Farewell to Holbrook & Schindler’s (1989) “Song-Specific Age”? – Little Evidence for Lifelong Influence of Age-Specific Musical Preferences

Reinhard Kopiez1, Jan Weigang1, Friedrich Platz2 and Nina Düvel1

Abstract
As predicted by Holbrook and Schindler (1989) in a seminal study, popular music from the charts released when a person is roughly 23.47 years old (so-called Song-Specific Age, SSA) has a particularly positive impact on that person’s song evaluations decades later. In our replication study, \(N = 162\) participants (\(M_{\text{age}} = 59.1\) years, \(SD = 17.3\)) indicated their preferences for 18 song excerpts randomly selected from a corpus of 87 German Top 10 chart hits, released between 1930 and 2017. The fitting of a quadratic curve (as in the original study) to the aggregated ratings revealed a much earlier overall SSA peak at 14.16 years (\(R^2 = .184\)). The best approximation to the original SSA peak of 23.47 years was found only for the elderly subgroup of participants aged 50+ years with an SSA value of 22.63 years, however, with a relatively low goodness-of-fit (\(R^2 = .225\)). To summarize, the original finding of an SSA peak in the phase of early adulthood (23.47 years) could not be confirmed in terms of a point estimate. Instead, various subgroups showed various SSA peaks. The decomposition of ratings on the song level by latent profile analysis revealed four basic rating patterns (constantly high, constantly low, increasing, and decreasing over time) that might explain the different findings of the overall course of SSA regression curves within our subgroups without reference to the concept of SSA. Results are discussed in favor of current dynamic models of lifelong changes in musical preferences. We conclude that today, the SSA proposition – at least in its original form – seems to be of only limited validity for the explanation of musical preferences.

Keywords
Autobiographical memory, latent class analysis, MEAM, music, music preference, reminiscence bump, replication

Submission date: 15 June 2020; Acceptance date: 22 February 2021

Literature
Musical preferences vary with age and evolve constantly throughout adulthood (Bonneville-Roussy et al., 2013, 2017; Bonneville-Roussy & Rust, 2018). This development is influenced by social networks (Bonneville-Roussy & Rust, 2018) as well as by music-specific audio features, such as timbre, dynamics, and tonal clarity (Bonneville-Roussy & Ererola, 2018). For these reasons, Bonneville-Roussy et al. (2017) argued that there is little basis for the hypothesis that musical preferences crystallize in early adulthood and remain static over the next decades. Knowledge about the age-related development of musical preferences is of fundamental interest to those who study the social and developmental psychology of music (Greasley & Lamont, 2016; North & Hargreaves, 2008). For example, North and Hargreaves (2008, p. 106ff.) assumed that the reason few young people like classical music and jazz could be due to the high complexity of such music. Classical music requires knowledge of genre characteristics and

1 Hanover Music Lab, Hanover University of Music, Drama and Media, Hannover, Germany
2 University of Music and Performing Arts, Stuttgart, Germany

Corresponding author:
Reinhard Kopiez, Hanover Music Lab, Hanover University of Music, Drama and Media, Hannover 30175, Germany.
Email: reinhard.kopiez@hmtm-hannover.de

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
more listening experience, which have not been acquired by, but come, with increasing age.

The developmental processes for acquiring musical knowledge are often described according to Erikson’s (1950) model of primary life stages with the phases of adolescence (12–19 years), young adulthood (20–39 years), and middle adulthood (40–65 years). In particular, the phase of late adolescence and young adulthood in the mid-20s is significant for the development of musical preferences. Beyond the perspective of music psychology, in-depth knowledge about the influence of early exposure to cultural products, such as music, movies, or books, on the development of later preferences might also be of interest for market researchers seeking to predict consumer behavior.

Against this background, the economists Morris B. Holbrook and Robert M. Schindler published a highly influential study (Holbrook & Schindler, 1989): Starting from the “tentative hypothesis […] of whether popular music preferences peak at a certain age” (p. 120) and based on ratings of 108 respondents between 16 and 86 years of age (mean age: 54.3 years), participants indicated their liking of musical excerpts of 30 seconds from 28 popular songs published between 1932 and 1986 (selected from two-year intervals) on a 10-point scale. The target variable was the SSA, which was calculated as the difference between the year of song release minus the year of a person’s birth. Hence, if a pop song released in 1990 was rated by a person born in 1980, this resulted in a liking value for the SSA category of 10 years because the person was 10 years old when the song was released. As a consequence of the calculation of difference values, SSAs could also be negative, which denoted ratings for songs released years before a person was born.

The new approach of Holbrook and Schindler (1989) was not the identification of the developmental phase of a preference peak but “achieving a more precise estimate of that age” (Holbrook & Schindler, 1989, p. 120). The resulting peak of the inverted U-shaped quadratic regression fit was 23.47 years (based on the aggregated data of the 124 SSA categories), meaning that music listened to at this particular age peak was likely to become a lifelong preference (see Figure 1). This relationship between age and musical preferences is called by us the SSA proposition. Holbrook and Schindler held onto this idea in the next decades and in subsequent studies: The existence of an inverted U-shaped pattern of sensitivity has also been applied and confirmed by Holbrook and Schindler for a variety of products, such as movies (Holbrook & Schindler, 1996), movie stars (Holbrook & Schindler, 1994), automobile models (Schindler & Holbrook, 2003), and a wide range of products from other categories (Holbrook, 1995; see North & Hargreaves, 2008, p. 108). In other words, it must be possible to judge Holbrook and Schindler’s (1989) following strong claim: “Meanwhile, we appear to have identified […] a peak in the development of preferences for popular music that occurs in early adulthood (in the present case, at about the 24th year)” (p. 124).

Various possible explanations for the SSA peak during early adulthood have been offered by Holbrook and Schindler in their papers (for details see North & Hargreaves, 2008, p. 110f.). First, the process of early imprinting (with reference to the animal kingdom in which a “learning process that occurs during a critical period in the life some animals,” see VandenBos, 2015, p. 528) can play an important role. Second, learned associations between songs and “rites of passage” at this stage of life might be of relevance. Third, mechanisms of nostalgia (Janata et al., 2007) might play a role (i.e., the music concerned was more accessible via radio airplay when the participants were younger). Fourth, nostalgia proneness (“Things used to be better in the old days,” see Holbrook & Schindler, 1996) can have an influence. Fifth, the phase of early adulthood might also be a period when musical prototypes become finalized, and preferences for prototypes manifest themselves. Finally, with reference to popular chart music, mere exposure effects play a central role, for example, in causing implicit memory effects (Peretz et al., 1998; Szpunar et al., 2004). Such omnipresence is typical for mainstream popular music, and the relationship between the liking of cultural products and exposure frequency is the key element of Zajonc’s (1968) mere exposure theory, which has been confirmed for musical stimuli in more recent research (Johnston, 2016; Szpunar et al., 2004).

Our study, which seeks to verify the knowledge gained by Holbrook and Schindler, is set against the background of the ongoing discussion on replicability of findings in the field of psychology and social science (so-called “replication crisis” see Open Science Collaboration, 2015; Pashler & Wagenmaker, 2012). The first replication attempt of the SSA proposition was conducted by Hemming (2013): Based on song ratings of a sample of 25 songs
(German Top 10 hits, released between 1960 and 2008 and selected in two-year intervals) made by 473 respondents on a 10-point liking scale, he found a regression peak at 13.47 years for the disaggregated and 8.59 years for the aggregated (averaged) SSA data (see Tables S4 and S6 in the Supplemental Material section). This finding did not confirm Holbrook and Schindler’s (1998) observed age peak of 23.47 years for the aggregated data (see also our own re-analyses of Hemming’s data in Table S6 in the Supplemental Material section). Hemming also applied data trimming by the seemingly arbitrary decision to exclude SSA categories with less than 50 ratings. This filtering criterion was used to avoid the inclusion of inappropriate weights (e.g., ratings of new songs by a small group of people at the negative and positive ends of the distribution) in the regression analysis.

The result of the quadratic regression fit to the trimmed and aggregated SSA values of Hemming’s (2013) data resulted in an age peak of 17.36 years, which was still far off the 23.47 years of the original study. Despite these obvious differences between the regression peak observed in Holbrook and Schindler’s (1989) original study and the replication attempt by Hemming (2013), in their commentary on Hemming’s replication, Holbrook and Schindler (2013) tried to confirm their previous finding. Using Hemming’s (2013) data set and after manually removing unreliable observations (elimination of SSA categories with very small sample sizes), Holbrook and Schindler (2013) found a cubic regression peak at 24.72 years ($R^2 = .657$), which was regarded as confirmation of the original finding.

Indeed, Hemming’s (2013) replication study showed some conceptual and analytical weaknesses: First, as mentioned by Holbrook and Schindler (2013), small sample sizes can cause a problem at the end tails of the SSA distribution, which can only be solved by “collecting big enough numbers of observations from people at the SSA extremes” (p. 308). In particular, in the case of a sample of a relatively young age, the proportion of people older than 70 years will be small. Although Hemming’s (2013) study comprised $N = 473$ participants, the average sample age was low ($M = 33.25$ years, $SD = 17.40$). Second, Hemming (2013) used only very basic methods of data analysis limited to quadratic regression fit in combination with a restrictive trimming method of SSA categories with less than 50 responses. Third, the selected music examples were limited to Top 10 hits from 1960 to 2008 (range: 48 years) in two-year intervals, while Holbrook and Schindler’s (1989) selection of musical stimuli covered a wider time span from 1932 to 1986 (range: 54 years). For the Hemming sample, this meant, for example, that participants aged 60 years (born in 1948) did not evaluate musical excerpts released during the first 12 years of their lives.

The most recent attempt to reveal the relationship between age and lifelong musical preferences was conducted by Stephens-Davidowitz (2018). Based on the ranking of more than 18,000 counts of selected songs from the platform Spotify, the author investigated to what extent the year we were born influences how frequently we listen to a particular song. Stephens-Davidowitz (2018) found that the strongest adult musical preferences set in at the age of 13 for women and 14 for men. In other words, the year we were born influences the music we like to listen to. The key years for musical preferences seemed to match those associated with puberty. However, the ranking peaks of 13 and 14 years are not congruent with the proposition of SSA. Other problematic aspects of the analysis were that rankings based on average listening behaviors were used instead of preference ratings (songs not included to the, 4,000 available ranks were all assigned to rank 4,001), only 631 out of 18,327 (= 3.4%) entries came from users aged 70+, and the number of users remained unknown. In other words, the frequency analysis represented songs familiar to the mainstream of Spotify users, but the analysis did not show true preference ratings.

Holbrook and Schindler’s predicted inverted U-shaped course of musical preference development has also been the subject of some controversy. First objections come from research on age-related changes in preferences for musical genres: In their comprehensive study, Bonneville-Roussy et al. (2013) summarized a variety of developmental influences on musical preferences throughout the lifespan. They found, for example, that the preference for rock and hip-hop music decreased with increasing age, while the preference for classical or country music increased. In addition, age trends in musical preferences were also associated with personality types. For example, the personality factor of openness was associated with a preference for classical and jazz music from early adulthood to middle adulthood. In their Music Preference in Adulthood Model, Bonneville-Roussy et al. (2017) assumed that musical preferences vary with age and continuously evolve during adulthood.

This critique also applies to similar static concepts of lifetime distributions of favorite cultural products: Janssen et al. (2007) suggested that books, movies, and records that become favorites among users between 20 and 26 years (peaks: books = 26.06 years, movies = 25.09 years, records = 24.48 years) can cause a reminiscence bump and are privileged through the subsequent decades. A second objection to the validity of a single regression peak for the relationship between popular songs and personal memories comes from autobiographical research. Platz et al. (2015) found that music-related autobiographical memories (so-called MEAMs) did not seem to be limited to the developmental phase between 15 and 24 years, but popular songs could be associated with autobiographical memory over many decades of life up to the sixth decade. A third objection to the assumption of a uni-modal distribution comes from the model of Cascading Reminiscence Bumps by Krumhansl and Zupnick (2013): Based on song sections from the Billboard charts between 1955 and 2009, participants ($M_{\text{age}} = 20.1$ years) reported their song-related
memories and liking for the songs. Surprisingly, ratings were high for music from the early 1980s (when their parents were aged about 20 years) and increased again for music from the participants’ own age (about 20 years). In other words, musical preferences seemed to be influenced by popular chart songs not only from the participants’ own early adolescence but also from the same phase in their parents’ lives. This pattern of an early peak and a late one of preferred and remembered music was regarded by Krumhansl and Zupnick (2013) as a process of cultural transmission over generations and denoted as cascading reminiscence bumps, and it could be confirmed in a later study by Jakubowski et al. (2020). However, this pattern of a double reminiscence bump could not be confirmed in a more recent study by Spivack (2019) investigating the recognition rates for 152 songs from Billboard lists (released between 1940 and 2015) by so-called millennials (participants with median age of 20 years): Recognition rates were highest for music from the current millennium, declined to a stable plateau for songs from 1960s to the 1990s, and showed a further decrease with a gradual drop-off for the 1940s and 1950s.

Finally, our investigation goes beyond the limits of a technical replication of a single study and considers the perspectives for other overlapping fields, such as music-related autobiographical memories, music preference research, and studies on memory for cultural products. For example, while looking for music-evoked autobiographical reminiscence bump effects, Platz et al. (2015) only found a small effect of an increased memory for popular songs released in the SSA range between 15 and 24 years, and he could not identify a clear SSA peak for liking. In a more recent study by Jakubowski et al. (2020), participants evaluated songs from a list of 111 titles (French charts released between 1950 and 2015) for autobiographical salience, familiarity, and liking. Although no pronounced peak in liking could be observed for all age groups, the oldest group of participants (aged 56–82 years) showed a maximum for the broad range of SSA categories from 5 to 19 years.

Another argument for the non-static application of the SSA concept comes from Rathbone et al. (2017), who found evidence that the reminiscence for music from the past is moderated by personal significance of a song. Based on a list of song titles (56 most successful songs from the UK music charts released between 1950 and 2005), participants selected the five most significant songs. The overall distribution of SSAs (called Age at Release by the authors) revealed an SSA peak between 15 and 19 years (Study 1). However, when responses were elicited by the request for particular song-related memories (Study 2), the SSA peak for songs with personally significant song release dates moved to the early teenage years (10 to 14 years). Songs that were personally non-significant did not show a clear SSA peak in recognition rates. The study by Loveday et al. (2020) joins this line of research on the role of personally relevant music during the developmental phase of adolescence and early adulthood (10–30 years): Based on self-selected songs suggested by guests of a long-running radio show, it was found that 50% of selected songs had a personal significance (links to memories of a person) released during the ages of 10 to 30 with a peak between 10 to 20 years. The influence of individual variables such as sex or musical preference for genres on the location of the age peak for song preference was observed by Zimprich (2020): In a sample of persons aged 70–75 years, a maximum point (memory bump) for the recognition of popular songs from 1945 to 1995 was found at an SSA of about 17–19 years. To summarize, all these studies showed a high diversity of age peaks covering a wide age range (an observation which is supported by the systematic review by Munawar et al., 2018).1

**Research Aims**

The main aim of this study was the conceptual (not verbal) replication of Holbrook and Schindler’s (1989) as well as Hemming’s (2013) studies on the relationship between a person’s song preference and his or her age. This relationship between age and preference is represented by so-called SSA. Thus, the overall distribution of all SSA values was fitted by quadratic and cubic regression fits for the unambiguous identification of SSA peaks. To reach this aim, first, a data set needed to be acquired that covered the preference ratings of a wide age range of participants. In particular, it was important to include a sufficient number of elderly participants (between 60 and 80 years) so that subgroup analyses, such as those filtering out of SSA categories with too little responses, could be conducted. Second, we sought to identify and compare the identified peaks of regression fits to the peak value of 23.47 years as suggested for the aggregated SSA values in the original study by Holbrook and Schindler (1989) and 17.36 years as suggested by Hemming (2013). Third, based on more sophisticated statistical methods, we aimed to separate the influences of interleaved variables on the SSA value such as the participant’s age, the age of the song, and the particular song.

**Method**

We conducted an online survey study in which participants had to rate their liking for song excerpts of 20 seconds each from the domain of popular music charts by means of a continuous visual analogue scale (VAS; slider scale). The selection of musical materials and the procedure will be described in the following paragraphs.

**Materials**

**Selection of Audio Examples.** The audio examples and selection procedures used in this study were the same as in Platz et al. (2015) but were updated to include songs from 2011.
to 2017. This resulted in 87 songs (released between 1930 and 2017; for additional information on the selection criteria and a complete list of songs, see Table S1 in the Supplemental Material section). For the German Number One chart hits from 1930 to 2001, an important source was the anthology by Ehnet (2002), complemented for the years 2002 to 2017 by additional mentions in the charts.

The excerpts from the songs had a length of 20 s, starting 10 s before the beginning of the first chorus and including 1 s for fade-in and fade-out (Kopiez et al., 2015). The length of 20 s for musical examples was within the range of previous studies (see Altemmüller et al., 2014; Basso et al., 2004; Bigand et al., 2005; Cady et al., 2008; Janata, 2009) as well as the original study by Holbrook and Schindler (1989) and the replication study by Hemming (2013), which was 30 s. The duration of 20 s selected for this study corresponds to the time length of < 30 s used most frequently in the field of music and emotion for stimuli from the genres of film music, popular music, and jazz music although excerpts from the genre of Western Art Music show a wider range of durations (48% < 30 s and 26% 30–60 s, see Warrenburg, 2020). To counteract an increasing loudness prevalent in popular music (the so-called “loudness race”, see Deruty & Tardieu, 2014), and the resulting psycho-acoustical effects, we normalized all songs with Root Mean Square (RMS) by means of the software Audacity and the plugin RMS Normalize (https://forum.audacity-team.org/viewtopic.php?t=95421). This resulted in the same average sound pressure level for each song.

Control for Cognitive Impairment. To control for potential cognitive impairment of elderly participants in the sample and hence an influence on answer reliability, we implemented a short test for dementia: the “three-item word memory” of the “Mini-Cog” test (Borson et al., 2003). Based on the word lists of the Mini-Cog test (see https://mini-cog .com), the following three German items were presented for the memory test: Dorf, Küche, and Baby (village, kitchen, and baby in English).

Participants

Owing to the survey character of the study, no a priori power analysis was conducted. However, the sample size of $N = 108$ from the original study by Holbrook and Schindler (1989) was used as a minimum benchmark. Participants for the online survey were recruited by a call for participation circulated in the newsgroups and social media of local associations of older people in the region of Hanover. The personal data collection was conducted with participants from retirement centers in the region of Hanover, Germany.

On average, the resulting $N = 162$ participants had a mean age of 59.1 years (median = 59 years, $SD = 17.3$ years, range [14; 94] years; male = 35.8%, female = 64.2%, < 50 years: 22.2% [$n = 36$], ≥ 50 years: 77.8% [$n = 126$], ≥ 70 years: 30.2% [$n = 49$]; for details, see Table 1 and Figure S1 in the Supplemental Material section for the overall age distribution). The oldest participant was born in 1925, the youngest in 2005. A hearing aid was used by 5.5% ($n = 9$). For comparison, the sample of the Holbrook and Schindler (1989) study comprised $N = 108$ participants with a mean age of 54.3 years (range [16; 86] years, < 50 years: 43%, ≥ 50 years: 57%), meaning that our sample had a higher proportion of participants aged 50 years and more. For later analyses, we should also bear in mind that the sample used in the replication study by Hemming (2013) also differed in some important aspects: his $N = 473$ participants had a mean age of 33.3 years ($SD = 17.4$, range [6; 86] years), which was much younger compared with the current study and that by Holbrook and Schindler (1989) and differed in age distribution (see Figure S2 in the Supplemental Material section). In addition, Hemming (2013) used only 25 song sections covering the charts between 1960 and 2008 in two-year steps.

Sociodemographic Data. As a supplement to demographic standards (i.e., age and gender), information on the level of education was requested according to the International Standard Classification of Education (2012). Most participants (44.9%) had completed apprenticeships and vocational training (ISCED-2011, Levels 3 and 4) or university degree programs (38.3%, ISCED-2011, Levels 6–8). A minority (14.8%) held a Level 1 or 2 degree. This educational path is not unusual for survey participants from the war generation who often did not have the opportunity for subsequent vocational training and higher education after leaving school (see Table S2 and Figure S3 in the Supplemental Material section for details).

Study Design and Procedure

Study Design. To reduce the number of ratings of songs released before the participant was born (so-called negative

| Table 1. Descriptive data of the final sample in the current study ($N = 162$). |
|---|
| **$M_{age}$** | **Range$_{age}$** | **$SD_{age}$** | **$n$ (%)** |
| (a) Online study |  |  |  |
| Male | 56.4 | 23 to 94 | 17.2 | 53 (32.7) |
| Female | 54.5 | 25 to 85 | 13.8 | 75 (46.3) |
| Total | 55.3 | 23 to 94 | 15.3 | 128 (79.0) |
| (b) Personal presence (with or without assistance) |  |  |  |
| Male | 48.2 | 14 to 77 | 31.6 | 5 (3.1) |
| Female | 77.8 | 48 to 93 | 7.9 | 29 (17.9) |
| Total | 73.4 | 14 to 93 | 16.9 | 34 (21.0) |
| (c) Overall total (online study and personal presence) |  |  |  |
| Male | 55.7 | 14 to 94 | 18.6 | 58 (35.8) |
| Female | 60.7 | 25 to 93 | 16.3 | 104 (64.2) |
| Total | 59.1 | 14 to 49 | 17.3 | 162 (100) |
SSA values – see results section and Hemming, 2013; Holbrook & Schindler, 1989), and because of our interest in older adults evaluations of songs from their entire lifespan, a particular focus was on the acquisition of participants aged 50 years and more. Owing to the excessive time required in the case of a presentation of all songs (87 song examples, of 20 s duration each, $\times$ 0.5 minutes response time $\approx$ about 70 minutes), we decided to use an incomplete, randomized, blockwise design based on the randomized selection of 18 songs rated by each person (two songs drawn randomly from nine decades between 1930 and 2017). The maximum time required for participation in the survey was not to exceed 10 minutes, which was deemed an acceptable time span for the attentional span of the targeted population of elderly people. We considered all recommendations for internet experiments, such as recording of participation time, control for stability of internet connections and loading time, and reduction of drop-out risk by use of short participation time as suggested by Reips (2002, 2012), and we pretested the survey script on various tablets, computers and browsers.

Procedure

The survey was conducted in 2018/19 in two variations, both using the online research platform SoSciSurvey (https://www.soscisurvey.de). This survey was spread via social media and local public outlets for the targeted age group. In addition, the survey was conducted personally in local senior residences and recreation centers for the elderly using a hotspot connection to the same online platform. Participants were given a touchscreen laptop and headphones and could choose to be guided through the survey with extra instructions, which were scripted and included all instructions (for the visually impaired) as well as general guidance for people with little experience in the handling of computers. All participants received individualized surveys, and if they decided to use the extra instructions, this was noted as a control variable. However, guided assistance was demanded only by a minority of participants ($n = 11, 6.8\%$). For the personal interviews, participants started from a medium volume level, and loudness could be adjusted stepwise to a comfortable level.

The study was performed in accordance with relevant institutional and national guidelines (Deutsche Gesellschaft für Psychologie, 2016; Hanover University of Music, 2017) and with the principles expressed in the Declaration of Helsinki. Formal approval of the study by the Ethics Committee of the Hanover University of Music, Drama and Media was not mandatory as the study adhered to all required regulations. All participants gave their informed consent online by ticking a checkbox.

The entire procedure was as follows (see Figure S4 in the Supplemental Material section): After the welcome, the listening experiment started with a short introduction text about the study aim:

In this survey, you will be asked how much you like the short examples of music. You will use a continuous scale with a slider to indicate your personal evaluation of a song. In total, you will listen to 18 songs, followed by some questions regarding your personal background.

Next, participants gave informed consent by ticking a checkbox, followed by the “Mini-Cog” three-item word memory test (Borson et al., 2003). Participants were instructed to remember the three target words for later recall as mentioned previously. This was followed by a short explanation of the VAS and, in a first practice trial, participants had to move the slider to the extreme right position (see Figure S5 and S6 in the Supplemental Material section). In the second and third practice trial for the handling of the VAS, the slider had to be moved to two different target positions (see Figure S6 in the Supplemental Material section). Slider positions for all three practice trials were recorded for later filtering of participants. In a next step, participants were asked to enter all remembered words from the “Mini-Cog” memory test to a text field, followed by the adjustment of sound to a comfortable level by the example of a musical example (Hello by Lionel Richie from 1983). At this point, the main part of the survey started with the recording of subjective ratings for each of the 18 randomized musical samples. Finally, participants were asked for sociodemographic data (age, gender, level of education). The participation time for the survey was about 10 minutes for the online participation, 15 minutes for the physical presence condition (without assistance), and 20 minutes for the physical presence with assistance condition. No reimbursement was paid.

Results

Data Preparation

Data preparation was conducted stepwise: In total, the website was accessed by 291 and completed by 198 participants (see Figure S7 in the Supplemental Material section). The first criterion for data filtering was the correct handling of the VAS slider in the trial section. Only participants with at least two correctly positioned slider trials were considered. Incorrect performance of the targeting task could have been caused by limitations in fine motor skills, too small screens, or a misunderstanding of the VAS. The second filter criterion was the performance on the “Mini-Cog” word memory test: To exclude participants with mental impairment (which might have a negative influence on the remembering of songs), cases were only considered if at least two out of three items were remembered correctly.

In three cases, the participants either accidentally continued the slides before having time to learn the three words or were distracted by an outside interference, prolonging the time between learning and repeating, thus putting them in a worse position to pass the Mini-Cog. In these three
cases, the criterion for inclusion was lowered to one out of three. The resulting overall distribution of the “Mini-Cog” performances was as follows: only one item correct: 1.9% \((n = 3)\), only two items correct: 8.0% \((n = 13)\), all three items correct: 90.1% \((n = 146)\). Finally, the sample, comprising \(N = 162\) participants, produced \(N = 2,916\) valid SSA values, which was close to the \(N = 3,024\) valid rating values in the study by Holbrook and Schindler (1989) but lower than the \(N = 11,825\) SSA values in the study by Hemming (2013). In a last step and in line with the procedure used by Holbrook and Schindler (1989), preference ratings were standardized \((z\) transformed) within participants (across a participant’s rating of the 18 examples). These arrays can be referred to by following a cell reference with “#”, which precludes having to know the end of a range. This proves valuable for filtered data sets that vary in length as it is possible to calculate regressions and local maxima of cubic functions numerically. All calculations take place in one macro-enabled workbook (“.xlsm”) in combination with multiple worksheets (see the Supplemental Material section for the Excel workbook used for this data analysis). For ease of use, the workbook analyzes the current data set. Except for this data set, no values needed to be changed in any other worksheet. Additional parameter settings for the data analysis (e.g., data trimming) can be made via control buttons and dialog boxes.

**Choice of the Analysis Software**

For the data analysis, Microsoft Excel (as part of the Microsoft 365 Insider Program) was used. Statistic programs such as Jasp (JASP Team, 2020) and Jamovi (The jamovi project, 2020) have made great strides in advancing highly capable and visually appealing open-source solutions to statistical data analyses. However, they are still not able to integrate polynomial regressions into scatterplots. Moreover, as with SPSS, these programs do not offer the versatility in dynamic filtering that Excel offers with the use of researcher-developed macros. This dynamic and flexibility in data analysis enable the creation of the entire analysis before data are entered (from this or a different study), followed by the instantaneous visualization of results as well as the option of flexible data trimming. While similar workflows with the data sets might be possible in R, using Excel was the most sensible choice as it is widespread and met our expertise with the required processes of macro programming.

A particular feature added to the “Office Insider” version of Excel in 2019 is the “Dynamic Arrays” function. These arrays can be referred to by following a cell reference with “#”, which precludes having to know the end of a range. This proves valuable for filtered data sets that vary in length as it is possible to calculate regressions and local maxima of cubic functions numerically. All calculations take place in one macro-enabled workbook (“.xlsm”) in combination with multiple worksheets (see the Supplemental Material section for the Excel workbook used for this data analysis). For ease of use, the workbook analyzes the current data set. Except for this data set, no values needed to be changed in any other worksheet. Additional parameter settings for the data analysis (e.g., data trimming) can be made via control buttons and dialog boxes.

**Analysis of General Rating Behavior**

**Distribution of SSA Values.** The resulting \(N = 2,916\) SSAs from the 162 participants covered a wide range from -74 to 88 years (median = 15.0 years, see Figure 2), which was larger compared with the age range of Holbrook and Schindler’s (1989, SSA range = -39 to 85 yrs) and Hemming’s sample (2013, SSA range = -43 to 86 yrs). The skewness of the data distribution (as a measure of asymmetry) is -0.087, which can be regarded as hardly any skewness, and the kurtosis (as a measure of the “pointiness” of a distribution) was -0.64, meaning that the pointiness was flatter (when compared with the value of 0 for a normal distribution see Navarro & Foxcroft, 2019, p. 78). This characteristic of the distribution of SSA values was confirmed by a significant deviation from normal distribution (Shapiro-Wilks \(W = 0.99, p < .001\)). However, SSA categories up to the age of about 70 showed a sufficient number of counts (about \(n = 10\)) for further subgroup analyses of data.

To test for the influence of the particular data collection method (personal vs. internet), we conducted a cubic
regression fit for the disaggregated SSA values (see Figure 3). Although the curve fit for the internet-based procedure showed a marginally higher fit compared with the personal data collection method (internet: $R^2 = 0.0414$, personal: $R^2 = 0.0711$), both curves were nearly identical between the SSA range $[0; 88]$. Thus, we used the pooled data set from both sampling procedures for all further analyses.

### Options for Data Analysis

As there are numerous options for data analysis applied in previous studies, our analysis was not limited to one method. Table 2 gives an overview of the methods (and combinations) of data analysis procedures applied in previous studies on SSA – of particular interest are the data aggregation, method of regression fit, and data trimming (exclusion of cases). In our analysis, these options were extended. In addition, we considered the handling of outliers and analyses by age subgroups such as participants older than 50 or 70 years.

| Sample selection | Data aggregation | Criteria for data trimming (exclusion) | Method of regression fit | Reference |
|------------------|------------------|----------------------------------------|--------------------------|-----------|
| Total            | Yes              | –                                      | Quadratic                | Holbrook and Schindler (1989) |
| Total            | No               | –                                      | Quadratic                | Hemming (2013) |
| Total            | Yes              | < 50 SSA values                        | Quadratic                | Holbrook and Schindler (2013) |
| Total            | No               | < 50 SSA values                        | Quadratic                | Holbrook and Schindler (2013) |
| Total            | Yes              | Manually (removal of very small counts in SSA categories) | Cubic                   | Holbrook and Schindler (2013) |

### Disaggregated vs. Aggregated Data

In general, there are two options for the handling of data: first, the use of disaggregated $z$-transformed (raw) SSA ratings, and, second, the use of (also $z$-transformed) aggregated values (average of all ratings within one SSA category). Holbrook and Schindler (1989) as well as other authors (see Table 2) used both options for the calculation of the target variable “music preference,” but both methods differ in the resulting strength of the relationship between SSA and music preference. Owing to a larger dispersion, disaggregated preference ratings of individual cases showed a weaker regression fit (Holbrook & Schindler, 1989, p. 122). Thus, in line with Holbrook and Schindler...
(2013) and Hemming (2013), we applied both methods to the various steps of data analysis, but we emphasized the use of aggregated data for regression fit.

**Minimum Number of Responses per SSA Category**

In general, the trimming of data is recommended in the case of an unusual low number of cases in the tails of a distribution or outlier values. As emphasized by Holbrook and Schindler (2013), outliers of SSA values at the extreme ends of the SSA scale can have a severe influence on the quality of the regression fit and the identification of the preference peak. Thus, Holbrook and Schindler (2013) suggested the removal of “unreliable observations from the data array” (p. 305) or the application of “other plausible cut-off criteria” (e.g., very small sample sizes in SSA responses). However, the question remained as to how many responses should represent an SSA category. Holbrook and Schindler (1996) suggested an iterative procedure for data trimming: first, analyses should be run with cut-off values from \( n \geq 1 \) to \( n \geq 50 \) on the basis of disaggregated data. Second, a cut-off value should be selected at which the regression fit becomes stable but without raising the cut-off value too high to avoid unnecessary restrictions for the available SSA range. Third, a final analysis on the basis of aggregated data with the parameters of the best fitting model derived from the disaggregated regression fit should be conducted. In Holbrook and Schindler’s (1996) study, the best curve fit was obtained after the exclusion of SSA categories with < 10 observations. Our inspection of Holbrook and Schindler’s (2013) manual data trimming strategies applied to Hemming’s (2013) data revealed that SSA categories of -43, -42, -41, and -40 were excluded due to too low observations (between \( n = 2 \) and \( n = 5 \)) and the SSA categories of 75, 77, 79, 81, 83, 84, and 85 due to observations between \( n = 1 \) and \( n = 3 \).

Hemming (2013) made the arbitrary decision to remove all SSAs with less than 50 responses. As a consequence, the number of the remaining SSA categories to be considered for data analysis was reduced from 129 SSAs to only 78 (ranging from -28 to 49 years). Although it makes sense to reduce the data set for unreliable or underrepresented SSA categories, the reduction of SSAs by minimum numbers of observations should not be too conservative (increasing the risk of masking the assumed non-monotonicity in variable relationships) and should also apply to combinations of trimming criteria.

**Data Trimming (Handling of Outliers)**

The handling of outliers is a standard procedure in dealing with non-normal distribution, unequal variances, and extreme values (Kirk, 2013, p. 108). For the exclusion of outliers, the criterion of \( z \leq -2.5 \) and \( z \geq 2.5 \) is recommended (Kirk, 2013, p. 131) because these values represent only the upper and lower 0.62% of the normal distribution and therefore together have a probability of \( p = .0124 \). As a rule of thumb, it is recommended that trimming should reduce the data between 15% and 25% (Kirk, 2013, p. 108f.). The removal of “unreliable observations” (Holbrook & Schindler, 2013, p. 305) is also in line with procedures for data analysis suggested in other research in the field of SSA (Holbrook & Schindler, 1996).

**Quadratic vs. Cubic Regression Fit**

Although the original study by Holbrook and Schindler (1989) used only a quadratic regression fit, in their response to Hemming’s (2013) replication study, Holbrook and Schindler (2013) suggested the additional use of a cubic fit because the fit for the often skewed distribution of SSAs improves significantly with a cubic regression equation (Holbrook & Schindler, 2013, p. 306). The application of quadratic as well as of cubic regression fit seems to be a standard procedure in the modeling of consumer behavior in marketing research (Holbrook & Schindler, 1996). However, as there is no theoretical framework for the interpretation of cubic regression fit to SSA values, we decided to exclude this option from our study. We leave the use of this data filtering option up to the reader (see the Excel workbook in the Supplemental Material section).

**Analyses for Different Age Groups**

To control for the influence of generation and cohort effects, analyses were also conducted for the groups of 50+ and 70+ year olds. Although this subgroup division was expected to diminish the sample size, it had the longest exposure to the majority of songs from our selection and would thus reduce the proportion of negative SSA values.

**Resulting Regression Peaks**

For the identification of SSA peaks, we will now go through the various options for regression fit and data trimming. Finally, results will be summarized in comprehensive tables at the end of this section.

**Overall Peaks for (Dis)Aggregated Data**

In a first approach, the correlation between the preference ratings and the disaggregated SSAs was calculated for quadratic regression fit (see Figure 4). However, the resulting model fit from the regression curve for the disaggregated SSAs was small in terms of effect size benchmarks (Ellis, 2010, p. 41) quadratic fit: 32.18 years (\( R^2 = .0243 \)). The peak of 32.18 years also differed considerably from the peak of 23.47 years for the aggregated and 23.66 years for the quadratic fit of the disaggregated data found by Holbrook and Schindler (1989) as well as from the peak of 8.59 years (\( R^2 = .454 \)) for the aggregated and 13.47 years (\( R^2 = .025 \)) for the quadratic fit of the disaggregated data found by Hemming (2013). To stay in line with the original study by Holbrook and Schindler (1989) and most of the analyses conducted by Hemming (2013), our data analyses, as outlined in the following sections, were based on the
aggregated SSAs. However, the final comparison of regression peaks based on the most relevant combinations of additional filtering criteria such as aggregated, disaggregated, trimmed, and age-grouped data analyses is given in Table S4 in the Supplemental Material section.

Peaks for Data Trimmed by Minimum Number of Responses

In a first approach, only those SSAs with a minimum number of five SSA ratings were included. This resulted in a data trimming by -16% (from 152 to 128 valid SSA categories), which is within the trimming rate of -15% to -25% recommended by Kirk (2013, p. 108f.). When we trimmed the data by exclusion of SSA categories with less than 10 ratings, this resulted in a trimming rate of -29% (from 152 to 108 valid SSA categories). This conservative approach would have weakened the data basis for regression fit in our limited sample. More extreme trimming methods such as a minimum number of 50 ratings for a valid SSA category as applied by Hemming (2013) would have resulted in a significant reduction of SSA categories (in this case -39%; from 128 to 78 valid SSAs). Thus, restrictive trimming methods require a very large sample size with a large number of ratings for each SSA category. Figure 5 shows the resulting peaks: In the case of a minimum of five ratings (upper graphs), the peak for the regression fit was 27.07 years ($R^2 = .116$). Although the peak of fit was close to the peak of 23.47 years as suggested by Holbrook and Schindler (1989), the quality of the curve fit remained weak.
In the case of a more strict trimming method with a criterion of at least 10 ratings for a valid SSA category (see Figure 5, lower graphs), this resulted in a peak at 29.36 years ($R^2 = .379$) for the regression fit. As a result of this higher threshold for inclusion of data (from a minimum 5 to 10 ratings per SSA category), the effect sizes for the regression coefficients increased from small to large for the quadratic fit.

Peaks for Data Trimmed by Exclusion of Outliers

For this trimming of outliers, the method of exclusion by means of a threshold for extreme z-values was applied (see Kirk, 2013). In our case, SSA ratings with z-values smaller than -2.5 or larger than 2.5 were not considered for data analysis (Figure 6). The resulting data set was not further reduced. The regression peak for the quadratic regression was 14.16 years ($R^2 = .184$) and for the cubic regression 32.48 years ($R^2 = .285$). Both regression methods represented medium to large effect sizes (see Ellis, 2010, p. 41), but the observed peaks differed considerably from the original finding of 23.47 years.

Peak for Data of Different Age Groups

An additional analysis was for the age groups of 50+ and 70+ year olds. This allowed us to control for cohort and generation effects. Figure 7 shows the resulting regression fits and distributions of SSAs. The data sets for the two age groups were reduced by -19% (from 152 to 124 valid SSA categories) for the 50+ group and by -31% (from 152 to 105 valid SSA categories) for the 70+ group. For the 50+ group a quadratic regression peak of 22.63 years ($R^2 = .225$) was found. For the 70+ group, a quadratic regression peak of 32.37 years ($R^2 = .471$) could be identified. Regressions represented good fits with a nearly large effect size (group 50+) or a large effect size beyond the benchmark (see Ellis, 2010, p. 41). Except for the peak of the curve fit for the 50+ group (22.63 years), the other peak differed from the original finding of 23.47 years. We should bear in mind that the peak of the quadratic fit for the 50+ group showed the smallest difference (0.84 years) from the original finding, but its regression coefficient of $R^2 = .225$ did not reach the original curve fit coefficient of $R^2 = .706$ as observed by Holbrook and Schindler (1989).

Distribution of Peaks Resulting from Various Analysis Options

In the next step of data analysis, we compiled the various findings from different selection and trimming methods applied to the current data set into one big picture. For this purpose, all identified results from the application of different data trimming methods and subgroup analyses as well as their coefficients were summarized into Table S4.
The most interesting column is the “SSA Peak” column. On the basis of 14 peaks (aggregated data), the distribution of SSA peak values is shown in Figure 8a. The distribution covers a wide range of SSA peak values [14.16; 32.37 years] with a mean of 26.73 years.

The same procedure of SSA data aggregation was applied to the regression peaks obtained from Hemming’s (2013) replication study. On the basis of valid 10 peaks (see Table S6 in the Supplemental Material section), the distribution of SSA peak values is shown in Figure 8b. The distribution covers a wide range of SSA peak values [8.28; 20.83 years] with a mean of 12.40 years. The average peak values between the current study and Hemming’s peak values differ significantly ($t = 6.2$, $df = 22$, $p < .001$, $d = 2.57$, 95% CI [1.44, 3.66]).

**Multilevel Regression Analysis**

Owing to the nested structure of our data (18 out of 87 songs were repeatedly rated by each participant), we conducted several cross-classified multilevel regression analyses. On the lowest within level (I), the raw preference rating score served as the target variable and the SSA as linear and quadratic predictor variable. The age and year of the song release (being in the Top 10 German charts) were selected as the higher level (II) between predictor variables. By using cross-classified multilevel regression analyses, we aimed to decompose the nested structure of our sample’s response data in order to determine and quantify the potential dependencies of the hierarchical structure on the intercept of the target variable (preference rating) and the slope of the predictor variables (SSA and SSA$^2$).
Furthermore, our aim was to identify the impact of both background variables, participant’s age (the difference between the year of data collection and year of birth) and the year of song release (year in which the song was ranked in the Top 10 German charts). According to Holbrook and Schindler (1989), both variables should be able to significantly contribute to the prediction of an overwhelming amount of systematic variance within the distribution of preference ratings. In addition, we aimed to verify the latent premise of the SSA, namely, that the proportions of variance explained by both variables should be approximately equally distributed so that they can be merged equally into the new variable by subtraction from each other. For the cross-classified regression model analyses, we used a Bayesian approach along with non-informative priors as implemented in the software Mplus (V 8.3, see Muthén & Muthén, 1998–2017).

In a first step, we conducted an intercept-only model (see Model no. 1 in Table S7 in the Supplemental Material section) with no predictor variables on the lowest within level but with the variables “person” and “song” as Level-II background clusters. We found an estimated total variance

**Figure 7a to d.** Relationship between SSA and musical preference for participants of different age groups of age 50+ (A and B) and 70+ (C and D). A and C: regression curve fits (quadratic fit), B and D: distributions of SSA values. Hatched areas highlight the included SSAs. Both model fits differ significantly ($z = -4.09, p < .001$; see Table S3 for details). For standard errors of the regression coefficients, see Table S5 in the Supplemental Material section.

**Figure 8a and b.** Histograms of SSA regression peaks resulting from various methods of data trimming (aggregated data). A: data from the current study ($N = 14$ peaks, $M_{SSA} = 26.73$ years), B: data from Hemming’s (2013) study ($N = 10$ peaks, $M_{SSA} = 12.40$ years).
of 873.99 (random part) and an intraclass correlation for the person (between) level of 22.8% (199.19/873.99) and for the variable “song” as second nested level-II cluster of only 8.9% (77.7/873.99). Together, both cluster variables accounted for 31.7% of the estimated variance of participants’ preference ratings. With respect to the SSA, the unequal ratio between the proportion of both background variables for the estimated variance could therefore result in a biased estimator that might thus be stronger driven by the heterogeneity of the participants than by the songs.

Although the introduction of the variable “SSA” in Model 2 resulted in the inclusion of a significant low-level predictor, only a marginal model improvement was obtained compared with the intercept-only model (no. 1). Furthermore, the variables “age” as well as “year of song release” as between-level predictors did not show any significant model improvement (Model 3). Their non-significant influence might indicate that neither song year nor year of birth represented important characteristics of the nested structure of music and persons as between-level cluster variables in a linear regression. A small improvement could be observed when the SSA was modified into a random factor, varying in its strength across both cluster variables (Model no. 4 and no. 5). Although increasing the flexibility of the predictor variable’s slope led to a strong model improvement (\(D_{DIC,M_5} = 344.9\)), the proportion of expected variance with respect to the total variance (< 0.02%) was also of less importance and might thus be negligible. Compared to Model 2, the introduction of the variable “SSA” as an additional quadratic term (Model no. 6) resulted only in a marginal model improvement (\(D_{DIC,M_6} = 47.91\)), which could be further increased by introducing both predictors (SSA and SSA²) as random factors allowing for variation in the slope across persons and songs (\(D_{DIC,M_7} = 338.23\)).

In summary, we would have expected to reveal a much higher proportion of the expected variance due to heterogeneity in participants and songs compared with the amount of variance explained by the SSA as revealed by Holbrook and Schindler (1989). With regard to their operationalization of the SSA by subtracting the year of song release from the year of birth, we would also have expected a much more balanced ratio between both decomposed variance proportions.

**Data Analysis by Latent Profile Analysis**

The previous analyses did not result in a consistent picture, and none of the SSA regression peaks met the suggested peak of 23.47 years as observed in the original study. Therefore, in the final step of data analysis we wanted to try a different approach to resolve observed inconsistencies. The main objection toward the use of the SSA is the confounding of a person’s biological age and a song’s year of release. In other words, an SSA value of 30 could be the result of a combination of a song released in 2000 and a person being born in 1970 or of a song released in 1960 and a person being born in 1930, for instance. Accordingly, although it seemed to be an advantage for research to use the SSA as the only target value, it turned into a problematic coefficient: The influence of the particular songs’ features (e.g., genre, style), the generation of the listener and other influential variables on the rating of a song could not be separated.

To separate the influences of the mixed-up variables, a Latent Profile Analysis with song ratings as target variable and the year of song release was conducted. As suggested by Hemming (2013), the inverted U-shaped relationship between SSA and preference rating could also be the result of two complementary preference profiles: of preference ratings by, first, a group of listeners aged under about 40 years with ratings increasing with year of song release; and second a group aged over 40 years with ratings decreasing with the year of song release. According to Hemming’s observation, young people should have strong preferences for the latest music available, and “the older people get, the stronger their preference for the oldest music available is” (p. 301). The turning point for the change of rating patterns should be 37.51 years.

Thus, the main aim of our statistical decomposition approach was the explanation of the overall course of the regression curve as an outcome of a limited number of typical rating patterns (class profiles). This approach put clear emphasis on the song level as a predictor, and the release year was kept constant.

First, raw song ratings were aggregated into decades for every person. Second, based on the software Mplus (Version 8.3, see Muthén & Muthén, 1998–2017), we looked for a limited number of prototypical rating patterns (latent classes). The SSA proposition predicted only one inverted U-shaped rating pattern. Finally, we decided on one of the models that differed in their number of profile classes. As can be seen in Table 3, the best model fit was obtained from a 4-class solution: According to the principle of parsimony (proportion of explained variance and number of required parameters), the BIC values were lowest for this model. In addition, the values for Log Likelihood and AIC did not change significantly when compared to the 5-class solution. The validity of the 4-class solution was also supported by the high entropy value and the smallest class size of 10% (as a rule of thumb, the minimum size should be at least 5% of the total sample size, see Marsh et al., 2009; Masyn, 2013).

The profile patterns of the four identified classes are shown in Table 4 and Figure 9. Profile 1 can be regarded as a subgroup, the members of which did not prefer music from the charts at all. The ratings of this subgroup remained in the lowest vicinity of the rating scale. Profile 2 shows a tendency for a better evaluation of music from former times. Within this subgroup, ratings decreased when the song’s release was more recent. This pattern would correspond to Hemming’s (2013) suggested preference profile...
for people older than approximately 40 years. On the contrary, Profile 3 shows a tendency toward increasing preference, the more recent a song’s release was; in other words, the more recent the release year, the higher the preference for a song. This pattern would correspond to Hemming’s (2013) suggested preference profile for people younger than approximately 40 years. Finally, Profile 4 reflects a constantly positive rating of music from the charts independent of the release year. Based on these four latent profiles, it became clear that the level of songs as a unit of

Table 3. Overview of model fit indices for the Latent Profile Analysis (boldface indicates best class fit).

| Model | No. of classes | Log Likelihood | AIC | BIC | SABIC | Entropy | Smallest class | LMR p value | LMR meaning | BLRT p value | BLRT meaning |
|-------|----------------|----------------|-----|-----|-------|---------|---------------|-------------|-------------|--------------|--------------|
| 1     | 1   | -6690.889     | 13417.778 | 13473.355 | 13416.371 |        |               |             |             |              |              |
| 2     | 2   | -6559.351     | 13174.702 | 13261.154 | 13172.512 | .808   | 48%           | .0023       | 2 > 1       | < .001       | 2 > 1         |
| 3     | 3   | -6515.308     | 13106.616 | 13223.944 | 13103.645 | .831   | 17%           | .4228       | 3 < 2       | < .001       | 3 > 2         |
| 4     | 4   | -6461.405     | 13018.810 | 13167.015 | 13015.057 | .857   | 10%           | .1155       | 4 < 3       | < .001       | 4 > 3         |
| 5     | 5   | -6443.333     | 13002.665 | 13181.746 | 12998.130 | .882   | 4%            | .5444       | 5 < 4       | < .001       | 5 > 4         |

Table 4. Characteristics of the latent profiles for the 4-class model.

| Variables (Decades) | Profile 1 (n = 58; 35.80%) | Profile 2 (n = 43; 26.54%) | Profile 3 (n = 45; 27.78%) | Profile 4 (n = 16; 9.88%) | Variance |
|---------------------|-----------------------------|----------------------------|-----------------------------|-----------------------------|----------|
|                     | M   | SE  | M   | SE  | M   | SE  | M   | SE  | Estimate | SE  |
| 1930s               | 28.070 | 2.892 | 57.045 | 4.028 | 37.175 | 6.474 | 64.137 | 6.370 | 371.801 | 40.967 |
| 1940s               | 26.432 | 2.627 | 53.591 | 4.102 | 35.199 | 5.453 | 83.378 | 6.438 | 263.438 | 50.464 |
| 1950s               | 23.011 | 2.554 | 59.774 | 5.326 | 41.663 | 5.453 | 83.378 | 6.438 | 263.438 | 50.464 |
| 1960s               | 25.778 | 2.261 | 64.123 | 6.364 | 48.666 | 4.297 | 77.360 | 5.138 | 287.482 | 50.066 |
| 1970s               | 29.782 | 2.820 | 56.208 | 6.282 | 54.458 | 4.348 | 74.561 | 5.377 | 263.438 | 50.464 |
| 1980s               | 39.139 | 2.619 | 48.335 | 5.654 | 66.138 | 3.113 | 79.485 | 6.297 | 234.151 | 50.780 |
| 1990s               | 45.895 | 2.666 | 42.020 | 5.359 | 76.627 | 4.267 | 76.740 | 5.287 | 254.756 | 47.767 |
| 2000s               | 30.835 | 2.809 | 38.794 | 5.549 | 55.586 | 2.773 | 75.192 | 5.102 | 336.666 | 44.357 |
| 2010s               | 38.184 | 2.901 | 39.190 | 4.139 | 65.688 | 5.066 | 62.674 | 8.536 | 345.925 | 49.719 |

Note. Values in boldface indicate the highest mean preference rating within a decade (max. = 100) across all latent profiles. Estimates (and SE) for the variances of indicators were held constant across all latent profiles.

Figure 9. Latent profile plot for the 4-class solution.
investigation should not be ignored. However, the influence of a particular song on musical preference judgments remained obscure if the concept of SSA was used.

**Discussion**

In this replication study, we tried to verify the age peak of 23.47 years as identified in Holbrook and Schindler’s (1989) original concept of SSA. According to Holbrook and Schindler, music from the charts listened to at this particular age in early adulthood and late adolescence is evaluated positively over the total lifespan. However, we found little empirical support for the SSA proposition. The following discussion of our findings is guided by two perspectives: first, the best fit of the regression curve peak and the difference of these observed peaks from the original finding of 23.47 years, and second, the quality of the curve fit as represented by the goodness-of-fit measure $R^2$, which measured the strength of the relationship between the (quadratic) regression model and the dependent variable (preference ratings for songs) on a scale from 0 to 100%. Holbrook and Schindler’s (1989) finding of an SSA peak at 23.47 years and a quadratic goodness-of-fit measure of $R^2 = .706$ shall serve as benchmarks.

For the regression peak fit, the smallest difference from the original finding (23.47 years) was observed for SSA = 22.63 years of the subgroup of participants aged 50+ years in our study (aggregated data, no data trimming; see Table S4 in the Supplemental Material section). However, the parameter adjustment for the SSA value of 22.63 years significantly differed from the original study: The peak of 22.63 years was not obtained from the total sample of $N = 162$ but from the ratings of a subgroup of participants aged 50+ years ($n = 124 \ (76.5\%)$). This closest approximation to the original SSA peak did not improve the fit quality and only reached a goodness-of-fit index of $R^2 = .225$. According to Ellis (2010, p. 41), this index value can be regarded as of medium effect size (benchmarks for $R^2 = .02, .13, .26$ for small, medium, and large effect sizes, respectively). The second closest SSA peak fit was found for the full sample after trimming for SSA categories with at least five values ($n = 128 \ (79\%)$) and revealed an SSA value of 27.07 years. However, in this case the parameter adjustments also differed from the original study (resulting goodness-of-fit index was $R^2 = .116$).

As an overall tendency, the goodness-of-fit index increased with the increasing age of the subgroups (see Table S4 in the Supplemental Material section). When we used the same statistical procedures as in Holbrook and Schindler’s (1989) study, an SSA peak of 14.16 years (full sample, aggregated data) was identified ($R^2 = .891$). This peak value was much lower when compared with the SSA peak of 23.47 in the original study. Finally, careful ocular inspection of the distribution of all regression peaks from ours as well as from Hemming’s (2013) replication study revealed a considerable range of peaks resulting from the application of reasonable methods of data trimming and curve fitting (see Figure 8). The resulting peak from regression analyses seemed to be extremely sensitive to smallest methodological changes in data analysis: The peak for the untrimmed sample was 14.16 years while the subgroup analysis of people aged 50+ resulted in a peak of 22.63 and for the group aged 70+ in a peak of 32.37 years. If the SSA were a stable concept, it would be robust against marginal changes in data analysis.

Similar inconsistent results were observed for the re-analysis of Hemming’s (2013) data set: The quadratic regression fit (comparable to the original study) resulted in a peak at SSA = 8.58 years ($R^2 = .454$). After the trimming of included SSA categories by a minimum of 10 SSA ratings per category, the best quadratic fit was observed for SSA = 20.83 years ($R^2 = .891$), the smallest difference from the original finding of 23.47 years. As in the analysis of our own data, a high dispersion of SSA peaks resulting from various but reasonable trimming methods was also found for Hemming’s (2013) data set (see Figure 8 and Table S6 in the Supplemental Material section).

Our search for a reliable and exact value of the SSA has proven challenging and, thus, it seems more promising to regard musical preference as a lifelong development. For example, the Music Preferences in Adulthood Model (MPAM, Bonneville-Roussy et al., 2017) offers such an alternative view on the complexity of age-related dynamic changes in musical behavior. Based on a large sample of responses of more than 4,000 adults to musical examples from a comprehensive database of 280 short clips from 27 genres, Bonneville-Roussy et al. (2017) posited that musical preferences vary with age. For example, preferences for rap and rock music decrease whereas preferences for classical, unpretentious, and jazzy music increase over time (between 18 and 65 years).

In a subsequent study, Bonneville-Roussy and Rust, (2018) investigated different sources of social influences and developed an integrative model of the psychological determinants of musical preferences in adulthood. They emphasized the role of social networks and interpersonal disposition toward conformity as important for the lifespan development of musical preferences. Finally, the acoustical features of music also play an important role for the development of preferences in adulthood: Bonneville-Roussy and Eerola (2018) conducted acoustical analyses of the genre clips preferred by different age groups and found that, in particular, the feature of tonal clarity contributed to the explanation of musical preferences in the middle adulthood group (40 to 65 years).

Of course, our results were limited by the selection of songs: due to the older population in our sample, participants rated only a random selection of 18 out of 87 songs (two songs per decade). Thus, based on this incomplete design, we could not avoid the influence of the variable “song” on ratings although the selection of sound examples
in our study was randomized. However, the magnitude of the random factor “song” can be estimated from previous research: As revealed by Zimprich and Wolf (2016), every recognized song in their study which fell under the category of “bump years” (the age range between 10 and 30 years) decreased the mean of the memory distribution by 1.77 years. In contrast, Zimprich (2020) reported that the number of recognized songs resulted in a higher and later bump. These random effects cannot be avoided as long as no standardized set of musical stimuli and no research design is used; however, we point out that our multilevel regression analysis (see Column 1 of Table S7 in the Supplemental Material section) only revealed a relatively small contribution of 8.9% for this random part of the total variance. The use of an incomplete design and a randomized selection of songs has also been applied in other studies, such as that of Spivack (2019), who presented a much smaller subsample of songs to each of their N = 643 participants (7/152 from seven decades = 4.6% of the total number of songs, released between 1940 and 2015) than we did (18/87 from eight decades = 20.7% of the total number of songs, released between 1930 and 2017).

Another important limitation caused by the standardized use of mainstream popular music for the determination of SSAs is that other musical genres fall between the cracks and remain unconsidered. When this happens, the musical material from the mainstream might be meaningless or cause unpredictable aversions in listeners. For example, for a follower of punk rock, it is unlikely that one of the Number One titles of the German charts from 1985 (e.g., “You’re my Heart, You’re my Soul” by Modern Talking) will cause positive memories. In addition, the classical concept of SSA becomes doubtful against the background not only of different musical genres (e.g., classic, or jazz) but also of the increasing individualization of music usage in the age of digital music streaming.

Although the SSA concept of critical periods for the imprinting of lifelong preferences might not be valid for the domain of music, it might be valid for other products. For example, the concept has been applied to non-musical domains such as men’s preferences for women’s fashion styles (Schindler & Holbrook, 1993, mean Style-Specific Age = 32.93 yrs), the appearances of male and female movie stars (Holbrook & Schindler, 1994, mean Star-Specific Age = 13.9 yrs), and males’ preferences for car makes (Schindler & Holbrook, 2003, mean Product-Specific Age = 25.74 yrs). In a comprehensive approach, Holbrook (1995) found evidence for the validity of the SSA concept for 21 other product categories such as soft drinks, cereals, novels, and toothpastes. Holbrook (1995) concluded that older respondents prefer cultural products that were experienced earlier on in life. Preferences for these products are assumed to be formed in the transition between late adolescence and early adulthood.

But how can the observed inconsistencies between later studies and Holbrook and Schindler’s study be resolved? We should bear in mind that the age peak of 23.47 years for maximum song preference as predicted by the SSA proposition could not be replicated in two subsequent replication studies by Hemming (2013 and the current study). An alternative explanation for the observed age-related changes in song evaluation was obtained from the more sophisticated Latent Profile Analysis approach. The advantage of this method of data analysis is that a song’s release year is kept constant. However, under these circumstances, the maximum peak of music preference can be explained as the result of the rating patterns of four subgroups and seems to be independent of the participants’ ages. This finding is in line with the study by Jimenez et al. (2020) on differences in feelings among participants when remembering a song (in this case Adele’s song “Hello” from 2015), which was not driven by the participants’ age differences.

Two failed replications of the predicted exact value of an SSA age of 23.47 years as the maximum peak of liking could also be interpreted as the result of other influential variables: First, all three studies used other compilations of songs, which makes studies hardly comparable; second, the age distribution of the samples differed significantly between the studies, and the number of participants aged 60+ was insufficient in two out of three cases; third, the original finding of Holbrook and Schindler (1989) could be a random effect caused by the regional and country-specific composition of the sample (this aspect is emphasized, for example, by López-Cano et al., 2020); fourth, unfortunately, residual fit analyses were reported neither by Holbrook and Schindler (1989) nor by Hemming (2013), so it remains unclear whether the distributions of music preference judgments of all songs were a composition of latent subgroups of different size and, thus, responsible for the different SSA estimation for the peak in lifelong music preference. Following from this, future studies should investigate lifelong music preferences by means of mixture-modeling approaches including covariance analyses for the investigation of individual predictors (i.e., the year of birth could partly determine SSA) as well as multilevel analyses for contextual factors such as country-specific differences in musical preferences, musical background or preferences for musical genres other than mainstream pop music.

To summarize, one or two failed replications are not sufficient for the falsification of a finding and, thus, we cannot rule out that human cultural and aesthetic behavior might be imprinted during critical periods in late adolescence and early adulthood. Indeed, the historical context of listening biographies should also not be disregarded. What is implied by musical “mainstream” has changed significantly over the lifespan of the elderly participants; for example, the change toward the English language in popular music and unique circumstances in the post-war era in Germany. However, a concept that is regarded as a fundamental psychological mechanism should be sufficiently robust against those influential factors. Of course, more
replication studies will be required to verify the precise date – if there is any at all. Whatever the outcome of future replications will be, the age range of such phases should not be understood in terms of static physical laws, but as being influenced by the dynamic changes of social-psychological determinants (e.g., social networks, levels of conformity) which are typical for modern societies (Bonneville-Roussy & Rust, 2018). Thus, if such sensitivity peaks exist at all, they should be updated for each generation. In our replication study, the prediction of the SSA proposition of a clearly identifiable sensitivity peak around a very limited period during late adolescence and early adulthood (point estimate 23.47 years) could not be confirmed. Strictly speaking, nowadays the SSA proposition seems to be only of limited validity for the domain of musical preferences.

Action editor
Alexandra Lamont, Keele University, Department of Psychology.

Peer review
One anonymous reviewer.
Jan Hemming, University of Kassel, Germany, FB01, Music Institute.

Contributorship
RK conceived the study; JW conceived the study and conducted the data analyses; FP conducted additional data analyses; RK, FP, and ND wrote the paper.

Acknowledgments
We would like to thank Jan Hemming (University of Kassel, Germany) for making his data set available. The audio examples used in this study are available from the corresponding author on reasonable request. We are also grateful to the senior centers MGH Langenhagen, Seniorenclub Bothfeld and Begegnungstätte Silbernkamp, Hanover, Germany. We are also grateful to two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

Declaration of conflicting interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs
Reinhard Kopiez https://orcid.org/0000-0003-3356-3478
Friedrich Platz https://orcid.org/0000-0002-9758-9434
Nina Düvel https://orcid.org/0000-0003-0877-0483

Note
1. Promptly after the completion of our manuscript, the study by Senn et al. (2021), regarding the influence of familiarity, taste, and SSA on the groove experience, was published. No support was found for an influence of Song-Specific Age on the urge to move to song. The groove experience did not show a reminiscence bump.

2. In his anthology, Ehnet (2002) based his selections on the following criteria: From 1930 to 1955, the three bestselling songs of a year were used, but no ranking information was available from the music industry; from 1956 to 2001 the German Top 30 sales chart rankings were included. From 1956 to 1989, the official sales charts were based on sales of singles, but the chart calculation in Germany was modified in the subsequent years. From 1989 to 2001, a combination of sales charts and airplay was used. For recent years, the following were considered in the rankings: since 2001 additional online sales (e.g., by Amazon), since 2002 sales of music videos, and since 2004 digital downloads (e.g., iTunes). For the period following 2001, we used additional sources on charts published by the official German chart provider Media Control (for 2002 to 2004: http://mix1.de/chartsarchiv/jahrescharts2002.htm; for 2005: http://viva.tv/charts/viva-single-jahrescharts [websites for 2002 to 2005 are no longer available]; for 2006 to 2010: http://musikindustrie.de/jahrescharts; for 2011 to 2017: https://www.offiziellecharts.de/charts/single-jahr/for-date-2011 [chart rankings were based on sales numbers of 2,800 German music traders and data were collected by the Consumer Research Association, GiF].

Supplementary materials
Additional material is available from the Supplemental Material section of this paper. Readers are encouraged to try their own analyses with the application of additional trimming options. The Excel workbook including the macros is available from https://osf.io/fywea/

References
Altenmüller, E., Siggel, S., Mohammadi, B., Samii, A., & Münte, T. F. (2014). Play it again, Sam: Brain correlates of emotional music recognition. Frontiers in Psychology, 5, Article 114. https://doi.org/10.3389/fpsyg.2014.00114
Basso, D., Basso, L., & Olivetti Belardinelli, M. (2004). “Do you remember?” - How memory is involved in song recognition. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), Proceedings of the 8th International Conference on Music Perception & Cognition, Evanston, IL, USA, 3-7 August. Causal Productions.
Bigand, E., Vieillard, S., Madarell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. Cognition and Emotion, 19(8), 1113–1139. https://doi.org/10.1080/02699930500204250
Bonneville-Roussy, A., & Erøla, T. (2018). Age trends in musical preferences in adulthood: 3. Perceived musical attributes as intrinsic determinants of preferences. Musicae Scientiae, 22(3), 394–414. https://doi.org/101177/1029864917718606
Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle
adulthood. *Journal of Personality and Social Psychology, 105*(4), 703–717. https://doi.org/10.1037/a0033770

Bonneville-Roussy, A., & Rust, J. (2018). Age trends in musical preferences in adulthood: 2. Sources of social influences as determinants of preferences. *Musicae Scientiae, 22*(2), 175–195. https://doi.org/10.1177/1029864917704016

Bonneville-Roussy, A., Stillwell, D., Kosinski, M., & Rust, J. (2017). Age trends in musical preferences in adulthood: 1. Conceptualisation and empirical investigation. *Musicae Scientiae, 21*(4): 369–389. https://doi.org/10.1177/1029864917691571

Bonneville-Roussy, A., & Rust, J. (2018). Age trends in musical preferences in adulthood: 2. Sources of social influences as determinants of preferences. *Musicae Scientiae, 22*(2), 175–195. https://doi.org/10.1177/1029864917704016

Borson, S., Scanlan, J. M., Chen, P., & Ganguli, M. (2003). The Mini-Cog as a screen for dementia: Validation in a population-based sample. *Journal of the American Geriatrics Society, 51*(10), 1451–1454. https://doi.org/10.1046/j.1532-5415.2003.51465.x

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

Deruty, E., & Tardieu, D. (2014). About dynamic processing in mainstream music. *Journal of the Audio Engineering Society, 62*(1/2), 42–55. https://doi.org/10.17743/jaes.2014.0001

Deutsche Gesellschaft für Psychologie. (2016). *Berufsethische Richtlinien* [Guidelines for professional ethics]. https://www.dgps.de/fileadmin/documents/Empfehlungen/berufsethische_richtlinien_dgps.pdf

Ehnet, G. (2002). *Hit Bilanz. Deutsche Chart Singles 1956-2001 (mit den Hits von 1930-1955)* [Hit balance. German chart singles 1956-2001 (including Hits from 1930-1955). Taurus Press.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Erikson, E. H. (1950). *Childhood and society*. Norton.

Greasley, A., & Lamont, A. (2016). Musical preferences. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Music psychology* (2nd ed., pp. 263–281). Oxford University Press.

Hanover University of Music, D. a. M. (2017). *Leitlinien guter wissenschaftlicher Praxis* [Guidelines for good scientific practice]. https://www.musikwissenschaft.hmtm-hannover.de/fileadmin/www.musikwissenschaft/Downloads/HMTMH_Regeln_guter_wissenschaftlicher_Praxis.pdf

Hemming, J. (2013). Is there a peak in popular music preference at a certain song-specific age? A replication of Holbrook & Schindler’s 1989 study. *Musicae Scientiae, 17*(3), 293–304. https://doi.org/10.1177/1029864913493800

Holbrook, M. B. (1996). The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results. *Journal of Consumer Research, 16*(1), 119–124. https://doi.org/10.1086/209200

Holbrook, M. B., & Schindler, R. M. (1994). Age, sex, and attitude toward the past as predictors of consumers’ aesthetic tastes for cultural products. *Journal of Consumer Research, 16*(3), 412–422. https://doi.org/10.2307/3152228

Holbrook, M. B., & Schindler, R. M. (1996). Market segmentation based on age and attitude toward the past: Concepts, methods, and findings concerning nostalgic influences on customer tastes. *Journal of Business Research, 37*(1), 27–39. https://doi.org/10.1016/0148-2963(96)00023-9

Holbrook, M. B., & Schindler, R. M. (2013). Commentary on “Is there a peak in popular music preference at a certain song-specific age? A replication of Holbrook & Schindler’s 1989 study”*. *Musicae Scientiae, 17*(3), 293–304. https://doi.org/10.1177/1029864913486665

Jakubowski, K., Eerola, T., Tillmann, B., Perrin, F., & Heine, L. (2020). A cross-sectional study of reminiscence bumps for music-related memories in adulthood. *Music & Science, 3*, 1–13. https://doi.org/10.1177/205924320965058

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex, 19*(11), 2579–2594. https://doi.org/10.1093/cercor/bhp008

Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterization of music-evoked autobiographical memories. *Memory, 15*(8), 845–860. https://doi.org/10.1080/09658210701734593

Janssen, S. M. J., Chessa, A. G., & Murre, J. M. J. (2007). Temporal distribution of favourite books, movies, and records: Differential encoding and re-sampling. *Memory, 15*(7), 755–767. https://doi.org/10.1080/09658210701539646

JASP Team. (2020). *JASP* (Version 0.12.2 [Computer software]). https://jasp-stats.org

Jimenez, I., Kuusi, T., & Doll, C. (2020). Common chord progressions and feelings of remembering. *Music & Science, 3*, 1–16. https://doi.org/10.1177/205924320916849

Johnston, R. R. (2016). The effect of repetition on preference ratings for select unfamiliar musical examples: Does preference transfer? *Psychology of Music, 44*(3), 514–526. https://doi.org/10.1177/096573561577248

Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Sage.

Kopiez, R., Platz, F., Müller, S., & Wolf, A. (2015). When the bump and enculturation. *Holbrook, M. B., & Schindler, R. M. (1989). Some exploratory findings on the development of musical tastes. Journal of Consumer Research, 16*(1), 119–124. https://doi.org/10.1086/209200

Kopiez, R., Platz, F., Müller, S., & Wolf, A. (2015). When the bump and enculturation. *Psychology of Music, 43*(3), 359–374. https://doi.org/10.1080/0965735613511505

Krumhansl, C. L., & Zupnick, J. A. (2013). Cascading reminiscence bumps in popular music. *Psychological Science, 24*(10), 2057–2068. https://doi.org/10.1177/0956797613486486

López-Cano, M. A., Navarro, B., Nieto, M., Andrés-Pretel, F., & Latorre, J. M. (2020). Autobiographical emotional induction in older people through popular songs: Effect of reminiscence bump and enculturation. *PLoS One, 15*(9), Article e0238434. https://doi.org/10.1371/journal.pone.0238434
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, 73*(11), 1969–1976. https://doi.org/10.1177/1747021820940300

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person-and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 191–225. https://doi.org/10.1080/10705510902751010

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. Little (Ed.), *The Oxford handbook of quantitative methods* (pp. 551–611). Oxford University Press.

Munawar, K., Kuhn, S. K., & Haque, S. (2018). Understanding the reminiscence bump: A systematic review. *PLoS One, 13*(12), Article e0208595. https://doi.org/10.1371/journal.pone.0208595

Muthén, L. K., & Muthén, B. O. (1999-2017). *Mplus user’s guide* (8th ed.). Muthén & Muthén. https://www.statmodel.com/html_ug.shtml

Navarro, D. J., & Foxcroft, D. R. (2019). *Learning statistics with Jamovi: A tutorial for psychology students and other beginners* (Version 0.70). https://doi.org/10.24384/hg3c-7p15

North, A., & Hargreaves, D. (2008). *The social and applied psychology of music*. Oxford University Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716–4717. https://doi.org/10.1126/science.aac4716

Pashler, H., & Wagenmaker, E.-J. (2012). Replicability in psychological science: A crisis of confidence? [Special issue]. *Perspectives on Psychological Science, 7*(6), 528–530.

Peretz, I., Gaudreau, D., & Bonnel, A.-M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition, 26*(5), 884–902. https://doi.org/10.3758/bf03201171

Platz, F., Kopiez, R., Hasselhorn, J., & Wolf, A. (2015). The impact of song-specific age and affective qualities of popular songs on music-evoked autobiographical memories (MEAMs). *Musicae Scientiae, 19*(4), 327–349. https://doi.org/10.1117/1029864919839172

Rathbone, C. J., O’Connor, A. R., & Moulin, C. J. A. (2017). The tracks of my years: Personal significance contributes to the reminiscence bump. *Memory & Cognition, 45*, 137–150. https://doi.org/10.3758/s13421-016-0647-2

Reips, U.-D. (2002). Standards for internet-based experimenting. *Experimental Psychology, 49*(4), 243–256. https://doi.org/10.1027/1618-3169.49.4.243

Reips, U.-D. (2012). Using the internet to collect data. In H. Cooper (Ed.), *APA handbook of research methods in psychology: Vol. 2. Research designs* (pp. 291–310). American Psychological Association.

Schindler, R. M., & Holbrook, M. B. (1993). Critical periods in the development of men’s and women’s tastes in personal appearance. *Psychology and Marketing, 10*(6), 549–564. https://doi.org/10.1002/mark.4220100607

Schindler, R. M., & Holbrook, M. B. (2003). Nostalgia for early experience as a determinant of consumer preferences. *Psychology and Marketing, 20*(4), 275–302. https://doi.org/10.1002/mar.10074

Senn, O., Bechtold, T. A., Florian, H., & Kilchenmann, L. (2021). Taste and familiarity affect the experience of groove in popular music. *Musicae Scientiae, 25*, 45–66. https://doi.org/10.1177/1029864919839172

Spivack. (2019). Who remembers the Beatles? The collective memory for popular music. *PLoS One, 14*(2), Article e0210066. https://doi.org/10.1371/journal.pone.0210066

Stephens-Davidowitz, S. (2018, Feb. 10). The songs that bind. *The New York Times*. Retrieved April 9, 2020, from https://www.nytimes.com/2018/02/10/opinion/sunday/favorite-songs.html

Szpunar, K. K., Schellenberg, E. G., & Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 370–381. https://doi.org/10.1037/0278-7393.3.2.370

The jamovi project. (2020). *jamovi Version 1.2* [Computer software]. https://www.jamovi.org

UNESCO. (2012). *International Standard Classification of Education [ISCED] 2011*. UNESCO Institute for Statistics. http://uis.unesco.org/sites/default/files/documents internacional-standard-classification-of-education-isced-2011-en.pdf.

VandenBos, G. R. (2015). Imprinting. In G. R. VandenBos (Ed.), *APA Dictionary of psychology* (2nd ed., pp. 528). American Psychological Association.

Warrenburg, L. A. (2020). Choosing the right tune: A review of music stimuli used in emotion research. *Music Perception, 37*(3), 240–258. https://doi.org/10.1525/MP.2020.37.3.240

Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2, Pt. 2), 1–27. https://doi.org/10.1037/h0025848

Zimprich, D. (2020). Individual differences in the reminiscence bump of very long-term memory for popular songs in old age: A non-linear mixed model approach. *Psychology of Music, 48*(4), 547–563. https://doi.org/10.1177/0305735618812199

Zimprich, D., & Wolf, T. (2016). The distribution of memories for popular songs in old age: An individual differences approach. *Psychology of Music, 44*(4), 640–657. https://doi.org/10.1177/0305735615578708