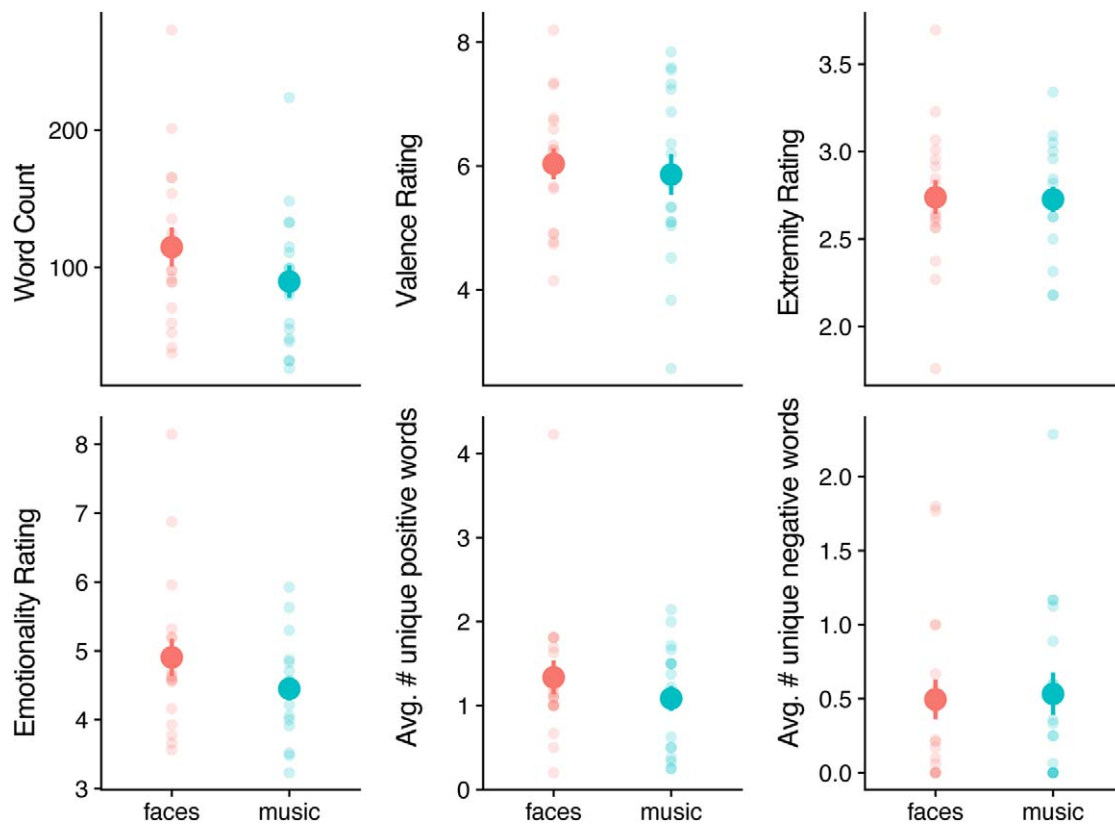
FIGURE 3. Evaluative Lexicon (EL) data. Word count is the average number of words per memory; valence, extremity, and emotional ratings are the average ratings on each variable as denoted by the EL software, and the average word counts for positive and negative emotional words. Individual subjects' data is plotted in transparent circles; average data is plotted in opaque circles; error bars depict standard error of the mean.

relative importance of the features, rather than the total number of features.

While both the AI and LIWC were successfully able to differentiate between MEAMs and face-evoked memories above chance, they provided complementary information as to what distinguishes these two memory types. Replicating our prior work using an updated statistical analysis (Belfi et al., 2016, 2018), face-evoked memories contained a significantly greater number of external details than MEAMs. External details are those that are not related to the episode itself, and are frequently semantic statements. This finding that MEAMs are less semantic than face-evoked memories may seem inconsistent with other recent work, which found that MEAMs contain *more* semantic content than picture-evoked memories (Baird et al., 2018). This contradiction may be due to the nature of the pictures used in these two studies. Here, our photos consisted of pictures of famous persons, while other work used photos of famous events. It may be the case that music serves as a contextual cue for autobiographical memory retrieval.

For example, episodic memories encoded during music listening show a context-dependent memory effect, such that recall is improved when listening to the same music at retrieval as at encoding (Balch, Bowman, & Mohler, 1992; Balch & Lewis, 1994). This effect may explain why MEAMs contain a greater proportion of episodic content than face-evoked memories. It also may explain why other work comparing MEAMs to memories evoked from pictures of *events* did not find the differences seen here. Instead, it may be the case that images of events, rather than persons, serve as a similar contextual cue as music. Future research could explore in detail the possible nature of music as a contextual cue, perhaps by comparing MEAMs to autobiographical memories evoked by other commonly occurring environmental sounds, or pictures of frequently-visited locations.

One perhaps surprising result of the present work was that the EL failed to distinguish between MEAMs and face-evoked memories. The EL was chosen because it provides a more detailed description of emotional

content than the LIWC (Rocklage et al., 2017). Despite this added information about the emotional content, the models trained on the EL data failed to distinguish between MEAMs and memories evoked by faces, suggesting no difference in these emotional characteristics. This is also consistent with our findings from the LIWC data: although the LIWC model successfully distinguished between MEAMs and face-evoked memories, the affective categories were not significant predictors in our models. Therefore, the findings from both the LIWC and EL converge to suggest little difference in the affective nature of MEAMs and face-evoked memories.

This lack of apparent difference in the emotional content of MEAMs and face-evoked memories may be surprising, given the fact that prior research has suggested that MEAMs tend to be quite emotional, based on subjective ratings of felt emotions (Cuddy, Sikka, Silveira, Bai, & Vanstone, 2017b; Janata et al., 2007). However, the present findings are consistent with work investigating MEAMs using the LIWC, which found that the frequency of affective words in MEAMs was quite low, similar to the present results (e.g., 1-5%; Cuddy et al., 2017; Janata et al., 2007). Recent work comparing MEAMs to memories evoked using lifetime period cues found that MEAMs contained fewer negative emotional words, suggesting that MEAMs may be *less* emotional on this particular dimension (as identified by the LIWC; Zator & Katz, 2017). Again, these results suggest that MEAMs may differ in emotional content from memories evoked by other cues, but it likely depends on the cue used as a comparison. It may be the case that the *experience* of recalling MEAMs evokes more, or stronger, emotions in the individual, but that the actual content of the memories is not more emotional. Future work could compare subjective ratings of emotions between MEAMs and memories evoked by other cues, to better clarify whether the experience of recalling MEAMs is particularly emotional.

While there were no clear differences in the emotional content of MEAMs and face-evoked memories, there were differences in the perceptual content: MEAMs contained a greater number of auditory and physical perceptual words than face-evoked memories. This suggests that memory cue modality has a strong influence over the experience of recalling the memory. An interesting venue for future research may be to directly compare memories evoked by a song to memories evoked by an image of the particular musical artist (here, we did not include images of musical artists in the famous faces stimulus set). One prior study sought to address this question by presenting participants with either a song,

the written lyrics to the song, or an image of the musical artist (Cady et al., 2007). However, one key difference from the present work is that Cady et al. (2007) used the song *name* as the initial memory cue; that is, participants were asked to read a list of song titles and specifically to choose a song that evoked the strongest autobiographical memory. The additional stimulus (either image, lyrics, or song) served as an "elaborative" cue during which participants were asked to describe their memory. It is therefore unclear whether memories evoked spontaneously between an image of a musical artist, versus the music itself, would differ in their vividness or specificity.

One limitation of the present study is the relatively small sample size, which limits our ability to investigate influences of individual differences on the characteristics of MEAMs. Prior work has indicated that autobiographical memories change across the lifespan, such that older adults tend to produce memories that are more semantic in nature and less temporally specific (Levine et al., 2002; Piolino et al., 2010). While the present study included participants across a wide age range of middle-age to older adults, we also did not include adults in the typical "younger" age range (typically 30–35 years of age or younger), so we are unable to fully explore the possibility of age-related differences in MEAMs here.

Overall, our results here suggest that both the AI and LIWC are successful at differentiating MEAMs from memories evoked by images of famous persons. These effects suggest that MEAMs contain fewer semantic details, fewer visual details, and more auditory details, physical details, and greater authenticity than memories evoked by images of faces. These results suggest that both manual and automated coding methods are useful for work investigating differences between MEAMs and memories evoked by other sensory cues, and can provide complementary information regarding such differences.

Author Note

Correspondence concerning this article should be addressed to Amy M. Belfi, Department of Psychological Science, 135 H-SS, 500 W. 14th Street, Rolla, MO 65409. E-mail: amybelfi@mst.edu

References

BAIRD, A., BRANCATISANO, O., GELDING, R., & THOMPSON, W. F. (2018). Characterization of music and photograph evoked autobiographical memories in people with Alzheimer's disease. *Journal of Alzheimer's Disease, 66,* 693–706.

BALCH, W. R., BOWMAN, K., & MOHLER, L. A. (1992). Music-dependent memory in immediate and delayed word recall. *Memory and Cognition, 20*(1), 21-28.

BALCH, W. R., & LEWIS, B. S. (1994). Music-dependent memory: The roles of tempo change and mood mediation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1354–1363.

BELFI, A. M., KARLAN, B., & TRANEL, D. (2016). Music evokes vivid autobiographical memories. *Memory, 24,* 979–989.

BELFI, A. M., KARLAN, B., & TRANEL, D. (2018). Damage to the medial prefrontal cortex impairs music-evoked autobiographical memories. *Psychomusicology: Music, Mind, and Brain, 28,* 201-208.

BRAINARD, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

CADY, E. T., HARRIS, R. J., & KNAPPENBERGER, J. B. (2007). Using music to cue autobiographical memories of different lifetime periods. *Psychology of Music, 36,* 157–177.

CUDDY, L. L., SIKKA, R., SILVEIRA, K., BAI, S., & VANSTONE, A. (2017). Music-evoked autobiographical memories (MEAMs) in Alzheimer disease: Evidence for a positivity effect. *Cogent Psychology, 4,* 1–20.

DAMASIO, H., GRABOWSKI, T. J., TRANEL, D., HICHWA, R. D., & DAMASIO, A. R. (1996). A neural basis for lexical retrieval. *Nature, 380,* 499-505.

EL HAJ, M., FASOTTI, L., & ALLAIN, P. (2012). The involuntary nature of music-evoked autobiographical memories in Alzheimer's disease. *Consciousness and Cognition, 21,* 238–246.

EL HAJ, M., POSTAL, V., & ALLAIN, P. (2012). Music enhances autobiographical memory in mild Alzheimer's disease. *Educational Gerontology, 38,* 30–41.

FORD, J. H., ADDIS, D. R., & GIOVANELLO, K. S. (2011). Differential neural activity during search of specific and general autobiographical memories elicited by musical cues. *Neuropsychologia, 49,* 2514–2526.

GARDNER, R. S., VOGEL, A. T., MAINETTI, M., & ASCOLI, G. A. (2012). Quantitative measurements of autobiographical memory content. *PLoS ONE, 7*(9).

HALPERN, A. R., TALARICO, J. M., GOUDA, N., & WILLIAMSON, V. J. (2018). Are musical autobiographical memories special? It ain't necessarily so. *Music Perception, 35,* 561–572.

HOLBROOK, M. B., & SCHINDLER, R. M. (1989). Exploratory findings on the development of musical tastes. *Journal of Consumer Research, 16,* 119–124.

ILIEV, R., DEHGHANI, M., & SAGI, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition, 7,* 265–290.

JANATA, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex, 19,* 2579–2594.

JANATA, P., TOMIC, S. T., & RAKOWSKI, S. K. (2007). Characterization of music-evoked autobiographical memories. *Memory, 15,* 845–860.

JANSSEN, S. M. J., CHESSA, A. G., & MURRE, J. M. J. (2007). Temporal distribution of favourite books, movies, and records: Differential encoding and re-sampling. *Memory, 15,* 755–767.

LEVINE, B., SVOBODA, E., HAY, J. F., WINOCUR, G., & MOSCOVITCH, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging, 17,* 677–689.

MEHL M.R. (2007). Quantitative text analysis. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 141–156). Washington, DC: American Psychological Association.

NEWMAN, M. L., PENNEBAKER, J. W., BERRY, D. S., & RICHARDS, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29,* 665–675.

NORTH, A. C., & HARGREAVES, D. J. (1999). Music and adolescent identity. *Music Education Research, 1,* 75–92.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., ET AL. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12,* 2825–2830.

PELLI, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437-442.

PENNEBAKER J. W., BOYD, R. L., JORDAN, K., & BLACKBURN, K. (2015). *The development and psychometric properties of LIWC2015.* Austin, TX: University of Texas at Austin.

PIOLINO, P., COSTE, C., MARTINELLI, P., MACÉ, A.-L., QUINETTE, P., GUILLERY-GIRARD, B., & BELLEVILLE, S. (2010). Reduced specificity of autobiographical memory and aging: Do the executive and feature binding functions of working memory have a role? *Neuropsychologia, 48,* 429–40.

PIOLINO, P., DESGRANGES, B., & EUSTACHE, F. (2009). Episodic autobiographical memories over the course of time: Cognitive, neuropsychological and neuroimaging findings. *Neuropsychologia, 47,* 2314-2329.

ROCKLAGE, M. D., & FAZIO, R. H. (2015). The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude valence, extremity, and emotionality. *Journal of Experimental Social Psychology, 56,* 214–227.

ROCKLAGE, M. D., RUCKER, D. D., & NORDGREN, L. F. (2018). The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior Research Methods, 50,* 1327-1344.

RUBIN, D. C., & SCHULKIND, M. D. (1997). Distribution of important and word-cued autobiographical memories in 20-, 35-, and 70-year-old adults. *Psychology and Aging, 12,* 524-535.

SCHULKIND, M. D., HENNIS, L. K., & RUBIN, D. C. (1999). Music, emotion, and autobiographical memory: They're playing your song. *Memory and Cognition, 27,* 948–955.

SEABOLD, S., & PERKTOLD, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference* (pp. 57–61). Austin, TX: SciPy 2010.

SHELDON, S., & DONAHUE, J. (2017). More than a feeling: Characterizing the impact of emotional cues on the access and experience of autobiographical memories. *Memory and Cognition, 45,* 731–744.

TAUSCZIK, Y., & PENNEBAKER, J. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 1,* 24-54.

TRANEL, D. (2006). Impaired naming of unique landmarks is associated with left temporal polar damage. *Neuropsychology, 20,* 1–10.

WECHSLER, D. (2001). *Wechsler Test of Adult Reading.* San Antonio, TX: Psychological Corporation.

ZATOR, K., & KATZ, A. N. (2017). The language used in describing autobiographical memories prompted by life period visually presented verbal cues, event-specific visually presented verbal cues and short musical clips of popular music. *Memory, 25,* 831–844.