Music Evokes Fewer but More Positive Autobiographical Memories Than Emotionally Matched Sound and Word Cues

Kelly Jakubowski & Tuomas Eerola

Department of Music, Durham University, UK

Author Note

This work was supported by a Leverhulme Trust Early Career Fellowship (ECF2018-209) awarded to the first author. Ethical approval was granted by the Durham University Music Department Ethics Committee. We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Kelly Jakubowski, Department of Music, Durham University, Palace Green, Durham, DH1 3RL, United Kingdom (Email: kelly.jakubowski@durham.ac.uk). https://orcid.org/0000-0002-4954-7117

Abstract

Anecdotal propositions that music is “special” as a memory cue have been partially supported by research demonstrating that music can evoke qualitatively different autobiographical memories than various other cues. However, it is unknown whether such differences in memory qualities may be attributed to inherent differences in properties of the retrieval cues. Across four online experiments, we tested whether unfamiliar musical retrieval cues exhibit differential effects on autobiographical memories when compared against matched emotional cues (environmental sounds and words). The music consistently elicited fewer memories in comparison to sound and word cues. Music also evoked relatively positive memories regardless of its valence, with negative-sounding music reliably bringing back more positive memories than negative sounds/words. Words elicited memories rated as more unique than music-cued memories. These results have implications for the use of music as a memory cue in both research and practical contexts, from music therapy to advertising.

Keywords: autobiographical memory, music-evoked autobiographical memory, emotion, retrieval cues

General Audience Summary

It is commonly believed that music is a particularly powerful stimulus for bringing back memories from one's past. However, explanations as to why music may be "special" in this regard are lacking. One notable feature of music is its ability to convey a range of emotions, and it is well established that emotional features of memory retrieval cues can impact on both the emotions and other qualities of autobiographical memories evoked by these cues. As such, we tested how music compares to other, equivalently emotional stimuli as an autobiographical memory cue. In two experiments unfamiliar instrumental music was compared against another auditory stimulus (environmental sounds), and in two subsequent experiments it was compared against a visual stimulus (single words). After matching the music to the sounds/words on their emotional valence and arousal, these stimuli were presented to participants as autobiographical memory cues. The music consistently evoked fewer memories than sound/word cues, and music-cued memories were retrieved more slowly in three of the experiments. However, music evoked relatively positive memories regardless of its emotional valence, with negative-sounding music reliably eliciting more positive memories than negative sounds/words. Music-evoked memories were phenomenologically similar to sound-evoked memories, but words elicited memories of events rated as more unique than music-cued memories. These findings indicate that unfamiliar music is not necessarily the ideal retrieval cue in situations where the aim is to elicit many memories, or bring them to mind quickly. However, music may be a particularly effective medium for evoking positive lifetime memories, regardless of the emotional valence of the music itself. These results have implications for the way in which music is used in commercial contexts, such as shops and advertisements, as well as in clinical settings where music may be utilized as a tool for positive reminiscence.

Music Evokes Fewer but More Positive Autobiographical Memories Than Emotionally Matched Sound and Word Cues

"It is extraordinary how music sends one back into memories of the past." This quote from author Amantine Lucile Aurore Dupin (known as George Sand) exemplifies the widespread belief that music is somehow "special" in its ability to conjure up memories and associated feelings from our lives. This idea has been partially supported by research showing that music can evoke qualitatively different autobiographical memories than other types of retrieval cues in both healthy and clinical populations (Baird et al., 2018, 2020; Belfi et al., 2016, 2020; El Haj et al., 2012; Jakubowski, et al., 2021; Zator & Katz, 2017). The present work aims to critically examine the extent to which these differences between music-evoked autobiographical memories (MEAMs) and other autobiographical memories emerge in healthy adults, specifically by taking into account emotional features of the retrieval cues as a potential factor that may drive differences in memory accessibility and content. Such comparisons are essential for furthering our understanding as to *why* music might be a salient cue for memories that are particularly vivid, emotional, and valued.

Comparing Autobiographical Memory Cues

Music has previously been compared against several different retrieval cue types. Popular music was found to evoke fewer but more episodically vivid memories than photographs of famous faces (Belfi et al., 2016), MEAM descriptions comprised more motion-related and spatial terms than memories cued by words referencing life periods and famous events (Zator & Katz, 2017), and MEAMs were rated as more vivid, positive, and emotionally intense, and contained more social content than television-evoked memories (Jakubowski et al., 2021). Despite this accumulating evidence that MEAMs differ from other memories on a number of dimensions, there is still limited evidence as to *why* these differences emerge. One possibility is that music tends to be paired with different types of events—for instance, music may accompany events that are highly emotional, social, and embodied more often than various other cue types. Another possibility, which is not mutually exclusive, is that music itself is inherently different from other types of retrieval cues, and these differences between cues can lead to differences in the retrieved memories. This latter proposition serves as the main motivation for the present research.

It is well established that differences in properties of retrieval cues can lead to differences in associated autobiographical memories. For instance, the modality of cue presentation (e.g., visual, auditory, olfactory) has been shown to impact on the quantity, content, and phenomenological characteristics of the memories (Congleton et al., 2020; Herz, 2004; Herz & Schooler, 2002). Research comparing different word cues has shown that more imageable and concrete words typically evoke more memories, which are older, more specific, and retrieved faster, in comparison to less imageable and more abstract words (Robinson, 1976; Rubin & Schulkind, 1997; Uzer, 2016; Uzer et al. , 2012; Williams et al., 1999). A study comparing MEAMs elicited via hearing a song, seeing the lyrics, seeing a picture of the artist/album cover, or seeing the song title (Cady et al., 2008) revealed that the picture condition elicited memories that were less emotional and produced fewer feelings of reliving than some of the other conditions. One potential explanation for this is that pictures of artists and album covers may be associated with multiple songs, thereby eliciting more generic memories than a specific song.

Emotional Features of Autobiographical Memory Cues

To date, research comparing music to other memory retrieval cues has predominantly focused on matching the cues in terms of time period of likely exposure (e.g., comparing music and events from the same year; Zator & Katz, 2017). Music differs from other cue types in several additional ways—perhaps most notably in its capacity to convey emotions (Juslin & Sloboda, 2010). Although several cue types invoked in previous comparisons to music—including famous events, television shows, odors, and paintings (Herz, 1998; Jakubowski et al., 2021; Zator & Katz, 2017)—can also potentially communicate a range of emotions, a primary gap here is that no previous study has attempted to match music and other retrieval cues on their emotional properties. This is an important step in understanding whether music still confers any retrieval advantages over cues that are similarly emotional or, conversely, whether some of the differences between MEAMs and other autobiographical memories observed in previous literature may be at least partially attributed to inherent differences in emotional features of the retrieval cues.

The impact of emotions on autobiographical memory processes is well documented (Holland & Kensinger, 2010). Here, we briefly review previous findings in the context of the two-dimensional, circumplex model of emotions (Posner et al., 2005; Russell, 1980), in which emotions are conceptualized according to their valence (i.e., variations in pleasure/displeasure or positivity/negativity) and arousal (i.e., variations in activation/deactivation). Both the valence and arousal of autobiographical events have been shown to impact retrieval qualities when these events are recalled (Berntsen, 2002; Ford et al., 2012; Talarico et al., 2009; Talarico et al., 2004). More pertinent to the present study is a body of research demonstrating that emotional factors *at retrieval* impact on how autobiographical memories are accessed. One example is the mood-congruence effect, in which it is easier to access memories of a similar emotional tone to one's current mood (Blaney, 1986; Bower, 1981; Singer & Salovey, 1988). Two studies have used unfamiliar music to investigate how the valence and arousal of retrieval cues influence properties of associated autobiographical memories (Schulkind & Woldorf, 2005; Sheldon & Donahue, 2017). Both studies concluded that the valence of the music cues impacted memory valence (more positive memories were recalled in response to positive cues and vice versa), but that cue arousal did not impact memory arousal. In Sheldon and Donahue's (2017) study, memory arousal was measured via a rating scale for “emotional intensity,” which follows various studies in the autobiographical memory literature (e.g., Ford et al., 2012; Talarico et al., 2009; Talarico et al., 2004). However, they also collected ratings of the “energy” of the memories which, interestingly, *did* significantly increase in relation to high arousal cues. Previous emotion studies have indicated that “energy” may be a more suitable measure of arousal than intensity, whilst intensity is better conceptualized as an independent property related to the degree to which particular emotions are experienced (Reisenzein, 1994; Yik et al., 1999). As such, our experiments included measures of both “emotional intensity” and “energy” of memories for consideration in relation to cue arousal.

The Present Research

Across four online experiments, we preselected music and other cues that were matched on the emotions (valence and arousal) they conveyed (see Table 1). In the first two experiments, music was compared to other common environmental sounds (e.g., nature, crowd, mechanical sounds; hereafter referred to as “sounds” for brevity) as an autobiographical memory cue. This allowed us to compare music to another auditory stimulus, taking advantage of their similarly dynamic, temporal nature. Both music and sound cues were chosen to be previously unfamiliar to participants, although both cue types belonged to familiar musical genres/categories of sounds (e.g., “rock music,” “bird sounds”). In two further experiments, we compared autobiographical memories cued by music versus single, well-known words, given the extensive evidence of shared processing resources between music and language, including in affective priming tasks (e.g., Patel, 2008; Steinbeis & Koelsch, 2011).

Table 1. Overview of stimuli used in the four experiments.

Experiment	Musical stimuli	Comparison stimuli
1	Instrumental film music* (Eerola & Vuoskoski, 2011)	Environmental sounds* (SoundEffects+)
2	Instrumental music of various genres* (Aljanaki et al., 2017)	Environmental sounds* (SoundEffects+)
3	Instrumental music of various genres* (Aljanaki et al., 2017)	Well-known, concrete words (Balota et al., 2007)
4	Instrumental music of various genres* (Aljanaki et al., 2017)	Well-known, abstract words (Balota et al., 2007)

* = Selected to be unfamiliar excerpts, but sourced from relatively familiar musical genres and categories of sounds (see details in Method sections of Experiments 1 and 2).

In our experiments, musical excerpts were selected to be unfamiliar, to control for familiarity differences across participants and excerpts. Schulkind and Woldorf (2005) and Sheldon and Donahue (2017) found that unfamiliar music can be an effective cue for autobiographical memories, and previous studies of chart-topping pop music have shown that even unfamiliar songs sometimes elicit autobiographical memories (Janata et al., 2007). Another study revealed that not only unfamiliar music but also unfamiliar odors and unfamiliar visual images (paintings) can elicit autobiographical memories (Herz, 1998).

Our experiments addressed three categories of research questions. The first questions dealt with how **effective** the different cue types were; specifically, we compared the number of memories evoked, and the speed and intentionality with which these were accessed. The second questions related to the **phenomenological characteristics** of the memories; we examined whether ratings of vividness, uniqueness, social content, and importance of the memories differed as a function of cue type. The final questions addressed the **emotional valence and arousal** of the memories; we tested whether the valence and arousal of the cues predicted the valence and arousal of the memories, in an attempt to replicate previous findings, and sought to reveal new evidence on whether different types of emotional cues (music/sounds/words) operate similarly in this regard. Overall, these experiments provide necessary critical insights into the nature of MEAMs as compared against memories evoked via similarly emotional retrieval cues.

Experiment 1

Method

Design

We manipulated cue type (music/sound), cue valence (positive/negative), and cue arousal (high/low). The dependent measures were the number of memories retrieved, retrieval times (i.e., how long it took to retrieve a memory in response to a cue), and ratings of the following: retrieval intentionality (the degree to which the memory came to mind in a voluntary or involuntary manner), valence and arousal of the memories (with arousal measured via both intensity and energy ratings), vividness, uniqueness, socialness, and importance of the memories. Participants also provided a short description of the memory (one sentence) and reported their age during the remembered event. For music stimuli, participants were asked to assess whether they had ever heard the piece of music before participating in this experiment. See Appendix A for full question list.

Participants

Participants were recruited on the basis that they were aged 18 years or older and fluent in English. They were required to confirm that they had no previous history of any of the following: stroke, severe head injury, brain tumor/injury, any other neurological condition that may contribute to cognitive impairment, severe depression or anxiety, alcohol abuse or dependence, or recurrent substance abuse or dependence. In total, 114 participants completed the study, two of whom were excluded due to technical difficulties in sound playback. This left a final sample of 112 participants, aged 18 to 69 years (*M* = 29, *SD* = 12; 54 female and 57 male, 1 chose not to report a gender). Most participants classified themselves as non-musicians (64%) or amateur musicians (29%), with 34% reporting no previous formal training in music and another 30% reporting up to two years of previous musical training. Participants were based in 26 countries of current residence, with most living in the UK (27%), Poland (13%), Portugal (9%), Greece (8%), or the US (5%). The most frequently reported native languages were English (34%), Polish (13%), Portuguese (10%), and Greek (8%). Five participants reported mild hearing loss, three of whom were wearing hearing aids during the study; all others reported normal hearing. Participants were either volunteers (*N* = 42, recruited via social media and university-wide email lists) who were entered in a prize draw for a gift voucher (£25) or panelists from Prolific, an online research participant recruitment tool (https://www.prolific.co; *N* = 70), who were compensated for their participation (£3.75).

Materials/Stimuli

Musical stimuli were selected from the instrumental film music stimulus set compiled by Eerola and Vuoskoski (2011). This set comprises 110 musical excerpts that have been pre-rated on perceived valence and arousal using 9-point scales and have been assessed to be generally unfamiliar to participants in previous research. Nine musical stimuli were used in the present experiment; details of the selection procedure for these are provided below.

Sound stimuli were sourced from a professional sound effects database, SoundEffects+ (https://www.soundeffectsplus.com). We selected an initial set of 39 sounds that we perceived to vary in terms of valence and arousal. We also focused on selecting sounds that were relatively dynamic and complex in nature, often containing multiple sound sources (e.g., a crowd with various talking and traffic noise, a nature scene with birds and insects), to keep the sounds similar to music in these regards. In order to obtain independent ratings of valence and arousal for each sound clip, a pilot study was run in which the 39 sounds were rated on 9-point Likert scales on the emotions they conveyed in terms of valence (1 = negative, 9 = positive) and arousal (1 = low energy, 9 = high energy). Participants in the pilot study were 116 volunteers (ages 18 to 70 years, $M = 35$, $SD = 13$, 73 female), all of whom rated all sounds for both valence and arousal. Mean ratings for all 39 sound stimuli are available on the Open Science Framework (OSF): https://osf.io/dhk8z/. From these initial ratings it became clear that very few stimuli fell conclusively into the negative valence/low arousal category, suggesting that everyday sounds do not convey "sadness" or similar emotions particularly well. As such, in this study we focused on the remaining three valence/arousal quadrants, and selected three sound stimuli from each quadrant that best exemplified the category and were different enough from one another to be distinctive. Examples of the selected sounds include a sound labelled in SoundEffects+ as "beach at night ambience" containing crickets and gentle waves (positive valence/low arousal), a sound labelled "parade crowd ambience" containing cheering and ringing bells (positive valence/high arousal), and a sound labelled "spaceship computers beeping" comprising noisy and rapidly changing electronic sounds (negative valence/high arousal). See Appendix B for labels of all selected sounds.

Two-dimensional distances were computed to find the closest musical stimulus from the full Eerola and Vuoskoski (2011) stimulus set to each of the 9 sound stimuli in terms of both mean valence and mean arousal ratings; one-to-one matching was performed without replacement such that each sound was matched to one unique piece of music. The mean valence and arousal ratings for all 18 selected stimuli are visualized in Figure 1. Overall, there was no significant difference between the two cue types in mean valence ($t(16) = 0.08$, *p* = .94) or mean arousal ratings ($t(16) = -0.01$, $p = .99$) when the selected sound and music cues were compared in independent-samples t-tests.

We presumed both the specific pieces of music and the specific sound effect clips that we selected were unlikely to have been heard before by our research participants, but that both the music and sounds would be similar to music/sounds participants had encountered in their everyday lives (i.e., they would be generally familiar with the genres of music and types of sounds we used). To test this further, we ran a separate experiment with 60 participants (ages 18 to 63 years, $M = 34$, $SD = 11$, 35 female, all native English speakers), who were recruited via the same online platform we used to source the majority of the participants in the main experiment (Prolific). For each music/sound stimulus, participants were asked to rate whether they had ever heard that *specific* piece of music or sound clip before, on a 5-point scale from “1 (Never)” to “5 (Frequently)”. For each stimulus they were also asked to rate the degree to which they had encountered *similar-sounding* music/sounds before, on the same 5-point scale. Half of the participants rated the musical stimuli and half rated the sound stimuli. For ratings of familiarity with a specific piece of music/sound clip, the median rating for 17 of the 18 stimuli was 1. One stimulus (musical excerpt 91) received a median rating of 1.5, although only one participant correctly identified the excerpt as coming from “one of the Alien films.” Furthermore, the music and sound stimuli did not significantly differ in mean ratings of familiarity with the specific excerpts in an independent-samples t-test ($M_{music} = 1.53$, $SD_{music} = 0.29$; $M_{sounds} = 1.55$, $SD_{sounds} = 0.13$; $t(11) = -0.24$, $p = .81$). For ratings of whether participants had encountered similar-sounding music/sounds before, both the music and sounds attained mean ratings near the midpoint of the scale, with ratings again not significantly differing across the two stimulus types ($M_{music} = 2.77$, $SD_{music} = 0.38$; $M_{sounds} = 2.81$, $SD_{sounds} = 0.58$; $t(14) = -0.18$, $p = .86$).

All music and sound stimuli were edited to a 30-second duration for the experiment. This meant the music stimuli, which were all initially around 15 seconds in duration, were typically played twice on a loop.

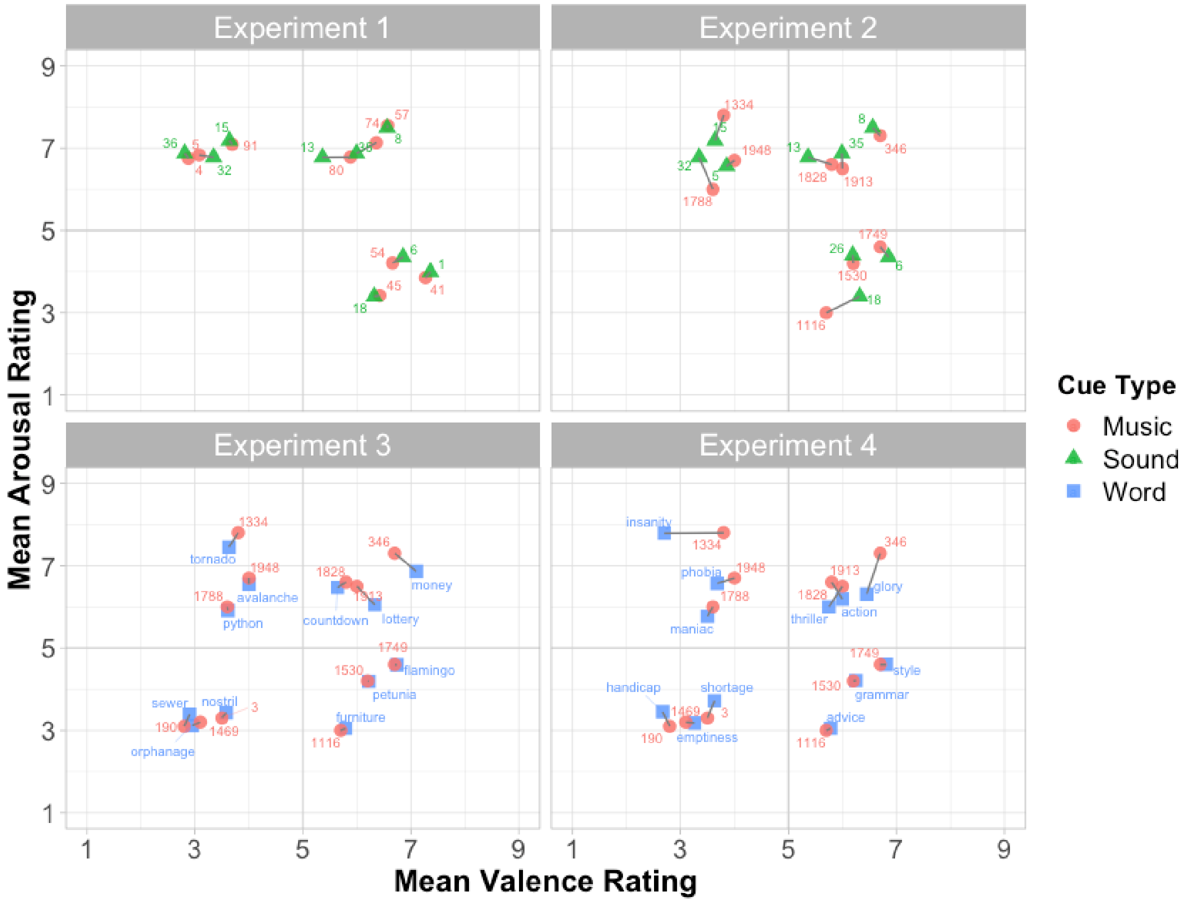

Figure 1. Selected cues for Experiments 1-4. Mean valence and arousal ratings represent pre-ratings of each cue from the pilot study (Experiment 1 & 2 sounds) or existing datasets. Numeric labels for the music cues are those assigned by the authors of the existing datasets (Eerola & Vuoskoski, 2011 for Experiment 1; Aljanaki et al., 2017 for Experiments 2-4). Sound cues are numerically labelled to match the dataset from our pilot study, available on OSF. Word cues are labelled using the words used as stimuli. Grey connectors between points denote the pairs of cues that were matched in our stimulus selection procedure.

Ratings of retrieval intentionality, valence, arousal (intensity and energy), vividness, uniqueness, socialness, and importance were recorded for each retrieved memory using 5-point Likert scales (see Appendix A). Memory retrieval time was measured as the amount of time between the initial presentation of a stimulus and the time a participant clicked a button reading "I have recalled a memory"; participants were instructed to click this button as soon as a memory came to mind. Participants entered their age during the remembered event (hereafter "age at event") via a dropdown menu with age choices from 1 to 100 years, plus a "Don't know" option. To assess musical expertise, we included one question from the Ollen Musical Sophistication Index (Ollen, 2006) in which participants were asked to classify their level of musicianship; this question has been shown to be a suitable proxy for more detailed measures of musical sophistication in recent research (Zhang & Schubert, 2019). In addition we used one question from the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014) to probe the number of years participants had engaged in formal instrumental/vocal music lessons (see Appendix A).

Procedure

The experiment was presented online, hosted on Qualtrics. After providing informed consent, participants completed demographic questions (e.g., age, gender, see Appendix A). Next, they completed a sound check, in which they heard two clips (one music, one sound) that were not used in the main experiment; they were asked during this check to ensure the volume of their computer was turned up to a comfortable level, and asked to keep their device set at this same volume throughout the experiment. They then saw instructions for the main experiment, in which they were informed that they would hear a series of different types of sounds, and be asked to use each sound as a cue to help them think of an autobiographical memory. An autobiographical memory was defined as "a memory of an event that you were personally involved in, involving a specific place and time, that lasted no longer than one day." An example was provided as follows: "For example, the sound of a waterfall might bring back a memory of a boat trip around Niagara Falls that you took last July with your sister." Each music/sound clip started playing automatically, and would play for a maximum duration of 30 seconds. If a memory was evoked, participants were asked to click a button reading "I have recalled a memory" as soon as the memory came to mind. They then completed the memory description and rating questions, and familiarity question for the music trials[1]. There was no time limit imposed on completing these questions, and no character limit set for the memory descriptions. If no memory came to mind, they were asked simply not to click the "I have recalled a memory" button; if no button press was made during the 30-second presentation of a stimulus the experiment automatically, immediately advanced to the next trial. One practice trial was provided, using a sound stimulus that was not included in the main experiment. In the main experiment, participants heard the 9 music and 9 sound cues in a randomized order. Following completion of the main experiment, participants completed the two questions probing their previous musical training/musicianship level. The median time taken to complete the experiment was 26.5 minutes.

Results

Data for this and all experiments presented in this article are available on OSF: https://osf.io/dhcjb/. In total, 996 memories were evoked by the music and sound cues (49.4% of the cues evoked memories). The 18 individual stimuli varied in the number of memories evoked from 33 to 95 memories each. For the music cues, all three emotion categories

[1] Since the music was from commercially-released films, we assessed familiarity for the music trials in order to further validate (beyond our pilot studies) that the specific pieces of music were wholly unfamiliar to this participant sample. As the sound stimuli were sourced from a comparatively obscure database of professional sound effects, we did not further probe the familiarity of the sound clips to minimize the duration of the experiment (but see pilot results on the relative unfamiliarity of these in the Materials/Stimuli section).

evoked a similar number of memories (positive valence/high arousal: 129 memories, positive valence/low arousal: 127 memories, negative valence/high arousal: 122 memories; $\chi^2(2) = 0.33$, $p = .85$), whereas the sound-cued memories were less evenly distributed, with the negative valence/high arousal cues evoking fewer memories (positive valence/high arousal: 234 memories, positive valence/low arousal: 230 memories, negative valence/high arousal: 154 memories; $\chi^2(2) = 50.99$, $p < .001$). See Appendix C for a summary of the percentages of cues that evoked memories across all experiments reported in this article.

Preliminary analyses revealed that, in line with our expectations, the music cues were generally highly unfamiliar to participants; participants indicated that they thought they had heard the music before on 6% of all music trials completed by all participants. To account for the repeated-measures nature of the data and the fact that different participants reported different numbers of memories, the majority of the analyses we performed utilized linear mixed effects models, with each individual participant and stimulus included as random effects in the models. In a linear mixed effects model, we found no difference in the participant's age at event as a function of cue type (M_{music} = 20.00 years, SD_{music} = 10.50; M_{sounds} = 20.88 years, SD_{sounds} = 11.22; β = 0.82, 95% CI [-0.31, 1.91], SE = 0.56; $t(20) = 1.47$, $p = .16$). This indicates that music and sound cues did not show systematic differences in terms of the time period from which the associated memories were drawn.

The following analyses focus on our primary research questions of comparing the two cue types in terms of their effectiveness and phenomenological characteristics, and testing whether cue type and cue valence/arousal influenced the valence/arousal of the memories evoked.

Cue Efficacy

A paired-samples t-test revealed that sounds (M = 5.52, SD = 2.40) elicited significantly more autobiographical memories than music (M = 3.38, SD = 2.75), t(111) = -

10.03, $p < .001$. In a linear mixed effects model we found that the sounds ($M = 10.94$ s, $SD = 6.51$) evoked memories more quickly than the music ($M = 13.93$ s, $SD = 6.98$), β = -3.25, 95% CI [-4.28, -2.22], $SE = 0.52$; $t(19) = -6.27$, $p < .001$. Similarly, participant ratings of retrieval intentionality indicated that sounds ($M = 2.17$, $SD = 1.16$) evoked memories more spontaneously than music ($M = 2.47$, $SD = 1.24$), β = -0.30, 95% CI [-0.49, -0.11], $SE = 0.10$; $t(20) = -3.11$, $p = .005$. These findings indicate that emotional sounds are more effective cues for autobiographical memories than similarly emotional music.

Phenomenological Characteristics of Memories

Next, we examined whether ratings of phenomenological characteristics of the autobiographical memories varied as a function of cue type. As shown in Table 2, sound-evoked memories were rated as more vivid and social, with no significant differences in ratings of memory uniqueness or importance.

Table 2. Results of four linear mixed effects models using cue type (music/sound) to predict ratings of vividness, uniqueness, social content, and importance of memories (Experiment 1).

Dependent measure	Music cues mean rating (SD)	Sound cues mean rating (SD)	Beta (SE)	95% CI	t-test
Vividness	2.89 (1.14)	3.17 (1.14)	0.26 (0.11)	[0.04, 0.49]	$t(16) = 2.31$, $p = .034$*
Uniqueness	2.83 (1.23)	2.89 (1.17)	0.09 (0.07)	[-0.05, 0.23]	$t(948) = 1.20$, $p = .23$
Social content	2.50 (1.29)	3.07 (1.30)	0.47 (0.22)	[0.04, 0.91]	$t(17) = 2.14$, $p = .048$*
Importance	2.67 (1.21)	2.77 (1.23)	0.10 (0.13)	[-0.16, 0.35]	$t(16) = 0.77$, $p = .45$

Note: * = $p < .05$. All models include both ‘participant’ and ‘stimulus’ as random effects. In the model predicting ‘Uniqueness,’ due to the minimal residual variance in one of the random factors (stimulus), the degrees of freedom reflect the altered variance structure but the model operates similarly.

Emotional Valence and Arousal of Memories

Finally, we examined whether the valence and arousal of the cues impacted the valence and arousal of the associated memories, and whether these effects were consistent across both cue types. We fitted a linear mixed effects model with cue type, cue valence, and their interaction as predictors of memory valence. This revealed a significant main effect of cue type (β = -0.51, 95% CI [-0.81, -0.21], *SE* = 0.16; $t(19) = -3.21$, $p = .005$), with music evoking more positive memories than sounds, no significant main effect of cue valence (β = 0.23, 95% CI [-0.03, 0.49], *SE* = 0.14; $t(21) = 1.66$, $p = .11$), and a significant interaction between cue type and cue valence (β = 0.62, 95% CI [0.26, 0.98], *SE* = 0.19; $t(17) = 3.25$, $p = .005$). Post hoc tests using the Tukey method of p-value adjustment revealed that valence ratings for memories cued by music did not significantly differ as a function of cue valence ($p = .37$), but positive sounds evoked more positive memories than negative sounds ($p < .001$); negative music also elicited more positive memories than negative sounds ($p = .02$), with no difference in memory valence ratings between positive music and positive sounds ($p = .73$) (see Figure 2).

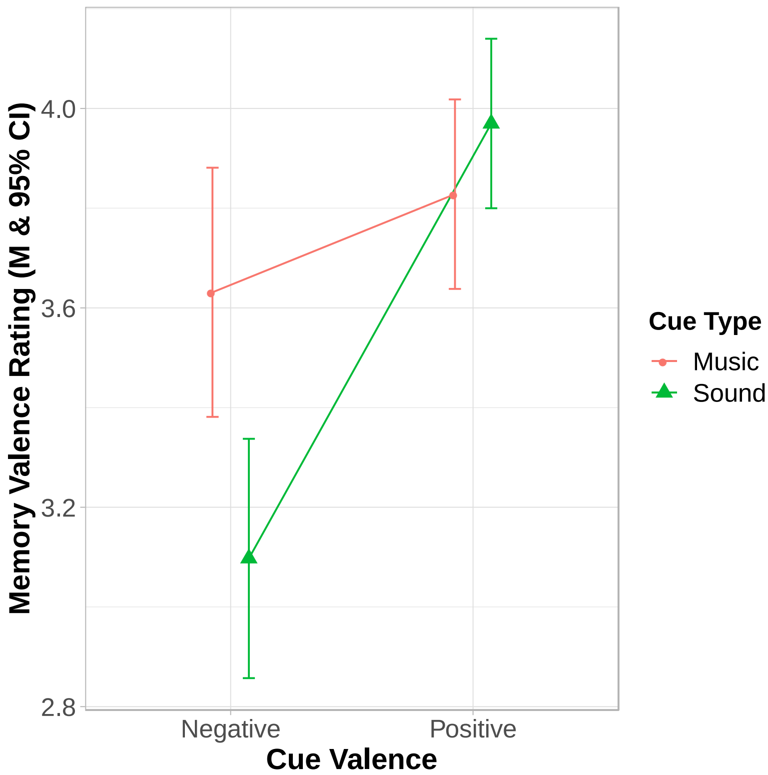

Figure 2. Mean ratings of memory valence by cue type and cue valence. Error bars denote 95% confidence intervals. Higher memory valence ratings indicate more positive memories.

To examine arousal effects, we fitted two separate models – one for each of the dependent variables (intensity, energy) used to measure arousal. The model predicting ratings of intensity of the memories showed no significant effects of cue type (β = 0.17, 95% CI [-0.07, 0.40], *SE* = 0.13; *t*(14) = 1.34, *p* = .20), cue arousal (β = 0.25, 95% CI [-0.04, 0.55], *SE* = 0.16; *t*(17) = 1.57, *p* = .13), or their interaction (β = -0.29, 95% CI [-0.68, 0.11], *SE* = 0.21; *t*(13) = -1.34, *p* = .20). The model predicting ratings of how energetic the memories were also showed no significant effect of cue type (β = 0.40, 95% CI [0.04, 0.76], *SE* = 0.20; *t*(13) = 2.03, *p* = .062), cue arousal (β = -0.37, 95% CI [-0.82, 0.08], *SE* = 0.24; *t*(14) = -1.52, *p* = .15), or interaction (β = 0.05, 95% CI [-0.57, 0.67], *SE* = 0.34; *t*(13) = 0.14, *p* = .89).

In sum, sound-evoked memories showed the predicted pattern of positive cues evoking more positive memories, but music-evoked memories did not significantly differ in valence in accordance with the valence of the music. Specifically, negative music evoked more positive memories than negative sounds. Memory arousal (when measured as either emotional intensity or energy) was not significantly predicted by cue arousal.

Post Hoc Coding of Memory Types

To better understand why the sound-evoked memories were rated higher than the music-evoked memories on several properties (e.g., vividness, social content, and marginally higher on energy), we examined the written descriptions of the memories. While the sound-evoked memories seemed to show a straightforward relationship between the sound and the associated memory in the vast majority of cases (e.g., a sound clip featuring a campfire primarily evoked memories involving camping, barbecues, or sitting by woodburning stoves), the music appeared to be evoking memories in several different ways. We coded the music-evoked memories into the following categories:

1. **Media**: The music reminded a person of an instance of watching a film or TV show, or playing/watching someone play a video game, such as “Pirates in the Carribean [sic] movie in the cinema” and “It reminds me when I was playing video games specifically the legend of Zelda.” (126 instances)
2. **Musical**: Something about the instrument or timbre of the music cued a memory. For instance, a clip of piano music cued the memory “sitting in the audience with my wife hearing my son play piano in public for his first time.” (90 instances)
3. **Other**: These memories do not fit the above categories. It was not always possible to discern the link between the music and the memory based on the participant’s short description. In various cases, emotion seemed to be the link between the two, for instance, a foreboding piece of music cued the memory “Whenever I had an essay to hand in while studying at my university” and a triumphant piece cued “It was in Riga I was in boxing championship where I got first place.” However, in other cases the

link between the music and memory was not entirely apparent, for instance, a peaceful sounding piece of music cued the memories “An argument with a family member” and “Sitting in my room studying for exams.” (115 instances).

We then examined whether the properties of the music-evoked memories varied depending on which category the memory fell into. The descriptive results in Figure 3 suggest that the “media” memories represent relatively generic memories (less energetic, important, social, unique, and vivid) in comparison to the other two memory categories. This suggests that the choice to use a dataset comprising solely film music as musical cues for this experiment may have had the unintended consequence of eliciting rather non-specific and mundane memories of watching films (or engaging in similar multimedia activities such as playing video games). As such, in Experiment 2 we sought to further explore the questions investigated here by replicating this approach using different music from a wider range of genres. We assumed that this wider range of music might elicit a wider range of associated memories, making the music-evoked memories more comparable to the sound-evoked memories. We thereby predicted that the features of music- and sound-evoked memories would be more similar in Experiment 2.

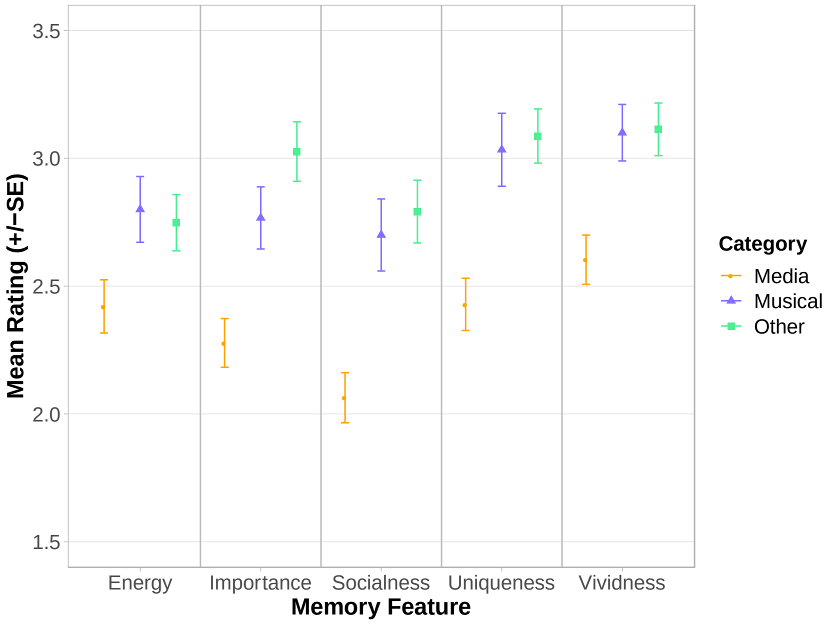

Figure 3. Mean ratings (+/- SE) of energy, importance, socialness, uniqueness, and vividness of music-evoked memories, grouped according to the three coded categories.

Experiment 2

Method

Design

The design was identical to Experiment 1.

Participants

Participants were recruited following the same exclusion criteria as Experiment 1 (no history of stroke, severe head injury, etc.). In total, 100 participants completed the study, with 1 excluded due to technical difficulties in sound playback. The final sample comprised 99 participants, aged 18 to 64 years (*M* = 28, *SD* = 10; 52 female and 46 male, 1 chose not to report a gender). Similar to Experiment 1, most participants were non-musicians (70%) or amateur musicians (26%), with 41% reporting no formal training in music and 32% reporting up to two years of previous musical training. Participants were sampled from 12 countries of current residence, with most residing in the UK (36%), Portugal (20%), or Poland (20%).

Similarly, the most commonly reported native languages were English (37%), Polish (21%), and Portuguese (20%). One participant was wearing a hearing aid, with no other participants reporting any hearing problems. All participants were recruited via Prolific and were compensated (£3.75). No participants who took part in Experiment 1 were recruited into Experiment 2.

Materials/Stimuli

Musical stimuli were selected from the MediaEval Database for Emotional Analysis in Music (DEAM; Aljanaki et al., 2017). This database comprises 1,802 royalty-free musical pieces/excerpts from a wide variety of genres. These pieces of music were rated as largely unfamiliar to participants in the initial curation of this corpus, and the database also comprises participant ratings of perceived valence and arousal for each piece of music. For the present work, excerpts were manually selected to represent the three valence/arousal quadrants used in Experiment 1, with the additional criteria that excerpts should not contain intelligible lyrics (or contain only short sections of lyrics that could be cut from the final excerpt) and should cover a wide range of genres. The final selection comprised 9 pieces of music from the following genres: classical, hip-hop, standard jazz, experimental jazz, country, pop, rock (2 excerpts), and electronic music. One excerpt contained some singing, but using nonsense syllables rather than intelligible lyrics.

Two of the sound stimuli were exchanged to provide a better match in terms of valence and arousal ratings to the new musical stimuli; the other 7 sounds were retained from Experiment 1 (see Figure 1 for all stimuli). All sounds and music were edited to 30 seconds in duration. The sound and music stimuli did not significantly differ in mean valence ($t(16) = 0.07$, $p = .95$) or mean arousal ratings ($t(16) = -0.17$, $p = .87$) when compared in independent-samples t-tests.

As in the first experiment, we collected familiarity data for the stimuli via a separate experiment with 60 participants (ages 18 to 63 years, $M = 35$, $SD = 11$, 37 female, all native English speakers), recruited via Prolific. We followed an identical procedure to that reported in Experiment 1, in which we collected ratings of whether participants had ever heard a *specific* piece of music or sound clip before, as well as ratings of the degree to which they had encountered *similar-sounding* music/sounds before (both rated on a scale from "1 (Never)" to "5 (Frequently)"). Half of the participants rated the musical stimuli and half rated the sound stimuli. For ratings of familiarity with a specific piece of music/sound clip, 17 of the 18 stimuli received a median rating of 1. One sound stimulus (number 26 "Rain and Thunder Ambience") received a median rating of 2; however, when probed where they had heard the clip before participants gave generic responses such as "Heavy rains and storm in movies" or provided links to similar but not identical sound clips on YouTube. Music and sound stimuli did not significantly differ in mean ratings of familiarity with the specific excerpts ($M_{music} = 1.45$, $SD_{music} = 0.19$; $M_{sounds} = 1.61$, $SD_{sounds} = 0.23$; $t(15) = -1.66$, $p = .12$), or ratings of whether participants had encountered similar-sounding music/sounds before ($M_{music} = 2.93$, $SD_{music} = 0.63$; $M_{sounds} = 2.79$, $SD_{sounds} = 0.65$; $t(16) = 0.47$, $p = .65$) in independent-samples t-tests.

Procedure

The procedure was identical to Experiment 1. The median time taken to complete the experiment was 23.7 minutes.

Results

In total, 932 memories were reported in response to the music and sound cues (52.3% of the cues evoked memories; with 29 to 86 memories reported per stimulus). For the music cues, memories were similarly distributed across the emotion categories (positive valence/high arousal: 125 memories, positive valence/low arousal: 134 memories, negative valence/high arousal: 121 memories; $\chi^2(2) = 1.22$, $p = .54$), whereas the sound-cued memories were less evenly distributed, with the negative valence/high arousal cues again evoking the fewest memories (positive valence/high arousal: 197 memories, positive valence/low arousal: 223 memories, negative valence/high arousal: 132 memories; $\chi^2(2) = 62.77$, $p < .001$; see Appendix C).

The musical stimuli from the DEAM dataset were highly unfamiliar, as anticipated, with participants reporting that they thought they had heard a piece of music before on 5% of all music trials completed across all participants. The memories evoked by the two cue types (music/sounds) did not differ in terms of the participant's age at the event in a linear mixed effects model with 'participant' and 'stimulus' as random effects ($M_{music} = 18.80$ years, $SD_{music} = 7.56$; $M_{sounds} = 18.94$ years, $SD_{sounds} = 9.30$, $\beta = -0.05$, 95% CI [-1.72, 1.60], $SE = 0.83$; $t(16) = -0.06$, $p = .96$).

Cue Efficacy

We followed the same analysis protocol as in Experiment 1 and found that sounds again evoked significantly more autobiographical memories in comparison to music ($M_{music} = 3.84$, $SD_{music} = 2.65$, $M_{sounds} = 5.58$, $SD_{sounds} = 2.39$; $t(98) = -8.06$, $p < .001$). However, the difference in mean retrieval times between music- and sound-evoked memories was not statistically significant ($M_{music} = 11.12$ s, $SD_{music} = 6.98$, $M_{sounds} = 10.29$ s, $SD_{sounds} = 6.33$; $\beta = -1.08$, 95% CI [-2.19, 0.04], $SE = 0.56$; $t(18) = -1.92$, $p = .070$), nor was the difference in retrieval intentionality ratings ($M_{music} = 2.24$, $SD_{music} = 1.19$, $M_{sounds} = 2.03$, $SD_{sounds} = 1.17$; $\beta = -0.18$, 95% CI [-0.36, -0.005], $SE = 0.09$; $t(19) = -2.05$, $p = .054$). Taken together, these findings indicate that sounds consistently evoked more memories across both Experiments 1 and 2, but that the musical cues utilized in Experiment 2 were more similar to the sounds in terms of the speed and spontaneity with which memories were evoked.

Phenomenological Characteristics of Memories

Music- and sound-evoked autobiographical memories in Experiment 2 did not significantly differ in ratings of vividness, uniqueness, social content, or importance of the memories (see Table 3). These results suggest that emotional music and emotional sounds evoke memories that are phenomenologically similar, and indicate that several of the differences found in Experiment 1 may be attributed to the limited selection of music used in that study (film music).

Table 3. Results of four linear mixed effects models using cue type (music/sound) to predict ratings of vividness, uniqueness, social content, and importance of memories (Experiment 2).

Dependent measure	Music cues mean rating (SD)	Sound cues mean rating (SD)	Beta (SE)	95% CI	t-test
Vividness	2.96 (1.12)	3.25 (1.23)	0.23 (0.14)	[-0.04, 0.51]	$t(16) = 1.68, p = .11$
Uniqueness	2.91 (1.22)	2.89 (1.23)	0.04 (0.10)	[-0.15, 0.24]	$t(17) = 0.41, p = .69$
Social content	3.15 (1.40)	2.96 (1.37)	-0.23 (0.24)	[-0.69, 0.24]	$t(16) = -0.96, p = .35$
Importance	2.63 (1.33)	2.75 (1.33)	0.14 (0.15)	[-0.15, 0.43]	$t(15) = 0.99, p = .34$

Note: All models include both ‘participant’ and ‘stimulus’ as random effects.

Emotional Valence and Arousal of Memories

In a linear mixed effects model to predict ratings of memory valence, we found a significant main effect of cue type ($\beta = -0.46$, 95% CI [-0.79, -0.12], $SE = 0.18$; $t(19) = -2.54$, $p = .020$), with music evoking more positive memories than sounds, no significant main effect of cue valence ($\beta = 0.08$, 95% CI [-0.21, 0.37], $SE = 0.16$; $t(19) = 0.50$, $p = .62$), and a significant interaction between cue type and cue valence ($\beta = 0.54$, 95% CI [0.13, 0.94], $SE =$

0.22; *t*(17) = 2.48, *p* = .024). As in Experiment 1, positive (*M* = 3.75, *SD* = 0.95) and negative (*M* = 3.69, *SD* = 1.00) musical cues did not significantly differ in terms of the valence of the memories they evoked (*p* = .96), but positive sounds (*M* = 3.85, *SD* = 1.04) evoked more positive memories than negative sounds (*M* = 3.19, *SD* = 1.01), *p* = .005, in post hoc tests using the Tukey method of p-value adjustment. Although memories cued by negative music again received higher (more positive) mean ratings of memory valence than memories cued by negative sounds, this difference did not reach statistical significance (*p* = .090). We found no significant difference in valence ratings of memories cued by positive music versus positive sounds (*p* = .91).

In a linear mixed effects model predicting ratings of emotional intensity of the memories, we found no significant effects of cue type (β = -0.14, 95% CI [-0.45, 0.17], *SE* = 0.17; *t*(15) = -0.84, *p* = .41), cue arousal (β = -0.11, 95% CI [-0.50, 0.27], *SE* = 0.21; *t*(15) = -0.55, *p* = .59), or their interaction (β = 0.28, 95% CI [-0.25, 0.80], *SE* = 0.29; *t*(13) = 0.97, *p* = .35). In contrast to Experiment 1, we did find that cue arousal was a significant predictor of energy ratings of the memories (β = -0.68, 95% CI [-1.10, -0.26], *SE* = 0.23; *t*(14) = -2.98, *p* = .009), with low arousal cues eliciting less energetic memories (*M* = 2.54, *SD* = 1.21) in comparison to high arousal cues (*M* = 3.07, *SD* = 1.23). This suggests that cue arousal does indeed impact on memory arousal, if arousal is conceptualized in terms of the energy content of the memories. Cue type did not significantly affect ratings of the energy of memories (β = -0.25, 95% CI [-0.59, 0.09], *SE* = 0.19; *t*(14) = -1.33, *p* = .20), and cue type did not interact with cue arousal in this model (β = 0.30, 95% CI [-0.28, 0.88], *SE* = 0.32; *t*(13) = 0.95, *p* = .36).

Interim Summary

Across two experiments, environmental sounds cued more autobiographical memories than music, but music cued memories that were more uniformly positive in valence (regardless of the music's emotional valence). Experiment 2 built on Experiment 1 by showing that diversifying the range of musical cues resulted in memories that were similar on several dimensions to sound-cued memories.

As environmental sounds have not been extensively studied as an autobiographical memory cue, we subsequently compared music against one of the most commonly employed cue types in the autobiographical memory research field (Koppel & Berntsen, 2015)—familiar words—to more firmly situate these findings in relation to previous literature and to broaden the comparison beyond the auditory domain. This also allowed us to incorporate the fourth valence/arousal quadrant (negative valence/low arousal) that was not captured by our sound stimuli. A number of parallels in the processing of music and language have been revealed to date (Patel, 2008). For instance, music elicits both affective and semantic priming effects over the same timescale as language (Koelsch et al., 2004; Steinbeis & Koelsch, 2011), suggesting that music and words might also serve as similar cues for autobiographical memories via affective and/or semantic associations.

Experiment 3

Method

Design

The design of Experiment 3 closely paralleled that of the first two experiments, but in this case the variable 'cue type' comprised a comparison of music to word cues. All of the dependent measures were the same as in Experiments 1 and 2, with the exception that we removed the retrieval intentionality ratings in order to reduce the overall experiment duration, given that we increased the number of stimuli from 18 to 24 to incorporate the negative valence/low arousal quadrant. In addition, memory retrieval times provide similar information to intentionality ratings, as involuntary memories are generally retrieved faster than voluntary memories (e.g., Schlagman & Kvavilashvili, 2008).[2]

Participants

We employed the same self-report health screening measures as in the first two experiments (no history of stroke, severe head injury, etc.). Due to the usage of English words as stimuli, we also recruited only participants who were of British nationality and spoke English as a first language. The final sample comprised 68 participants, aged 18 to 65 years (*M* = 32, *SD* = 11; 51 female and 16 male, 1 chose not to report a gender). Most participants were non-musicians (78%) or amateur musicians (21%). Most had no formal training in music (46%) or up to two years of previous musical training (35%). No participants reported any hearing problems; the only visual impairment reported was the requirement of corrective lenses (glasses/contact lenses), which were worn by 3 participants. Participants were recruited via Prolific and compensated for their participation (£3.75). No participants who took part in Experiments 1 or 2 were recruited for Experiment 3.

Materials/Stimuli

We used the same musical stimuli as in Experiment 2, but added three musical excerpts to represent the negative valence/low arousal emotion quadrant. These were also selected from the DEAM dataset (Aljanaki et al., 2017), and comprised 30-second musical excerpts with no lyrics from the following genres: blues, classical, folk. As in Experiment 2, we thereby presumed the specific pieces of music used would be highly unfamiliar to participants, but that the genres of music would be relatively familiar (see pilot experiment reported in Materials/Stimuli section of Experiment 2).

---

[2]In addition, for the data from Experiments 1 and 2 of the current article, we fit linear mixed effects models which showed that retrieval times significantly predicted intentionality ratings (Experiment 1: $\beta = 0.07$, 95% CI [0.06, 0.08], *SE* = 0.005; $t(985) = 12.97$, $p < .001$; Experiment 2: $\beta = 0.09$, 95% CI [0.08, 0.10], *SE* = 0.006; $t(919) = 14.57$, $p < .001$), such that longer retrieval times were associated with less spontaneous retrieval.

The word cues were single words selected from the English Lexicon Project (Balota et al., 2007). This database contains normative data for over 40,000 words, and incorporates information on both lexical characteristics and behavioral ratings of the words, including felt emotions and concreteness. To ensure similar words were chosen across the four emotion quadrants, we selected only words that were unmistakable nouns of 5 to 9 letters in length. We also filtered the dataset to include only words with an age of acquisition rating below 10 years, to ensure that these were easily comprehensible, well-known (familiar) words. In addition, for this experiment we utilized word cues that were relatively concrete (i.e., referring to a perceptible entity; Brysbaert et al., 2013), by selecting words that were rated higher than the median concreteness rating for the full dataset. Although there is no similar, existing measure of concreteness in music, relatively concrete words were selected here as they serve as a particularly rigorous comparator, given that concrete words have been found in previous research to be more effective cues for autobiographical memories than abstract words (e.g., Robinson, 1976; Uzer et al., 2012).

Once we had filtered the full dataset of possible words in terms of these parameters, we computed the distance in two-dimensional space between the mean valence and mean arousal rating of each word and the mean valence and mean arousal rating of each of the 12 pieces of music. For each piece of music, we exported the 10 closest words in terms of valence/arousal ratings; we then manually selected a word from this list that was as close in distance as possible, whilst accounting for various other factors (breaking ties, excluding words referencing concepts of a sexually explicit nature, etc.). The mean valence and arousal ratings for the 24 stimuli and the full list of word cues are displayed in Figure 1. The two cue types (music/words) did not significantly differ in mean valence ($t(22) = -0.08$, $p = .94$) or mean arousal ratings ($t(22) = 0.15$, $p = .88$) when compared in independent-samples t-tests.

Procedure

After giving informed consent, participants completed demographic questions (see Appendix A). They then completed a sound check, to ensure the volume of their computer was at a comfortable level, and were asked to keep their device set at this same volume throughout the experiment. The main experiment comprised a block of 12 music cues and a block of 12 word cues, which were presented in a counterbalanced order across participants. Music and word cues were randomized within these blocks. The task instructions were the same as in Experiments 1 and 2, that is, participants were asked to use each music/word cue to think of an autobiographical memory and to click a button reading “I have recalled a memory” as soon as a memory came to mind. Each block of trials began with a practice trial to familiarize them with the task. Each music clip played and each word was displayed for up to 30 seconds; if a participant did not recall a memory within 30 seconds the interface automatically advanced to the next trial. As before, participants completed the two questions probing their previous musical training/musicianship level following the main experiment. The median time taken to complete the experiment was 27.1 minutes.

Results

A total of 784 memories were reported (48.0% of the cues evoked memories), with 19 to 56 memories reported per stimulus. Music cues again did not significantly differ in terms of the distribution of memories across emotion categories (positive valence/high arousal: 91 memories, positive valence/low arousal: 86 memories, negative valence/high arousal: 91 memories; negative valence/low arousal: 69 memories; $\chi^2(3) = 6.61$, $p = .086$), whereas the word-cued memories were less evenly distributed, with the negatively valenced cues evoking the fewest memories (positive valence/high arousal: 144 memories, positive valence/low arousal: 123 memories, negative valence/high arousal: 90 memories; negative valence/low arousal: 90 memories; $\chi^2(3) = 41.81$, $p < .001$; see also Appendix C).

The age at event of the memories did not differ as a function of cue type (music/words) in a linear mixed effects model with 'participant' and 'stimulus' as random effects (M_{music} = 20.02 years, SD_{music} = 8.96; M_{words} = 20.45 years, SD_{words} = 10.64, β = -0.15, 95% CI [-1.99, 1.64], *SE* = 0.91; $t(24) = -0.17$, $p = .87$). The DEAM musical stimuli were again highly unfamiliar; participants reported that they thought they had heard a piece of music before on 2.6% of all music trials completed across all participants.

Cue Efficacy

The words evoked significantly more autobiographical memories than the music (M_{music} = 4.96, SD_{music} = 3.39, M_{words} = 6.57, SD_{words} = 2.93, $t(67) = -4.31$, $p < .001$). Word-evoked memories were also retrieved significantly faster than music-evoked memories (M_{music} = 12.61 s, SD_{music} = 7.16, M_{words} = 6.32 s, SD_{words} = 5.80, β = -6.36, 95% CI [-7.57, -5.14], *SE* = 0.62; $t(20) = -10.34$, $p < .001$).

Phenomenological Characteristics of Memories

Results comparing the phenomenological characteristics of the music- and word-evoked memories are displayed in Table 4. Music-evoked memories were rated as significantly more social and important than word-evoked memories, suggesting music may be more effective at eliciting certain types of memories (those that relate to particular people and are personally valued) in comparison to words. On the other hand, words evoked memories rated as significantly more unique (i.e., more likely to be a rare or once-in-a-lifetime event). This may be related to the fact that words elicited memories almost twice as quickly as music, as memories that are more directly and spontaneously retrieved tend to be more specific (e.g., Berntsen & Hall, 2004). The two memory types did not significantly differ in ratings of vividness.

Table 4. Results of four linear mixed effects models using cue type (music/word) to predict ratings of vividness, uniqueness, social content, and importance of memories (Experiment 3).

Dependent measure	Music cues mean rating (SD)	Word cues mean rating (SD)	Beta (SE)	95% CI	t-test
Vividness	2.98 (1.05)	3.02 (1.15)	0.05 (0.11)	[-0.17, 0.27]	$t(26) = 0.47$, $p = .64$
Uniqueness	2.70 (1.04)	2.92 (1.25)	0.25 (0.12)	[0.02, 0.48]	$t(21) = 2.13$, $p = .046$*
Social content	3.16 (1.28)	2.38 (1.15)	-0.77 (0.16)	[-1.07, -0.46]	$t(23) = -4.90$, $p < .001$***
Importance	2.71 (1.14)	2.35 (1.18)	-0.35 (0.14)	[-0.64, -0.07]	$t(24) = -2.49$, $p = .020$*

Note: * = $p < .05$, *** = $p < .001$. All models include both 'participant' and 'stimulus' as random effects.

Emotional Valence and Arousal of Memories

Similar to the first two experiments, we found a significant effect of cue type (β = -0.87, 95% CI [-1.18, -0.56], *SE* = 0.16; $t(23) = -5.25$, $p < .001$), with music evoking more positive memories than words, no significant effect of cue valence (β = 0.28, 95% CI [-0.03, 0.59], *SE* = 0.16; $t(23) = 1.70$, $p = .10$), and a significant interaction of cue type and cue valence (β = 0.57, 95% CI [0.14, 0.99], *SE* = 0.23; $t(21) = 2.50$, $p = .021$) in predicting memory valence. Pairwise post hoc tests with Tukey correction showed that positive words (*M* = 3.66, *SD* = 1.10) evoked more positive memories than negative words (*M* = 2.82, *SD* = 1.12), $p < .001$, but positive (*M* = 3.96, *SD* = 0.94) and negative music (*M* = 3.66, *SD* = 1.02) again did not significantly differ in their effects on memory valence ($p = .35$). Negative music evoked more positive memories than negative words ($p < .001$), but positive music and positive words evoked similarly valenced memories ($p = .26$).

Again, no significant effects of cue type ($\beta = -0.09$, 95% CI [-0.40, 0.21], $SE = 0.16$; $t(21) = -0.59$, $p = .56$), cue arousal ($\beta = -0.02$, 95% CI [-0.33, 0.29], $SE = 0.17$; $t(24) = -0.12$, $p = .90$), or their interaction ($\beta = -0.29$, 95% CI [-0.72, 0.14], $SE = 0.23$; $t(22) = -1.28$, $p = .22$) were found on the emotional intensity of the memories. However, in a model predicting energy ratings of the memories, we found significant main effects of cue type ($\beta = -1.06$, 95% CI [-1.30, -0.83], $SE = 0.12$; $t(19) = -8.65$, $p < .001$), with music evoking more energetic memories than words, and cue arousal ($\beta = -0.93$, 95% CI [-1.19, -0.68], $SE = 0.13$; $t(27) = -7.01$, $p < .001$), with high arousal cues evoking more energetic memories than low arousal cues. In addition, there was a significant interaction ($\beta = 1.00$, 95% CI [0.66, 1.34], $SE = 0.18$; $t(21) = 5.62$, $p < .001$), which is visualized in Figure 4. Pairwise post hoc tests with Tukey correction revealed that high arousal music evoked more energetic memories than low arousal music ($p < .001$), but energy ratings of the memories did not differ in relation to the arousal of the word cues ($p = .94$). Low arousal music and low arousal words evoked similarly energetic memories ($p = .96$), but high arousal music evoked more energetic memories than high arousal words ($p < .001$).

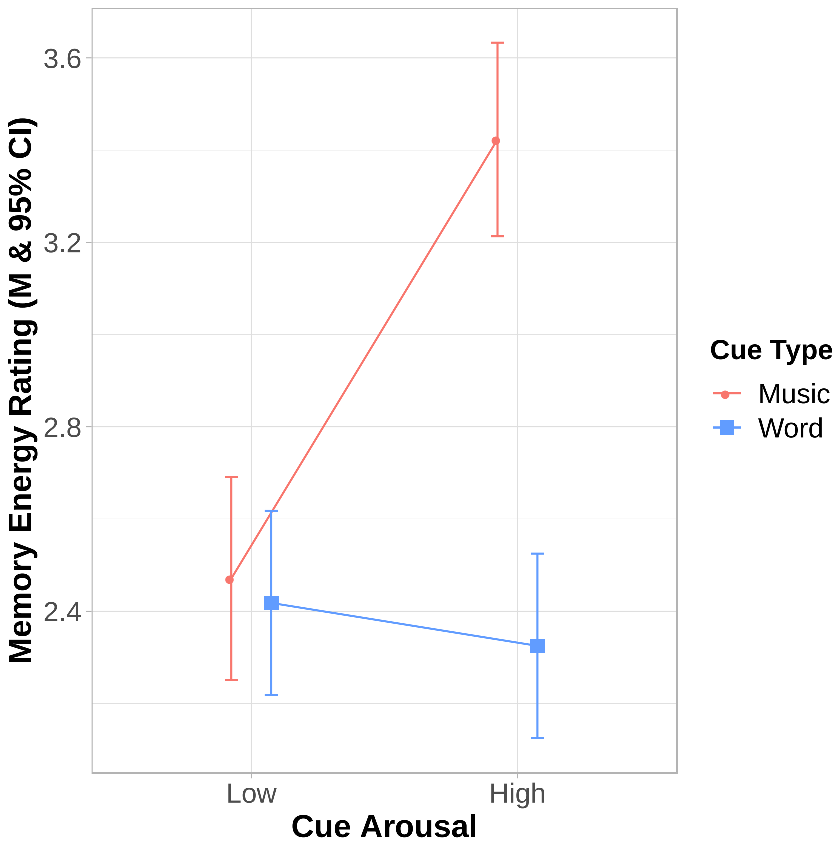

Figure 4. Mean ratings of memory energy by cue type and cue arousal. Error bars denote 95% confidence intervals.

Interim Summary

Several results in Experiment 3 replicated Experiments 1 and 2: unfamiliar music cued fewer memories than matched emotional cues and elicited similarly positive memories regardless of its emotional valence. However, we also found more substantial differences between ratings of the content and experience of the memories than we did when comparing music to sound cues. In a final experiment, we compared the same music cues used in Experiment 3 to more abstract words. Although analogous concreteness measures have not been established for music, music is often conceptualized as a relatively non-referential stimulus, with “floating intentionality” that varies across people and situations (Cross, 1999, 2014). We therefore anticipated that music might be more similar to abstract words as an autobiographical memory cue.

Experiment 4

Method

Design

The design was identical to Experiment 3.

Participants

Participant recruitment followed the same self-report health screening and screening by nationality (British) and first language (English) as Experiment 3. In total, 71 participants completed the experiment, aged 18 to 63 years (*M* = 35, *SD* = 13; 43 female and 26 male, 2 chose not to report a gender). Again, participants were primarily non-musicians (76%) or amateur musicians (23%) with no formal training in music (45%) or up to two years of previous musical training (22%).[3] Two participants reported mild hearing loss; the only visual impairment reported was the requirement of corrective lenses (glasses/contact lenses), which were worn by 3 participants. Participants were recruited via Prolific and compensated for participation (£3.75). No participants who took part in Experiment 1, 2, or 3 were recruited for Experiment 4.

Materials/Stimuli

The musical stimuli were identical to those used in Experiment 3. The word stimuli were selected following the same procedures as in Experiment 3: they were all unmistakable

[3] To understand whether the presence of participants with musical training may have confounded the results of any of these experiments, we reran all analyses reported in this article whilst excluding participants with 6 or more years of musical training (following the "6-year rule" for musicianship ascertained by Zhang et al., 2020). This resulted in the exclusion of 5-25 participants per experiment. All results were replicated in terms of their statistical (non-)significance with only two exceptions: in Experiment 1 there was no longer a significant effect of cue type on social content of the memories (p = .059) and in Experiment 2 cue type significantly affected retrieval intentionality ratings (p = .005), such that sounds evoked memories more spontaneously than music.

nouns of 5 to 9 letters in length with an age of acquisition rating below 10 years. However, in this case we selected words that were relatively abstract, by selecting only words rated *lower* than the median concreteness rating for the full dataset. Words were matched to the musical stimuli following the procedures outlined in Experiment 3. See Figure 1 for the full list of selected words, and mean valence and arousal ratings of both the music and word stimuli. In independent-samples t-tests, the words and music did not significantly differ in terms of mean valence ($t(22) = 0.19$, $p = .85$) or mean arousal ratings ($t(22) = 0.18$, $p = .86$).

Procedure

The procedure was identical to Experiment 3. The median time taken to complete the experiment was 29.4 minutes.

Results

A total of 875 memories were reported, that is, 51.3% of the cues evoked memories. The number of memories reported per stimulus ranged from 21 to 58. In this experiment, music cues did significantly differ in terms of the distribution of memories reported across emotion categories, which was due to the lower prevalence of memories reported for the negative valence/low arousal cues (positive valence/high arousal: 94 memories, positive valence/low arousal: 100 memories, negative valence/high arousal: 100 memories; negative valence/low arousal: 72 memories; $\chi^2(3) = 10.17$, $p = .017$). Word-cued memories did not significantly differ in terms of their distribution across the emotion categories (positive valence/high arousal: 124 memories, positive valence/low arousal: 135 memories, negative valence/high arousal: 121 memories; negative valence/low arousal: 129 memories; $\chi^2(3) = 2.20$, $p = .53$; see Appendix C).

The same analyses were run on this dataset as those reported in Experiment 3. Participants reported that they thought they had heard that particular piece of music before on only 2.1% of all music trials, and memories evoked by the two cue types did not significantly differ in terms of the participant's age at event (M_{music} = 23.62 years, SD_{music} = 10.92; M_{words} = 22.71 years, SD_{words} = 11.28, β = -1.12, 95% CI [-3.41, 1.18], SE = 1.16; $t(24) = -0.96$, p = .35).

Cue Efficacy

Similar to the first three experiments, abstract word cues evoked more memories overall in comparison to music cues (M_{music} = 5.15, SD_{music} = 3.28, M_{words} = 7.17, SD_{words} = 3.08, $t(70) = -6.76$, $p < .001$). Word-evoked memories were also retrieved significantly faster than music-evoked memories (M_{music} = 12.50 s, SD_{music} = 6.75, M_{words} = 6.50 s, SD_{words} = 5.38, β = -6.12, 95% CI [-6.99, -5.25], SE = 0.44; $t(20) = -13.89$, $p < .001$).

Phenomenological Characteristics of Memories

Music- and word-evoked memories were again compared in their phenomenological characteristics (see Table 5). As in Experiment 3, word-evoked memories were rated as significantly more unique, and were this time also rated as significantly more vivid, suggesting these more unique memories were also more richly reexperienced by participants in Experiment 4. In contrast to the findings of Experiment 3, no significant differences were found in ratings of social content or importance of the memory as a function of cue type, suggesting emotional abstract words serve as similar cues to emotional music in these regards.

Table 5. Results of four linear mixed effects models using cue type (music/word) to predict ratings of vividness, uniqueness, social content, and importance of memories (Experiment 4).

Dependent measure	Music cues mean rating (SD)	Word cues mean rating (SD)	Beta (SE)	95% CI	t-test
Vividness	2.80 (1.03)	3.27 (1.13)	0.46 (0.11)	[0.25, 0.67]	*t*(25) = 4.33, *p* < .001***
Uniqueness	2.84 (1.25)	3.13 (1.28)	0.33 (0.12)	[0.09, 0.56]	*t*(22) = 2.72, *p* = .012*
Social content	2.86 (1.36)	2.43 (1.28)	-0.37 (0.21)	[-1.07, -0.46]	*t*(23) = -1.80, *p* = .085
Importance	2.54 (1.28)	2.64 (1.32)	0.13 (0.16)	[-0.18, 0.44]	*t*(25) = 0.83, *p* = .41

Note: * = *p* < .05, *** = *p* < .001. All models include both 'participant' and 'stimulus' as random effects.

Emotional Valence and Arousal of Memories

The memory valence results replicated the first three studies. Specifically, there was a significant effect of cue type (β = -1.20, 95% CI [-1.56, -0.85], *SE* = 0.19; *t*(21) = -6.30, *p* < .001), with music evoking more positive memories than words, no significant effect of cue valence (β = 0.37, 95% CI [0.01, 0.73], *SE* = 0.19; *t*(23) = 1.90, *p* = .070), and a significant interaction of cue type and cue valence (β = 0.86, 95% CI [0.36, 1.36], *SE* = 0.27; *t*(21) = 3.21, *p* = .004) in predicting memory valence ratings. Positive (*M* = 3.89, *SD* = 0.89) and negative (*M* = 3.54, *SD* = 1.13) music cues elicited similarly valenced memories (*p* = .26), whereas positive words (*M* = 3.51, *SD* = 1.10) elicited more positive memories than negative words (*M* = 2.28, *SD* = 1.18), *p* < .001. Negatively valenced music elicited more positive memories than negative words (*p* < .001), while positive music and positive words did not significantly differ in their effects on memory valence (*p* = .29).

Emotional intensity was again not significantly predicted by cue type (β = 0.48, 95% CI [0.04, 0.92], *SE* = 0.24; *t*(21) = 2.04, *p* = .054), cue arousal (β = -0.09, 95% CI [-0.54, 0.36], *SE* = 0.24; *t*(23) = -0.37, *p* = .72), or their interaction (β = -0.15, 95% CI [-0.78, 0.47], *SE* = 0.33; *t*(21) = -0.46, *p* = .65). However, cue arousal was a significant predictor of the energy ratings of the memories (β = -0.98, 95% CI [-1.33, -0.63], *SE* = 0.19; *t*(24) = -5.30, *p* < .001), with high arousal cues eliciting more energetic memories (*M* = 3.00, *SD* = 1.29) than low arousal cues (*M* = 2.10, *SD* = 1.18). There was no significant effect of cue type (β = -0.11, 95% CI [-0.44, 0.23], *SE* = 0.18; *t*(21) = -0.60, *p* = .56) or significant interaction of cue type and cue arousal (β = 0.17, 95% CI -0.31, 0.65], *SE* = 0.25; *t*(21) = 0.68, *p* = .50) in predicting ratings of memory energy. This memory energy result replicates that found in Experiment 2, and shows some similarity to the results of Experiment 3. However, in Experiment 3 only the music showed the expected effect whereby high arousal music cued more energetic memories, whereas in Experiment 4 both cue types (music and words) elicited more energetic memories when high arousal cues were presented.

Discussion

Across four experiments, we compared unfamiliar instrumental music to unfamiliar environmental sounds and familiar (concrete and abstract) words matched on their emotional valence and arousal as cues for autobiographical memories. The music was consistently less effective than matched emotional cues in terms of the number of memories evoked. This result parallels some extant literature; for instance, Belfi et al. (2016) found that popular music evoked significantly fewer autobiographical memories than photographs of famous faces and Baird et al.'s (2018) healthy older participants reported significantly more memories in response to photographs of famous events than music. Taken together, these findings suggest that music may not be an optimal retrieval cue if one’s aim is simply to elicit as many autobiographical memories as possible.

A second result that emerged across all four experiments was an interaction between cue type and cue valence in their effects on memory valence. While negative sounds/words evoked more negative memories than positive sounds/words, unfamiliar music consistently evoked relatively positive memories regardless of the emotional valence of the music. This indicates that music may be somewhat unique, in comparison to other emotional memory cues, in its ability to evoke primarily positive lifetime memories, a result that has key implications for the therapeutic use of music. Although in contrast to two studies that found that memory valence did vary in accordance with the valence of musical cues (Schulkind & Woldorf, 2005; Sheldon & Donahue, 2017), our results are supported by evidence that many listeners intensely enjoy and experience positive emotions in response to sad and aggressive/violent music (Eerola et al., 2018; Taruffi & Koelsch, 2014; Thompson et al., 2019). These positive responses to negative music have been attributed to various explanations, including the lack of “real-life” consequences present when engaging with aesthetic objects such as music. One possible explanation for our results is that the negative-sounding (e.g., sad, angry) musical stimuli actually induced positive emotional responses, which then triggered the recall of similarly positive memories. Conversely, it could be that the negative music used here does actually induce negative emotional responses, but nevertheless tends to bring back memories associated with listening to such music in positive contexts (e.g., attending a hard rock gig with friends, hearing a sad yet beautiful piece of music in a concert). Future research should seek to further isolate and determine the temporal order of the associated responses (emotional responses to the music, recall of the memory, emotional responses to the memory).

We replicated and extended the findings of Sheldon and Donahue (2017) in that the arousal level of the music, sound, and word cues did not impact emotional intensity ratings of the retrieved memories. However, cue arousal *did* impact ratings of the energetic nature of the memories; in three of our four experiments higher arousal cues evoked significantly more energetic memories. This suggests that ratings of the energy content of memories may be a more suitable index of memory arousal than emotional intensity, but future research is needed that systematically compares different measures of emotional arousal (including various rating scales and physiological indices of emotional arousal, e.g., Salgado & Kingo, 2019) against one another.

After resolving the unintended consequences of film music eliciting rather generic memories of films/other media (Experiment 1), we found in Experiment 2 that unfamiliar emotional music elicited memories that were phenomenologically similar to memories evoked by unfamiliar emotional sounds (in terms of vividness, social content, etc.). This suggests that auditory cues of comparable emotional valence and arousal result in relatively similar autobiographical memory experiences. More differences were found between unfamiliar music- and familiar word-cued memories (Experiments 3 and 4). The most consistent difference was that words evoked memories rated as more unique, which were also rated as significantly more vivid in Experiment 4. As word-cued memories were also retrieved faster, this indicates that word cues were able to directly isolate more specific episodes from one's autobiographical memory store than music[4]. This aligns with previous findings that music tends to bring back memories of time periods and extended/repeated events more often than specific events (Baird et al., 2018; Janata et al., 2007; Loveday et al., 2020). Mazzoni et al. (2014) found that both complex and simple pictorial cues elicited fewer

[4] See also previous research demonstrating that more involuntarily retrieved memories are recalled faster and are more specific (Berntsen et al., 2013; Bernsten & Hall, 2004; Schlagman & Kvavilashvili, 2008).

involuntary autobiographical memories than word cues, and argued that word cues may allow more freedom for participants to generate their own details, whereas the details provided within pictorial cues may actually hinder recall (e.g., if a participant is shown a picture of a pink house, but only has memories pertaining to white houses). It may be that our unfamiliar music cues constrained recall in a similar way to these pictorial cues. Music also elicited memories rated as more social and important when compared against relatively concrete words (Experiment 3), but not in comparison to abstract words (Experiment 4). It may be that the specific concrete words we selected as stimuli did not reference concepts that were particularly social or highly valued by our participants, whereas the less referential nature of both the abstract words and music may have led to more similar memories being retrieved. However, future research is needed using a broader range of both word types to more comprehensively explore this topic.

Limitations and Future Directions

Several additional factors should be explored in future work. To control for potential familiarity differences between participants in the present study we used relatively unfamiliar musical excerpts and sound clips. However, word cues by their nature are familiar, and thus it is possible some of the differences that emerged between music- and word-cued memories may be attributed to previous exposure to the particular cues. In future research we aim to systematically manipulate prior familiarity with musical stimuli, in order to compare familiar versus unfamiliar musical cues, as well as familiar music versus familiar word cues. In addition, we aim to explore how familiarity and emotional features of retrieval cues interact; for instance, it may be that emotional aspects (such as arousal) of unfamiliar music guide access to emotionally congruent memories, but that familiar music that is already associated with particular autobiographical contexts does not show this emotional congruence effect between the cue and memory. Another related factor is musical preference. Experiments 2-4 utilized a wide range of musical genres in an attempt to elicit a wide range of memories. However, it is likely that participants preferred some of these genres more than others, and future studies could take account of this factor to unpack how it influences the number and type of memories evoked. Finally, we instructed participants to recall specific events that lasted no longer than one day; future work could employ broader definitions of autobiographical memory to probe whether some of the cues utilized here are more likely to bring back general memories of lifetime periods than specific events.

Conclusion

This research represents the first, to our knowledge, to make a controlled comparison between music and emotionally matched cues for autobiographical memory. Unfamiliar music consistently elicited fewer memories than both unfamiliar sounds and familiar words, refuting the common belief that music is "special" in its ability to access autobiographical memories. However, our results also suggest that music may be distinctive as a cue for relatively positive lifetime memories, regardless of the emotional expression of the cue. Beyond their research implications, these findings may inform applied uses of music in therapeutic and commercial contexts, by suggesting that music is a particularly effective vehicle for stimulating positive reminiscence.

Author Contributions

KJ conceived of the study, obtained funding, collected and analyzed the data, and wrote the first draft of the manuscript. TE contributed to stimulus selection and design, analyzing the data, creating figures, and writing the manuscript.

Acknowledgements

Thanks to Anita Ghosh (Durham University) for assistance with selection of the environmental sound stimuli and pilot testing.

Data Availability Statement

Valence and arousal rating data for the environmental sound stimuli associated with this article can be accessed at https://osf.io/dhk8z/ and associated data for all experiments at https://osf.io/dhcjb/.

References

Aljanaki, A., Yang, Y. H., & Soleymani, M. (2017). Developing a benchmark for emotional analysis of music. *PLoS ONE*, *12*(3), e0173392. https://doi.org/10.1371/journal.pone.0173392

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*(2), 693–706. https://doi.org/10.3233/JAD-180627

Baird, A., Gelding, R., Brancatisano, O., & Thompson, W. F. (2020). A Preliminary Exploration of the Stability of Music- and Photo-Evoked Autobiographical Memories in People with Alzheimer's and Behavioral Variant Frontotemporal Dementia. *Music & Science*, *3*. https://doi.org/10.1177/2059204320957273

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459. https://doi.org/10.3758/BF03193014

Belfi, A. M., Bai, E., & Stroud, A. (2020). Comparing methods for analyzing music-evoked autobiographical memories. *Music Perception*, *37*(5), 392–402. https://doi.org/10.1525/mp.2020.37.5.392

Belfi, A. M., Karlan, B., & Tranel, D. (2016). Music evokes vivid autobiographical memories. *Memory, 24*(7), 979–989. https://doi.org/10.1080/09658211.2015.1061012

Berntsen, D. (2002). Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Memory & Cognition, 30*(7), 1010–1020. https://doi.org/10.3758/BF03194319

Berntsen, D., & Hall, N. M. (2004). The episodic nature of involuntary autobiographical memories. *Memory & Cognition, 32*(5), 789–803. https://doi.org/10.3758/BF03195869

Berntsen, D., Staugaard, S. R., & Sørensen, L. M. T. (2013). Why am I remembering this now? Predicting the occurrence of involuntary (spontaneous) episodic memories. *Journal of Experimental Psychology: General*, *142*(2), 426. https://doi.org/10.1037/a0029128

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin, 99*(2), 229-246.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148. https://doi.org/10.1037/0003-066X.36.2.129

Cady, E. T., Harris, R. J., & Knappenberger, J. B. (2008). Using music to cue autobiographical memories of different lifetime periods. *Psychology of Music, 36*(2), 157–177. https://doi.org/10.1177/0305735607085010

Congleton, A. R., Nielsen, N. P., & Berntsen, D. (2020). Through the gateway of the senses: Investigating the influence of sensory modality-specific retrieval cues on involuntary episodic memory. *Psychological Research.* https://doi.org/10.1007/s00426-020-01304-5

Cross, I. (1999). Is music the most important thing we ever did? Music, development and evolution. In Yi, S. W. (Ed.), *Music, Mind and Science* (pp. 10–39). Seoul, Korea: Seoul National University Press.

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music, 42*(6), 809–819. https://doi.org/10.1177/0305735614543968

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Eerola, T., Vuoskoski, J. K., Peltola, H. R., Putkinen, V., & Schäfer, K. (2018). An integrative review of the enjoyment of sadness associated with music. *Physics of Life Reviews, 25,* 100–121. https://doi.org/10.1016/j.plrev.2017.11.016

El Haj, M., Postal, V., & Allain, P. (2012). Music enhances autobiographical memory in mild Alzheimer's disease. *Educational Gerontology, 38*(1), 30–41. https://doi.org/10.1080/03601277.2010.515897

Ford, J. H., Addis, D. R., & Giovanello, K. S. (2012). Differential effects of arousal in positive and negative autobiographical memories. *Memory, 20*(7), 771–778. https://doi.org/10.1080/09658211.2012.704049

Herz, R. S. (1998). An examination of objective and subjective measures of experience associated to odors, music, and paintings. *Empirical Studies of the Arts, 16*(2), 137–152. https://doi.org/10.2190/c43t-cjr2-9lpd-r0pb

Herz, R. S. (2004). A naturalistic analysis of autobiographical memories triggered by olfactory visual and auditory stimuli. *Chemical Senses, 29*(3), 217–224. https://doi.org/10.1093/chemse/bjh025

Herz, R. S., & Schooler, J. W. (2002). A naturalistic study of autobiographical memories evoked by olfactory and visual cues: Testing the Proustian hypothesis. *American Journal of Psychology, 115*(1), 21–32. https://doi.org/10.2307/1423672

Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of Life Reviews, 7*(1), 88–131. https://doi.org/10.1016/j.plrev.2010.01.006

Jakubowski, K., Belfi, A.M., & Eerola, T. (2021). Phenomenological differences in music- and television-evoked autobiographical memories. *Music Perception, 38*(5), 435-455.

Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterisation of music-evoked autobiographical memories. *Memory, 15*(8), 845–860.

https://doi.org/10.1080/09658210701734593

Juslin, P. N., & Sloboda, J. A. (Eds.). (2010). *Handbook of music and emotion: Theory, research, applications.* Oxford University Press.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302-307.

Koppel, J., & Berntsen, D. (2015). The peaks of life: The differential temporal locations of the reminiscence bump across disparate cueing methods. *Journal of Applied Research in Memory and Cognition*, *4*(1), 66–80. https://doi.org/10.1016/j.jarmac.2014.11.004

Loveday, C., Woy, A., & Conway, M. A. (2020). The self-defining period in autobiographical memory: Evidence from a long-running radio show. *Quarterly Journal of Experimental Psychology.* https://doi.org/10.1177/1747021820940300

Mazzoni, G., Vannucci, M., & Batool, I. (2014). Manipulating cues in involuntary autobiographical memory: Verbal cues are more effective than pictorial cues. *Memory & Cognition*, *42*(7), 1076-1085.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS One, 9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Ollen, J.E. (2006). A criterion-related validity test of selected indicators of musical sophistication using expert ratings (Unpublished doctoral dissertation). Ohio State University, Ohio, USA.

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715–734.

https://doi.org/10.1017/S0954579405050340

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67*(3), 525–539. https://doi.org/10.1037/0022-3514.67.3.525

Robinson, J. A. (1976). Sampling autobiographical memory. Cognitive Psychology, 8(4), 578–595. https://doi.org/10.1016/0010-0285(76)90020-7

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*(3), 524–535.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Salgado, S., & Kingo, O. S. (2019). How is physiological arousal related to self-reported measures of emotional intensity and valence of events and their autobiographical memories? *Consciousness & Cognition, 75*. https://doi.org/10.1016/j.concog.2019.102811

Schlagman, S., & Kvavilashvili, L. (2008). Involuntary autobiographical memories in and outside the laboratory: How different are they from voluntary autobiographical memories? *Memory & Cognition, 36*(5), 920–932. https://doi.org/10.3758/MC.36.5.920

Schulkind, M. D., & Woldorf, G. M. (2005). Emotional organization of autobiographical memory. *Memory & Cognition, 33*(6), 1025–1035. https://doi.org/10.3758/BF03193210

Sheldon, S., & Donahue, J. (2017). More than a feeling: Emotional cues impact the access and experience of autobiographical memories. *Memory & Cognition, 45*(5), 731–744. https://doi.org/10.3758/s13421-017-0691-6

Singer, J. A., & Salovey, P. (1988). Mood and memory: Evaluating the network theory of affect. *Clinical Psychology Review, 8*(2), 211–251. https://doi.org/10.1016/0272-

7358(88)90060-8

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience*, *23*(3), 604-621.

Talarico, J. M., Berntsen, D., & Rubin, D. C. (2009). Positive emotions enhance recall of peripheral details. *Cognition & Emotion*, *23*(2), 380–398. https://doi.org/10.1080/02699930801993999

Talarico, J. M., Labar, K. S., & Rubin, D. C. (2004a). Emotional intensity predicts autobiographical memory experience. *Memory & Cognition*, *32*(7), 1118–1132. https://doi.org/10.3758/BF03196886

Taruffi, L., & Koelsch, S. (2014). The paradox of music-evoked sadness: An online survey. *PLoS ONE*, *9*(10), e110490. https://doi.org/10.1371/journal.pone.0110490

Thompson, W. F., Geeves, A. M., & Olsen, K. N. (2019). Who enjoys listening to violent music and why? *Psychology of Popular Media Culture*, *8*(3), 218–232. https://doi.org/10.1037/ppm0000184

Uzer, T. (2016). Retrieving autobiographical memories: How different retrieval strategies associated with different cues explain reaction time differences. *Acta Psychologica*, *164*, 144–150. https://doi.org/10.1016/j.actpsy.2016.01.004

Uzer, T., Lee, P. J., & Brown, N. R. (2012). On the prevalence of directly retrieved autobiographical memories. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(5), 1296–1308. https://doi.org/10.1037/a0028142

Williams, J. M. G., Healy, H. G., & Ellis, N. C. (1999). The effect of imageability and predicability of cues in autobiographical memory. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *52*(3), 555-579. https://doi.org/10.1080/713755828

Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect:

Integration and beyond. *Journal of Personality and Social Psychology, 77*(3), 600–619. https://doi.org/10.1037/0022-3514.77.3.600

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*(6), 831–844.

Zhang, J. D., & Schubert, E. (2019). A single item measure for identifying musician and nonmusician categories based on measures of musical sophistication. *Music Perception, 36*(5), 457–467. https://doi.org/10.1525/MP.2019.36.5.457

Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, *48*(3), 389-409.

Appendix A: Demographic questions, questions probing memory features, and musical background questions, as used in in Experiments 1-4 (response options in italics)

Demographics:

Please enter your current age (in years)
Dropdown menu ranging from 18 to 100

Please enter your gender
Male/ Female/ Other/ Prefer not to say

In what country do you currently live?
Dropdown menu with country list (provided by Qualtrics)

What is your first (native) language?
Open text box

Do you currently suffer from any hearing impairments?
Yes/ No

**If participant responded ‘Yes’ to the hearing impairment question above, the following was displayed:
Please provide a short description of the hearing impairment, and any measures you are currently taking to correct it (wearing a hearing aid, etc.).
Open text box

Do you currently suffer from any visual impairments? [question used in Experiments 3 & 4 only]

Yes/No

**If participant responded ‘Yes’ to the visual impairment question above, the following was displayed:
Please provide a short description of the visual impairment, and any measures you are currently taking to correct it (wearing glasses, etc.).
Open text box

Memory features:

Please rate the **degree of control** you had over bringing the memory to mind, in terms of whether the memory came to you spontaneously or you tried deliberately to think of this memory. [question used in Experiments 1 & 2 only]
Completely spontaneous recall/ Somewhat spontaneous recall/ Neither spontaneous nor deliberate recall/ Somewhat deliberate recall/ Completely deliberate recall

How **vivid** is the memory in your mind? In other words, how clear is the image of the event in your mind?
1 (not at all vivid)/ 2 (a little vivid)/ 3 (somewhat vivid)/ 4 (very vivid)/ 5 (extremely vivid)

Please rate the **emotional content** of the memory, in terms of how negative or positive it is.
1 (very negative)/ 2 (somewhat negative)/ 3 (neither negative nor positive)/ 4 (somewhat positive)/ 5 (very positive)

Please rate the **emotional intensity** of the memory.
1 (not at all intense)/ 2 (a little intense)/ 3 (somewhat intense)/ 4 (very intense)/ 5 (extremely intense)

Please provide a short (1 sentence) description of the memory that you recalled, including details of what you were doing, who you were with, and where you were in the remembered event. (Note: If this is a memory you do not wish to share the content of for personal reasons, please simply write 'Private'.)
Open text box

How old were you (in years) during the event that you recalled? If unsure, please give your best estimate.
Dropdown menu ranging from 1-100, including 'Don't know' option

How **unique** is the event that you remembered?
1 (not at all- this type of event happens all the time)/ 2 (a little unique)/ 3 (somewhat unique)/ 4 (very unique)/ 5 (extremely unique- once in a lifetime event)

How **social** is the event that you remembered?
1 (not at all social)/ 2 (a little social)/ 3 (somewhat social)/ 4 (very social)/ 5 (extremely social)

How **energetic** is the content of this memory?
1 (not at all energetic)/ 2 (a little energetic)/ 3 (somewhat energetic)/ 4 (very energetic)/ 5 (extremely energetic)

How **important** is this memory to your life story?
1 (not at all important)/ 2 (a little important)/ 3 (somewhat important)/ 4 (very important)/ 5 (extremely important)

Have you ever heard the piece of music we used as the cue for this memory before today? [question displayed after music cues only]
Yes/No/Not sure

<u>Musical background:</u>

Which of the following best describes you?
Non-musician/ Music-loving non-musician/ Amateur musician/ Serious amateur musician/ Semi-professional musician/ Professional musician

I have had __ years of formal training on a musical instrument (including voice) during my lifetime.
0/ 0.5/ 1/ 2/ 3-5/ 6-9/ 10 or more

Appendix B: Sounds used as stimuli in Experiments 1 and 2 (names from SoundsEffects+ database)

Emotion category	Experiment 1	Experiment 2
Positive valence/high arousal	Parade Crowd Ambience People Talking Outside Fireworks Display	Parade Crowd Ambience People Talking Outside Fireworks Display
Positive valence/low arousal	Beach at Night Ambience Forest Birds and Bees Ambience Fireplace Crackling Ambience	Rain and Thunder Ambience Forest Birds and Bees Ambience Fireplace Crackling Ambience
Negative valence/high arousal	Tropical Birds Squawking Spaceship Computers Beeping Fireworks Exploding	Tropical Birds Squawking Spaceship Computers Beeping Factory Machinery Ambience

Appendix C: Percentages of cues that evoked memories by cue type, valence, and arousal for Experiments 1-4

Experiment	Cue type	Positive valence/high arousal (%)	Positive valence/low arousal (%)	Negative valence/high arousal (%)	Negative valence/low arousal (%)
1	Music	38.4	37.8	36.3	N/A
	Sound	69.6	68.5	45.8	N/A
2	Music	42.1	45.1	40.7	N/A
	Sound	66.3	75.1	44.4	N/A
3	Music	44.6	42.2	44.6	33.8
	Word	70.6	60.3	44.1	44.1
4	Music	44.1	46.9	46.9	33.8
	Word	58.2	63.4	56.8	60.6

Note: Percentages were calculated by summing the number of trials in which a memory was reported and dividing this by the total number of trials completed across all participants (computed separately for each cue category).