

Manuel Anglada-Tort, Thomas Baker, Daniel Müllensiefen

False memories in music listening: exploring the misinformation effect and individual difference factors in auditory memory

Journal article | Accepted manuscript (Postprint)

This version is available at <https://doi.org/10.14279/depositonce-7816>



This is an Accepted Manuscript of an article published by Taylor & Francis in "Memory" on 14 November 2018, available online: <http://www.tandfonline.com/10.1080/09658211.2018.1545858>.

Anglada-Tort, M., Baker, T., & Müllensiefen, D. (2018). False memories in music listening: exploring the misinformation effect and individual difference factors in auditory memory. *Memory*, 1–16.
<https://doi.org/10.1080/09658211.2018.1545858>

This paper is not the copy of record and may not exactly replicate the authoritative document published in the Taylor & Francis journal. Please do not copy or cite without author's permission.

Terms of Use

Copyright applies. A non-exclusive, non-transferable and limited right to use is granted. This document is intended solely for personal, non-commercial use.

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK

Technische
Universität
Berlin

False memories in music listening: Exploring the misinformation effect and individual difference factors in auditory memory

Manuel Anglada-Tort^{*}, Thomas Baker[#], and Daniel Müllensiefen[#]

^{} Audio Communication, Technische Universität Berlin, Berlin, Germany*

[#]Department of Psychology, Goldsmiths, University of London, London, UK

Correspondence concerning this paper should be addressed to Manuel Anglada-Tort,
Department of Audio Communication, Technische Universität Berlin, Berlin, Germany.

E-mail: m.angladatort@campus.tu-berlin.de

False memories in music listening: Exploring the misinformation effect and individual difference factors in auditory memory

The study of false memory has had a profound impact on our understanding of how and what we remember, as shown by the misinformation paradigm (Loftus, 2005). Though misinformation effects have been demonstrated extensively within visual tasks, they have not yet been explored in the realm of non-visual auditory stimuli. Thus, the present study aimed to investigate whether post-event information can create false memories of music listening episodes. In addition, we explored individual difference factors potentially associated with false memory susceptibility in music, including age, suggestibility, personality, and musical training. In two music recognition tasks, participants ($N=151$) listened to an initial music track, which unbeknownst to them was missing an instrument. They were then presented with post-event information which either suggested the presence of the missing instrument or did not. The presence of misinformation resulted in significantly poorer performance on the music recognition tasks ($d=.43$), suggesting the existence of false musical memories. A random forest analysis indicated that none of the individual difference factors assessed were significantly associated with misinformation susceptibility. These findings support previous research on the fallibility of human memory and demonstrate, to some extent, the generality of the misinformation effect to a non-visual auditory domain.

Keywords: false memory; misinformation effect; auditory memory; individual differences; music listening.

Introduction

To have a false memory is to recall something which did not happen (Smelser & Baltes, 2001). Motivated in part by its potentially serious implications - such as within courtroom testimonies - false memory has become one of the most widely explored topics in psychology in recent decades (see Brainerd & Reyna, 2005; Loftus, 2005; Neuschatz, Lampinen, Toggia, Payne, & Cisneros, 2017; Shaw, 2016; Scoboria et al., 2017, for reviews). From this wealth of research, it has now become clear that false memories can significantly influence our visual memory, from deteriorating the accuracy of eyewitness' memory (Davis & Loftus, 2007; Gabbert, Memon, & Allan, 2003; Liebman et al., 2002; Weingardt, Toland, & Loftus, 1994), to creating autobiographical memories of entire events that never occurred (Bernstein & Loftus, 2009a; Hyman, Husband, & Billings, 1995; Hyman & Kleinknecht, 1999; Wade, Garry, Read, & Lindsay, 2002). Nevertheless, it remains unclear to what extent memories of non-visual auditory stimuli, such as music, may prove exceptional or congruent within the broader realm of false memory. To what extent are false musical memories consistent with findings from the visual domain? The present study addresses this question by adapting for the first time the misinformation paradigm to use music instead of visual materials.

Two of the most prominent paradigms to induce false memory in the visual domain are the misinformation paradigm (see Frenda, Nichols, & Loftus, 2011; Loftus, 2005; Pickrell, McDonald, Bernstein, & Loftus, 2016, for reviews) and the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). The misinformation paradigm typically involves three stages: experiencing an event, receiving post-event misinformation, and a memory test. Findings across studies show that after exposure to post-event misinformation, the accuracy of individuals' memory

is altered (e.g., Loftus, Miller, & Burns, 1978; Loftus & Hoffman, 1989). In the original form of the DRM paradigm, participants are presented with a list of semantically related words (e.g., bed, night, rest, awake, dream, blanket, snore, nap). These words are all related by a word known as a “lure” (e.g., sleep), which is missing in the list. When participants are then required to remember as many words from the list as possible, they usually recall the lure as frequently as they recall the other presented words (Roediger & McDermott, 1995).

To the best of our knowledge, there are only two studies in the published literature that explored false memory in music, adapting a version of the DRM paradigm using music stimuli (Curtis & Bharucha, 2009; Vuvan, Podolak, & Schmuckler, 2014). In both studies, participants heard a melody followed by a test tone and were asked to indicate whether they had heard the test tone in the previously presented melody. Findings in both studies showed that participants were more likely to falsely remember context-congruent tones than context-incongruent tones. These studies demonstrate the generality of the DRM paradigm to a non-verbal auditory domain, suggesting that general theories for DRM false memory (see Brainerd & Reyna, 2005, for a review) may also apply to musical memory. However, there are no studies in the published literature that have applied the misinformation paradigm using music materials. This is surprising given the vast interest that misinformation effects have received in the last decades (Figure 1) and the implications that this paradigm has had for many disciplines, including psychology, social sciences, and law (see Brainerd & Reyna, 2005; Loftus, 2005; Neuschatz, Lampinen, Toggia, Payne, & Cisneros, 2017; Shaw, 2016; Scoboria et al., 2017, for reviews). The current research attempted to fill this gap by creating a musical version of the misinformation paradigm as implemented

in Loftus et al. (1978) seminar paper, “Semantic integrations of verbal information into a visual memory”.

Insert Figure 1 Here

Practical implications

There are both real-world and theoretical implications of studying the misinformation effect within music listening. When listening to music or attending a live music event, people are often exposed to post-event information about the musical episode, such as descriptions by peers, reports and reviews from concerts, or judgments from music auditions and competitions. In these situations, exposure to post-event misinformation could have an impact on listeners’ attitudes and behaviours towards music, influencing their preferences, taste, and habits. Evidence for this assumption comes from research on food preferences. For instance, suggesting to people that they had previously become ill after eating a certain food (e.g., boiled eggs or pickles) affects how much of that food the person consumes in the present (Bernstein & Loftus, 2009b; Bernstein, Laney, Morris, & Loftus, 2005; Bernstein, Scoboria, & Arnold, 2015). Interestingly, this influence in the individual’s current behaviour towards food can be reversed if the information suggests, instead, that he or she previously “loved” the food in question, increasing their current preference and consumption.

Under examination situations or competitions, music performances are often evaluated from memory. The outcome of these evaluative processes can be decisive for a musician’s career, determining, for example, whether a student is accepted in a prestigious music university or awarded in a competition situation. *The Queen Elisabeth Musical Competition* is one of the most well-known international competitions for

violin and piano, considered among the most demanding in the world. Members of the examination board include some of the worlds' greatest musicians, teachers, and music critics (Flôres & Ginsburgh, 1996). To win this competition would significantly change a musicians' future career. In the first two stages of the competition, the members of the jury listen to a number of different performances each day. It is only at the end of the day that the jury provide the final scores of the witnessed candidates. These evaluations reduce the total number of musicians to the final list of candidates. In the last stage of the competition, the jury listens to two candidates per day. Again, it is only at the end of each day that the jury provides their final scores. The grades are given without any discussion between judges and cannot be changed (Flôres & Ginsburgh, 1996; see Delhasse, 1985, for further information about the working of the competition). Thus, it is likely that the evaluative process in this and similar music competitions is susceptible to post-event misinformation.

There are even court cases where misinformation about music events might be influential, such as court decisions on music plagiarism. In the context of western pop music, melodic plagiarism has been a passionately debated phenomenon (Fruehwald, 1992; Müllensiefen & Pendzich, 2009), often because of the royalties gained from author's rights. In these cases, the goal of the courts and the Copyright Office is to judge whether the melodic material of two musical pieces (an original copyrighted piece and another performance by a different artist) is sufficiently similar or not. Although in most plagiarism cases there are music recordings that can be listened to repeatedly, there are also situations where recordings are not available. Thus, the evidence brought to court may instead simply rely on the memory of those who have attended concert performances from where music is alleged to have been plagiarised. For example, former *Thin Lizzy* guitarist Gary Moore had to pay damages after performing a guitar

solo on a recording considered to be plagiarised from a song written by the German band Jud's Gallery in 1974. This was even though the Jud's Gallery song was not available on record at the time of Gary Moore's recording (Graham, 2008). During court proceedings, Gary Moore was shown to be present at concert performances where the song was played, and it was argued that post-hoc information could have distorted his musical memory.

Theoretical implications

Beyond these implications, a musical version of the misinformation paradigm could provide valuable insights into the malleable nature of long-term representations of music. For instance, it could shed light on the interplay between *abstract structure* and *surface features*, the two types of musical information that people rely on when remembering music (Halpern & Müllensiefen, 2008; Peretz, Gaudreau, & Bonnel, 1998; Peretz & Zatorre, 2005; Schellenberg, Stalinski, & Marks, 2014; Trainor, Wu, & Tsang, 2004). A person's ability to remember music relies, to some extent, on surface features of the music. This includes the exact key (pitch level), precise tempo (speed), and timbre (the instrument on which the melody is performed). Nevertheless, the abstract structure of the music also plays an important role, including the relative pitch patterns (i.e., the pitch distance between tones regardless of their absolute pitch), relative durations (i.e., the ratios between durations of subsequent notes regardless of their absolute length), and contour (i.e., the sequence of ups and downs in a melody regardless of interval size). When processing the abstract structure of a given piece of music, listeners' commonly disregard surface characteristics. This is why people can effortlessly recognise "Happy Birthday" when played at almost any tempo, on any instrument, and sung at any pitch.

Based on this, one could expect misinformation effects to be likely in music listening: while a general abstract representation of the music piece may remain intact, specific perceptual features may be easily altered after exposure to post-event misinformation. However, we know that listeners (even when musically untrained) can remember the precise key, tempo, and timbre of familiar and unfamiliar music (Halpern & Müllensiefen, 2008; Levitin, 1994; Levitin & Cook, 1996; Poulin-Charronnat et al., 2004). These studies show that changes in these surface features significantly impair listeners' ability to recognise music. Thus, it remains unclear to what degree perceived surface features will be susceptible to misinformation effects.

Another important reason to study the misinformation paradigm within music listening tasks is to demonstrate the generality of the misinformation effect to the non-visual auditory domain, including theoretical accounts of false memory and general principles of memory. At a general level, the existence of misinformation effects in music could indicate that Bartlett's (1932) view on the reconstructive nature of memory and schema-based effects (see Alba & Hasher, 1983, for a review) also applies to memory for music. That is, musical events are not stored in memory like songs on a CD. Instead, musical events are reconstructed from memory using available schemata and knowledge structures. But demonstrating misinformation effects in music could also allow us to test and generalise more specific accounts of false memory to the music domain, such as *the source-monitoring framework* (SMF; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 2000), or *fuzzy trace theory* (FTT; Brainerd & Reyna, 2002; Reyna & Brainerd, 1995). Today, the extent to which these and other theoretical accounts of false memory apply to the music domain is largely unknown.

Individual differences

Finally, a musical version of the misinformation paradigm could constitute a suitable individual difference test to study listeners' susceptibility to false musical memories. In fact, this paradigm has been repeatedly used to examine individual differences and memory distortion using visual materials, showing that not everyone is equally susceptible to false memory (see Eisen, Winograd, & Qin, 2002; Loftus, 2005; Zhu et al., 2010, for reviews). The present study focused on four individual difference aspects that could be potentially associated with misinformation susceptibility in music, namely, age, personality, suggestibility, and musical training.

Firstly, there is evidence indicating that age matters. In general, young children and the elderly are more susceptible to misinformation effects than adolescents and adults (Davis & Loftus, 2005; see Wylie et al., 2014, for a review). Secondly, studies have identified associations between particular personality traits and false memory. For example, introverts are more likely to be affected by false memories than extroverts (Loftus, 2005; Porter, Birt, Yuille, & Lehman, 2000; Ward & Loftus, 1985), although this relationship was not supported by Liebman et al. (2002), who instead found associations between neuroticism and false memory occurrence. Even though findings on the relationship between personality variables and false memory are not always consistent, personality traits may be useful to increase our understanding of the processes underlying memory distortion (Frenda, Nichols, & Loftus, 2011). The third individual difference factor is suggestibility, which has also been linked to susceptibility to false memory. For example, false memory was positively associated with measures of social desirability (Tousignant, 1983, as cited in Schooler & Loftus, 1993) and hypnotisability (e.g., Barnier & McConkey, 1992). Though potential associations between false memory and these three individual differences (age, personality, and

suggestibility) have been demonstrated in past research (see Eisen et al., 2002; Loftus, 2005; Zhu et al., 2010, for reviews), the links between these individual differences and false memory in music have not yet been explored.

Musical training is another individual difference factor that could potentially mediate listeners' performance in a musical version of the misinformation paradigm. Traditionally, studies on musical memory have compared two groups of participants varying in their level of musical expertise (see Talami, Altoè, Carreti, & Grassi, 2017, for a review): musicians (i.e., participants with expertise playing a musical instrument, determined by the number of years of musical training or attendance to music conservatories or music schools) and nonmusicians (i.e., participants with little or no experience of playing a musical instrument). Using music stimuli (e.g., tones, chords, melodies), there is evidence indicating a superiority of musicians over nonmusicians in short-term memory tasks (Bidelman, Hutka, & Moreona, 2013; Monahan, Kendall, & Carterette, 1987; Pallesen et al., 2010; Williamson, Baddeley, & Hitch, 2010) and working memory tasks (Pallesen et al., 2010; Schulze, Mueller, & Koelsch, 2011). However, in the domain of long-term memory the evidence is less consistent. Cohen, Evans, Horowitz, and Wolfe (2011) and Weiss, Vanzella, Schellenberg, and Trehub (2015) found empirical evidence suggesting that musicians long-term memory for melodies is superior to nonmusicians, whereas Schiavio and Timmers (2016) found no differences between these two groups. The paradigm used in the current study can shed light on this issue by examining whether high levels of musical training reduce susceptibility to misinformation in music recognition tasks. In this study, musical training is measured on metrical scales and, therefore, it overcomes the traditional musician vs. nonmusician dichotomy.

Aims

The main aim of the present study was to investigate whether false memories can be induced within music listening tasks through the misinformation paradigm. As false memory has been consistently demonstrated within a wide range of visual scenarios (e.g., Brainerd & Reyna, 2002; Brainerd & Reyna, 2005; Liebman et al., 2002; Loftus, 2005; Neuschatz et al., 2017; Scoboria et al., 2017), it was hypothesised that participants will demonstrate susceptibility to false memories in music listening tasks when presented with misinformation. The second aim of the current research was to explore potential individual difference factors related to misinformation susceptibility in music listening. Although this second analysis was exploratory in nature, we had three hypotheses: (i) older participants will be more affected by post-event misinformation, (ii) suggestibility will be positively correlated with the effect of misinformation, (iii) and participants with more musical training will demonstrate less misinformation susceptibility. The assessment of the different measurements of personality traits and their potential to predict false memory occurrence was only exploratory within this study and no hypotheses were presented.

Method

Participants

A sample of 151 participants took part in the experiment. Of those, 143 disclosed their gender (80 female, 63 male) and age, ranging from 18 to 63 ($M= 31.04$, $SD= 8.10$). A total of 93 participants were allocated to the misinformation group and a total of 58 participants were allocated to the control group. Participants' mean score in the Gold-MSI musical training factor (Müllensiefen, Gingras, Musil, & Stewart, 2014) was 25.42 ($SD= 11.36$), which indicates an overall average level of musical training, corresponding to the 44th percentile of the data norms reported in Müllensiefen et al.

(2014). In the misinformation group, 88 participants disclosed their gender (49 female, 39 male) and age, ranging from ages 21-63 ($M = 33$, $SD = 8.27$). In the control group, 55 participants disclosed their gender (31 female, 24 male) and age, ranging from ages 18-49 ($M = 27.87$, $SD = 6.77$). Participation was on a voluntary basis and was unpaid.

An a-priori power analysis using an F-test for mixed within- and between-participants designs, with two groups (misinformation vs. control) and eight within-participant measurements (the total number of critical pairs), indicated a total sample size of at least 114 participants was necessary to detect a significant main effect of misinformation. Based on the estimated effect size from the results reported in Experiment 5 by Loftus et al. (1978), we set the effect size to .20. The significance level α and the power $1-\beta$ were set to .05 and .80, respectively. The power analysis was conducted using the G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007).

Design

The present study used a mixed within- and between-participants design. The within-participant variables were the type of clip (critical vs. noncritical clips), music piece (Bebop Jazz vs. Cool Jazz), and timbral manipulation (piano vs. drums). The between-participant variable was the presence of misinformation (misinformation vs. control group). The experiment had two parts. In each part, participants listened to and were tested on one of the two music pieces, which unbeknownst to them was missing an instrument (either piano or drums). The second part of the experiment used exactly the same procedure as the first, but altered the music piece and instrumental manipulation. Thus, all participants were tested two times, using two different music pieces and instrumental manipulations. The order of presentation of the two pieces of music and the instrumental manipulation was fully counterbalanced across participants.

Constructing an adapted musical version of the misinformation paradigm

The experimental design used in our study was inspired by Experiments 5 from Loftus et al. (1978), where participants were first shown a series of 20 slides depicting an auto-pedestrian accident: “A male pedestrian is seen carrying some items in one hand and munching on an apple held with the other. He leaves a building and strolls toward a parking lot. In the lot, a maroon Triumph backs out of a parking space and hits the pedestrian” (pp. 28-29). Four of the 20 slides were critical. Each version of the critical slides contained a particular object (“a pair of skis leaning against a tree”), whereas the other version contained the same identical slide with a changed detail (“a shovel leaning against a tree”). Participants only saw one version of the critical slides. Following a series of filler activities for approximately 10 minutes, participants read a three-paragraph description of the event. The description contained four critical sentences that either did or did not mention the incorrect critical objects (e.g., “if the participant had seen *skis* leaning against a tree, the statement might include a sentence that mentioned the *shovel* leaning against the tree). After another filler task of 10 min approximately, participants were given a two forced-choice recognition test, where they had to indicate which of the two slides on each pair was seen before. The test had an overall of 10 pairs of slides. Four of the 10 pairs were critical, containing a slide depicting the event as it corresponded with the incorrect information and another slide depicting the actual event. The authors found that the percentage of times a correct selection occurred was significantly lower (55.3%) when the information was misleading than when it was not (70.8%). This difference was statistically significant (Loftus et al., 1978).

The main challenge of constructing an adapted musical version of this paradigm was to use music stimuli instead of visual. In the exposure stage, instead of seeing 20 slides depicting an auto-pedestrian accident, participants listened to an instrumental piece of jazz music, which unbeknownst to them was missing an instrument, either

piano or drums. Instead of manipulating 4 critical pieces of information, we only manipulated one: the presence or absence of an instrument (piano or drums). While the visual presentation of 20 different slides allowed for different manipulations of critical information, the auditory presentation of a single piece of music did not. In the post-event information stage, participants read a descriptive text with a single critical piece of information, which could either suggest the presence of the missing instrument or did not. Finally, we tested participants using a two forced-choice recognition task analogous to Loftus et al. (1978). On this recognition task, participants had to decide which clip of the pair would correspond to the original recording they had heard. Four of the 10 pairs of clips were critical, containing a clip from the original track missing the target instrument (correct response) and a clip from the original track containing the target instrument (incorrect response). The order of presentation of the 10 trials as well as the order of presentation of the two clips within each trial was randomised for each participant.

Materials

Music stimuli

Two core tracks were used from the MedleyDB database (Bittner et al., 2014) by the artist “Music Delta”, namely the track “Bebop Jazz”, and the track “Cool Jazz”. MedleyDB is a dataset of annotated, royalty-free multitrack recordings, which were curated primarily to support music research (Bittner et al., 2014). Both tracks were instrumental jazz pieces, similar in style and using the same instrumentation. The original complete tracks both featured a drum set, a double bass, a piano, and a brass section. The original version of “Bebop Jazz” was 1 minute 43 seconds in length, and the original version of “Cool Jazz” was 1 minute 42 second in length. Alternative mixes were created using the *Ableton Live* software featuring the core tracks either without

drums or without piano (all other instruments and attributes remained the same). Each track was then shortened to 60 seconds, featuring 50 seconds of the original track from the start, and then a gradual 10-second fade out. In addition, we used a third track from the MedleyDB database called “Fusion Jazz”. This track was also an instrumental jazz track featuring drum set and saxophone, though it also featured electric bass, electric piano and synthesizer (rather than only acoustic instruments), creating a clear distinction with the original tracks.

To create the testing stimuli, the original and alternative mixes were used along with the third track to create three types of pairs of clips: (i) critical pairs (i.e., a clip from the original track either missing piano or drums and a clip from the same original track including the suggested target instrument), (ii) noncritical-brass pairs (i.e., a clip from the original track either missing piano or drums and a clip from the same original track missing the target instrument as well as an additional instrument, namely, the brass section), and (iii) noncritical-easy pairs (i.e., a clip from the original track missing piano or drums and a clip from a different hitherto unheard but stylistically similar track). The critical clips were used to measure the effect of post-event information on memory, whereas the noncritical clips were used to measure memory performance in the absence of misinformation. For each track and instrument condition, there was a total of 10 pairs of clips: four critical and six noncritical pairs (three noncritical-brass and three noncritical-easy pairs).

Post-event information

Descriptive texts consisting of a single paragraph of information on the original track were created. Except for the instrumentation referenced - which was incorrect only for the misinformation group - this text was wholly accurate and was identical across conditions. To increase credibility, the paragraph also included a web citation, pointing

to the Music Delta website. To ensure that participants read the post-event information, a sentence from the descriptive text was copied at the bottom of the survey page, with a blank space instead of one of the words for the participant to fill in and show they had read the text (the descriptive texts for all conditions, including the fill-in-the-blank sentence, may be found in Appendix A).

Individual difference factors

A variety of self-reporting questionnaires were used to measure individual difference factors. The Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014), a measurement instrument used to assess musical skills in the general population. The dimensions of musical sophistication measured by the Gold-MSI are (1) Active Engagement, Perceptual Abilities, Musical Training, Emotion, Singing Abilities, and General Musical Sophistication. Eight items from the Susceptibility to Persuade Strategies Scale (STPS; Kaptein, De Ruyter, Markopoulos, & Aarts, 2012; Eren, Unal & Temizel, 2014) that provide a measurement of general suggestibility in relation to two of the six persuasion principles identified by Cialdini (2001), namely, bias to authority and consensus. The Social Desirability Scale-17 (SDS-17; Stöber, 1999, 2001), which captures the tendency for individuals to self-describe with socially desirable attributes, as per Paulhus' impression management construct (1986). The Big Five Inventory (BFI; John & Srivastava, 1999), which measures an individual on the Big Five Dimensions of personality defined by Goldberg (1993), namely, Extraversion, Introversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. And the Revised Short Test of Music Preference (STOMP-R; Rentfrow & Gosling, 2003), measuring music preferences across 23 genres, which represent four higher-order dimensions of music preference. These four dimensions of music preference are named Intense and Rebellious, Upbeat and Conventional, Energetic and Rhythmic, and

Reflective and Complex (including preferences for jazz). These questionnaires were provided to fill the time between the exposure and the recognition test stages.

Procedure

The core procedure for the experiment (shown in Figure 2) involved participants listening to an initial track (“Cool Jazz” or “Bebop Jazz”), which unbeknownst to them was missing an instrument (either piano or drums). At this stage, participants were provided with the following instructions: “you will listen to a piece of music; you will later be tested on your memory for this piece of music, so please concentrate on the piece closely”. After listening to the piece of music, participants had to complete a filler task, consisting of filling several questionnaires for approximately 5-minutes. Following the filler task, participants had to read a descriptive text of the original track which either suggested the presence of the missing instrument in the original track (misinformation group) or did not (control group). Following another set of filler questionnaires for approximately 5-minutes, participants took the music recognition task, a 2-alternative-forced-choice recognition test of 10 pairs of clips, where participants had to choose the clip that was presented previously. In this recognition test, four pairs of clips were critical and six were noncritical. The order of the pairs, as well as the position of the two clips on the screen (left or right), were randomized for each participant. The entire procedure was repeated for a second time in part 2, where the initial track (“Cool Jazz” or “Bebop Jazz”) and the instrumental manipulation (piano or drums) was counterbalanced. Participants were sent the experiment to complete online via a single link, with an explanation that this was a test on memory and music, along with instructions providing an outline of the test structure. The experiment was constructed on the online survey software *Qualtrics* (Qualtrics, Provo, UT). The experiment was granted ethical clearance by the Ethics Committee of the Department of

Psychology, University of Goldsmiths, London, on 5 May 2017.

Insert Figure 2 here

Statistical analysis

Three participants were excluded from the subsequent analysis because they scored 0% in either the noncritical-easy clips or the noncritical-brass clips in the two parts of the experiment, indicating that they did not understand the experimental task. The average time to complete the experiment was 105.36 minutes ($SD= 342.88$). Five participants took more than two standard deviations from the mean to complete the survey and were excluded, resulting on a total average time of 54 minutes ($SD= 85.79$). Thus, the subsequent analysis included an overall of 143 participants, 88 in the misinformation group and 55 in the control group.

Matching procedure

Because individual differences in age have been associated with the susceptibility to false memories (see Wylie et al., 2014, for a review) and individual's musical training plays an important role in the ability to remember music (see Talami et al., 2017, for a review), we carried out a logistic regression analysis to investigate whether the two groups (misinformation vs. control) differed on age and musical training. The binary dependent variable was the group, and the two predictors were age and musical training. Table 1 shows the output of the model, which indicates that the two groups differed significantly on age (misinformation group, $M=32.96$, $SD= 8.39$; control group, $M=27.71$, $SD= 6.33$; $p= .002$), but not in musical training (misinformation group, $M=23.66$, $SD= 11.00$; control group, $M=28.40$, $SD= 11.53$; $p= .06$).

In order to correct for age differences, we used a nonparametric multivariate procedure to match the two samples on age and musical training, as implemented in the R package *MatchIT* (Stuart, King, Imai, & HO, 2011), which matches participants in two groups on the basis of several covariates at once. We used the nearest neighbour matching method and matched with the replacement method to enable one-to-many matching to accommodate the different sample sizes. The optimal solution to match participants in the two groups to correct for age differences excluded a total of 23 participants. After this procedure, we repeated the logistic regression analysis with the matched dataset. The two groups did not differ significantly in age (misinformation group, $M=29.97$, $SD= 4.74$; control group, $M=27.71$, $SD= 6.33$; $p= .09$), nor in musical training (misinformation group, $M=23.68$, $SD= 11.68$; control group, $M=28.92$, $SD= 11.50$; $p= .05$). Table 1 shows the output of the logistic regression model with the matched dataset. Thus, the final sample size for the main analysis regarding the effect of misinformation comprised a total of 120 participants (65 in the misinformation group and 55 in the control group).

Insert Table 1 here

The misinformation effect

The main analyses planned initially to measure misinformation effects in music listening followed the analysis strategy used in Experiment 5 by Loftus et al. (1978). This analysis aimed to compare our results, using music stimuli, with those reported in Loftus et al. (1978), using visual stimuli. This analysis only considered the critical pairs (i.e., a clip from the original track either missing piano or drums and a clip from the same original track including the suggested target instrument in the misinformation text). The

dependent variable was the total number of correct and incorrect responses, collapsed across all critical pairs in the two parts of the experiment. The independent variable was the experimental group (misinformation group vs. control group). An independent t-test was employed to test for significant differences between groups. We reported effect sizes in terms of Cohen's d because this measure can be used across several model types and is intuitively understood by most researchers. Moreover, we calculated the odds ratio to compare the effect sizes in our study and the ones reported in Loftus et al., (1978).

To expand this analysis using more advanced statistical techniques and also accounting for the timbral manipulation and different types of music clips, we performed an exploratory analysis using mixed-effects logistic regression as implemented in the R packages *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *car* (Fox et al., 2017). Mixed-effects logistic regression models have several advantages compared to ordinary logistic regression models. They can handle missing values and binomial or non-normal distributions, do not assume independence among observations, and can work with correlated observations. Mixed-effects logistic regression can also model random variability by assuming random intercepts for different relevant factors, such as participants' memory abilities, providing unbiased estimates of the coefficients of the predictor variables (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000).

This analysis was exploratory and different combinations of predictor and outcome variables were considered. In this paper, we report what we consider the most comprehensive model, including fixed effect factors for misinformation (presence vs. absence), type of clip (critical vs. noncritical), timbral manipulation (piano vs. drums), and the interaction between misinformation and type of clip. The binary response on each pair of clips (correct or incorrect) was the dependent variable. In addition, we specified a random intercept for participants because the individual ability of the

participants to perform on the recognition task contributed to the variance of the responses, inflating the overall variance of the data. The interaction term was used to study the effect of misinformation in the two types of pairs of clips, whereas the timbral manipulation factor was included to test whether memory performance was affected by the instrument under manipulation. The mixed-effects analysis was conducted using effects coding as opposed to the default treatment coding and Type-III Wald chi-square tests. The mixed-effect logistic regression model met the assumptions of linearity and normality. Collinearity was not an issue because the predictor variables were part of the orthogonal experimental design and, therefore, the association between these variables was 0. In addition, there were no correlations between residuals and the error variance of the residuals did not change across the range of fitted values.

Individual differences

A person-wise dependent variable was created, using the ratio of the individual's overall performance on the critical clips, divided by their performance on the noncritical clips. Ratio values of >1 suggested low susceptibility to false memories, whereas values of <1 indicated high susceptibility. For this analysis, we only used participants in the misinformation group, as they were the only exposed to incorrect information. We used the dataset from before the matching procedure was carried out, comprising a total of 88 participants. The subsequent analysis only includes the individual difference factors for which we had clear previous hypothesis based on the published literature (i.e., age, musical training, suggestibility, and personality; see Appendix B for a correlation matrix of the person-wise dependent variable and all 17 individual difference variables measured during the filler task phases of this study).

To examine which individual difference factors were associated with misinformation susceptibility, we used a data analysis method known as random forest

(Breiman, 2001), based on permutation tests, as implemented in the R package *party* (Hothorn, Bühlmann, Dudoit, Molinaro, & Van Der Laan, 2006; Hothorn, Hornik, & Zeileis, 2006; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008; Strobl, Malley, & Tutz, 2009). Compared to other classification and regression methods, random forests have several advantages, as they can handle complex interactions, large sets of predictor variables (even if they are highly correlated), and do not assume a linear relationship between predictors and dependent variable (see Hastie, Tibshirani, & Friedman, 2009). Moreover, random forest models use an in-built out-of-the-bag cross-validation mechanism that protects against alpha error inflations and overfitting.

The person-wise memory score was the dependent variable and the individual difference variables the predictors, including age, musical training, suggestibility (SPSS and SDS-17), and personality traits (i.e., Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness). The model was run with a size of 10,000 trees. The number of randomly preselected predictor variables to be chosen in each split was 4. The R^2 was calculated using the R package *caret* (Kuhn, 2008), which uses cross-validation and prevents model overfitting.

A measure of variable importance for each predictor was used to produce unbiased estimates even when there are significant correlations between predictor variables and/ or when the dependent variable is very unequally distributed (Janitza, Strobl, & Boulesteix, 2013). The variable importance score for each predictor variable describes how predictive the variable in question was relative to the other predictors. This includes their influence as main and in interaction effects. To select those variables that were associated with misinformation effects, we used a confidence interval criterion (Strobl et al., 2008; Strobl et al., 2009). This criterion indicates that only those variables whose importance score is positive and greater than the absolute value of the

lowest negative variable importance score should be selected.

Results

The percentage of times a correct response occurred was 64.23% ($SD= 22.95$) when the descriptive text contained incorrect information and 74.09% ($SD= 23.18$) when it did not. This difference was statistically significant, as indicated by an independent t-test, $t(118)= 2.33$, $p= .02$. The effect size was small to medium, $d= .43$. Based on the number of correct and incorrect selections in the presence and absence of misinformation, the odds ratio was also calculated, $OR= 1.59$, $\log \text{odds}= .46$, 95% CI [.87, 2.91].

To study whether the matching procedure affected the outcome of the main analysis regarding the misinformation effect, we repeated the same analysis using the dataset before the matching procedure was carried out ($N= 143$). Overall, the results were very similar. The percentage of times a correct response occurred was significantly lower when the descriptive text contained incorrect information ($M= 63.09$, $SD= 22.05$) than when it did not ($M= 74.09$, $SD= 23.18$), $t(141)= 2.63$, $p= .009$; $d= .49$; $OR= 1.61$, $\log \text{odds}= .48$, 95% CI [.88, 2.96].

The results of the exploratory analysis using mixed-effects logistic regression can be seen in Figure 3, which shows the proportion of correct selections in the presence and absence of misinformation in the two types of clips (critical vs. noncritical) and the two timbral manipulations (piano vs. drums). The mixed-effects logistic regression analysis revealed a main effect of misinformation, $\chi^2(1) = 5.96$, $p = .01$, type of clip $\chi^2(1) = 191.25$, $p < .001$, instrument $\chi^2(1) = 22.55$, $p < .001$, but the interaction between misinformation and type of clip was nonsignificant, $\chi^2(1) = .00$, $p = .99$. The classification accuracy of the model 1 was .84.

To test whether the order of the experiment (part 1 vs. part 2) had an effect on participants' performance on the music recognition task, we ran a mixed-effects logistic

regression analysis with order, misinformation, and the order-misinformation interaction as fixed factors. Participants ID was used as a random effect factor. The model indicated that the order of the experiment, $\text{clip } \chi^2(1) = .22, p = .63$, and the interaction term, $\text{clip } \chi^2(1) = .132, p = .25$, were nonsignificant, whereas the main effect of misinformation was significant, $\text{clip } \chi^2(1) = 6.88, p = .008$.

Because timbral manipulation was significant we performed mixed-effect logistic regression models to study the effects of misinformation in the two timbral manipulation. In the two models, misinformation and the interaction between misinformation and type of clip were the fixed factors, whereas participants the random effect factor. The mixed-effects logistic regression model with piano indicated a main effect of misinformation, $\chi^2(1) = 5.87, p = .01$, and type of clip $\chi^2(1) = 76.64, p < .001$, but no significant interaction between these two factors, $\chi^2(1) = .00, p = .97$. The classification accuracy of model 2 was .86. The model with drums revealed a main effect type of clip $\chi^2(1) = 110.50, p < .001$, but misinformation, $\chi^2(1) = 2.01, p = .16$, and the interaction between these two factors, $\chi^2(1) = .01, p = .92$, were nonsignificant. The classification accuracy of model 2 was .88.

Insert Figure 3 here

The random forests analysis revealed that none of the 9 individual difference factors were significantly associated with misinformation susceptibility. None of the positive variable importance scores were greater than the absolute value of the lowest negative variable importance score, indicating the person-wise dependent variable was not associated with any of the predictor variables. The overall R^2 of the random forest model was .19.

Discussion

The main aim of the present study was to investigate whether false memory can be induced within music listening tasks through the misinformation paradigm. The presence of post-event misinformation significantly deteriorated participants' performance in a music recognition task. Participants were more likely to select the wrong music clip (i.e., a clip containing an instrument that was never actually experienced) when the descriptive text included incorrect information suggesting the presence of the target instrument (36%) than when it did not (26%). This finding supports the malleable nature of long-term memory for music and suggests the existence of false musical memories. Regarding our initial question, *to what extent are false musical memories consistent with findings from the visual domain?*, the results reported in this study are congruent with research on the misinformation effect in visual recognition tasks (see Frenda et al., 2011; Loftus, 2005; Pickrell, et al., 2016, for reviews) and more broadly with the existence of false memories (Brainerd & Reyna, 2002; Brainerd & Reyna, 2005; Neuschatz et al., 2017; Shaw, 2016; Scoboria et al., 2017). Therefore, the findings reported here are also subject to the misinformation debate and related issues, such as whether impairment by post-event information exists or not (Loftus, Schooler, & Wagenaar, 1985; McCloskey & Zaragoza, 1985). To the best of our knowledge, this is the first published study demonstrating, at least partly, the generality of the misinformation effect to a non-visual auditory domain.

The present study attempted to create a musical version of the misinformation paradigm based on Experiment 5 in Loftus et al. (1978). Nevertheless, there are important differences to consider between these two studies. While we used a design in which the presence of misinformation was manipulated between-participants, Experiment 5 in Loftus' et al. (1978) used a repeated-measures design. Moreover, we used shorter filler tasks (approximately 5 minutes each) compared to Loftus et al. (1978;

approximately 10 minutes each); and allowed participants to listen to the clips multiple times. Despite these differences, the misinformation effect reported in Loftus et al. (1978) and the one observed in the current research were fairly similar in size, as indicated by the odds ratio (1.59 in the present study and 2.12 in Loftus' Experiment 5). Using visual stimuli, participants in Loftus' study (1978) selected the right clip 53% of the times in the presence of misinformation and 71% in the absence. In the present study, using musical materials, participants selected the right clip 64% of the times when the text included misinformation and 74% when it did not.

The initial analysis used to measure misinformation effects in music listening was based on Loftus' et al. (1978) and, therefore, only considered the proportion of wrongly selected clips on the critical trials. Accordingly, to establish a misinformation effect, the proportion of wrongly selected critical clips should be significantly higher on the misinformation group than in the control group. Using this criterion, the misinformation effect in our dataset was clear. Nevertheless, we carried out a second exploratory analysis using a more comprehensive statistical model (i.e., mixed-effects logistic regression) and accounting also for type of music clips (critical vs. noncritical) and type of misinformation (piano vs. drums). Note that the critical trials were designed explicitly to measure misinformation effects (i.e., including the target instrument in one of the two clips), whereas the noncritical clips were designed to measure general musical memory ability.

Thus, a second criterion could be used to determine whether the misinformation effect was established or not. That is, the increase in wrongly selected clips from noncritical to critical trials should be significantly higher only for those participants in the misinformation group. According to this, we were not able to establish a statistically significant misinformation effect. In other words, the interaction between type of clip

and misinformation was nonsignificant. However, there was a clear tendency supporting the interaction term between misinformation and type of clip (Figure 3). Moreover, it is likely that the presence of misinformation increased the overall difficulty of the test only in the misinformation group, as they had to read and process counterfactual information. Thus, the interference produced by the post-event misinformation could have resulted in an overall poorer performance in both critical and noncritical pairs compared to the control group. This second analysis was only exploratory and future research could improve our design by planning this type of analysis in advance and using a measurement of general long-term memory for music in which misinformation is not confounded, such as Gordon (1989) or Harrison, Collins, and Müllensiefen (2017). We also encourage future researchers to use within-participants designs to avoid the comparison of independent groups as well as use more comprehensive analysis strategies that take the performance on critical and noncritical clips into account within the same model.

In addition, the mixed-effects logistic regression analysis also revealed a significant main effect of instrumental manipulation. Participants performed significantly better when the misinformation paradigm involved a piano manipulation than when it involved drums. To explore this further we conducted two separate analyses for each instrument (piano and drums). Results indicated that misinformation regarding the presence of piano had a significant effect on participants' performance in the recognition task, whereas misinformation suggesting the presence of drums did not. A potential explanation for this could be that drums are perceived as a louder and more prominent instrument than piano, occupying a larger range of the frequency spectrum in a recorded track. Because drums are perceptually a more obvious element to be missing or introduced into a musical piece, this timbral manipulation may be less susceptible to

misinformation effects compared to a less prominent timbral manipulation, such as piano. However, this distinction on the two types of timbral manipulation did not affect the overall misinformation effect in the general model (including both piano and drums). This might be due to a clear trend supporting the misinformation effect in the drums manipulation as well (Figure 3).

The present research focused on musical timbre and instrumentation, which has been identified as a particularly important surface feature influencing the ability to remember music (Halpern & Müllensiefen, 2008; Poulin-Charronnat et al., 2004; Schellenberg & Habashi, 2015; Trainor et al., 2004). For example, research shows that participants' long-term memory for music deteriorates significantly when the instrumentation used at the recognitions test phase does not match the one used at the learning phase (Halpern & Müllensiefen, 2008; Poulin-Charronnat et al., 2004). In addition, there is evidence that information about the key and tempo is forgotten at a faster rate than information about the timbre (Schellenberg & Habashi, 2015), with infants even retaining memory of timbre after 1 week of daily exposure to music (Trainor et al., 2004). Results from our study indicate that post-event information about instrumentation significantly interfered with the actual surface information encoded at the exposure phase.

When creating an adapted musical misinformation paradigm, there were practical reasons to manipulate timbre instead of other surface features. In order to manipulate misinformation in a given piece of music, a musical feature was required that could be addressed specifically in the post-event descriptive text and could be later scored unambiguously as correct/ incorrect in the recognition test. This is most easily done by using a categorical feature such as the presence of an instrument which most people, even without musical training, are able to identify. Extracting other types of

information from a musical excerpt (e.g., key, mode, or tempo) requires special musical skills or training and would have limited our choice of participants and, therefore, the wider applicability of this study. Moreover, manipulating the presence or absence of an instrument in a piece of music is analogous to the missing objects manipulated in Loftus et al. (1978).

Nevertheless, there is a wide range of other musical aspects that could be manipulated, if these can be made congruent with the practical constraints of designing a task that does not require special musical trainings and skills. For example, the intonation accuracy of the vocalist, appropriateness of tempo, expressivity, emotionality, overall quality, and aesthetical aspects of the performance. Moreover, when creating false memories of music listening episodes, the perceived familiarity of the tracks and the jazz genre used may have played a role. Jazz music remains a relatively specialist and less popular music genre in the West, with jazz album sales accounting for only 2.1% of the total albums sold in the USA in 2014 (Nielsen, 2015). Thus, replications of this study should consider the use of a wider range of types of misinformation as well as music genres and styles in order to bolster the effects observed, and allow for a more nuanced understanding of what specific factors contribute to the fallibility of musical memory. Looking beyond misinformation there are clearly more experimental paradigms that could be employed to study false memory in music. For instance, Curtis and Bharucha (2009) and et al. (2014) studied false memories in music using an adapted version of Roediger and McDermott's paradigm (1995). The authors found that participants falsely remembered more notes in situations where the target note was congruent with the context (more expected) than when it was incongruent (less expected).

The existence of misinformation effects in music listening situations sheds light into the nature of long-term memory for music. Our results suggest that memory for music is, to some extent, malleable and susceptible to post-event misinformation about surface features, such as timbre. Participants failed to accurately monitor the source of information, misattributing information from the musical source to the verbal one. This finding is in line with the *source-monitoring framework* (SMF; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 2000) and *fuzzy trace theory* (FTT; Brainerd & Reyna, 2002, 2004; Reyna & Brainerd, 1995). According to FTT, there are two parallel types of memory, namely, verbatim and gist. While verbatim memory represents the surface details of physical stimuli, gist memory represents the main meaning or theme. These two types of memories are encoded separately and can be retrieved independently (Brainerd & Reyna, 2002). Thus, false musical memory may have occurred because verbatim (surface features) declined faster than gist (abstract structure) and integrated with schematic-gist information (Brainerd & Reyna, 2002). Although providing a comprehensive theoretical account of the effect of misinformation in music is beyond the scope of this study, we consider that efforts in this direction are essential to demonstrate the generality of theoretical accounts of false memory and general principles of memory to the non-visual auditory domain.

Finally, the present study aimed to explore potential individual difference factors related to the susceptibility to false memories in music listening. Based on previous literature (see Eisen et al., 2002; Loftus, 2005; Talmi et al., 2017; Zhu et al., 2010, for reviews), we explored the role of age, musical training, suggestibility, and personality traits. Contrary to our hypotheses, we found no evidence to support that any of these individual difference factors were significantly associated with misinformation susceptibility in music. It is important to mention, however, that this analysis was

purely exploratory and one would need to devise and carefully calibrate a proper individual difference test of misinformation susceptibility in music in order to confirm and generalise these results. Moreover, we did not investigate a sample of professional musicians and this finding may change when testing individuals with greater musical expertise. However, this result is in line with Anglada-Tort and Müllensiefen (2017), who also showed that musical expertise did not have a protective effect against a musical memory illusion. It also supports Schiavio and Timmers (2016), who did not find an advantage of musicians' long-term memory over nonmusicians.

Overall, our findings support, at least partly, previous research that post-event misinformation has a significant effect on the reliability of memory, suggesting that false memory can be induced in music listening tasks. When people listen to music or experience music in a live performance, they are normally exposed to related information at some point after the event. In our experimental setting, the presence of post-event misinformation about instrumentation impaired listeners' ability to remember music. Participants used verbal information that was never actually experienced to reconstruct a memory of a piece of music, demonstrating the generality of the misinformation effect to the non-visual auditory domain. Furthermore, a random forest analysis indicated that the misinformation effect occurred regardless of participants' levels of musical training, suggestibility, age, and personality traits. The existence of misinformation effects in music listening situations has implications for any area in which musical memory is involved, including aesthetics, music education, performance evaluation, preferences for music, marketing, and advertising. We conclude that memory for non-visual auditory stimuli can be fallible and the extent to which humans can memorise and remember music reliably should be, at least, questioned and further investigated.

Acknowledgements

This work was supported by a PhD studentship from the “Studienstiftung des Deutschen Volkes” (Bonn, Germany) awarded to Manuel Anglada-Tort.

Disclosure of interest

The authors report no conflicts of interest.

References

- Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203-231. DOI: 10.1037/0033-2909.93.2.203
- Anglada-Tort, M., & Müllensiefen, D. (2017). The repeated recording illusion: The effects of extrinsic and individual difference factors on musical judgements. *Music Perception: An Interdisciplinary Journal*, *35*(1), 92-115. DOI: 10.1525/mp.2017.35.1.94
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. DOI: 10.18637/jss.v067.i01
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Barnier, A. J., & McConkey, K. M. (1992). Reports of real and false memories: The relevance of hypnosis, hypnotizability, and context of memory test. *Journal of Abnormal Psychology*, *101*(3), 521.

- Bernstein, D., & Loftus, E. F. (2009a). How to tell if a particular memory is true or false. *Perspectives on Psychological Science*, 4, 370–374.
- Bernstein, D. M., & Loftus, E. F. (2009b). The consequences of false memories for food preferences and choices. *Perspectives on Psychological Science*, 4(2), 135-139.
- Bernstein, D. M., Laney, C., Morris, E. K., & Loftus, E. F. (2005). False memories about food can lead to food avoidance. *Social Cognition*, 23(1), 11-34. DOI: 0.1521/soco.23.1.11.59195
- Bernstein, D., Scoboria, A., & Arnold, R. (2015). The consequences of suggesting false childhood food events. *Acta Psychologica*, 156, 1–7.
- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 11(5), 164-169.
- Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford: Oxford University Press.
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS one*, 8(4), e60676.
- Bittner, R., Salamon, J., Tierney, M., Mauch, M., Cannam, C. & Bello, J.P. (2014, October). *MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research* (pp.155-160). Paper presented at the 15th International Society for Music Information Retrieval Conference, Taipei, Taiwan.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. DOI: 10.1023/A:1010933404324
- Cialdini, R.B. (2001). Harnessing the Science of Persuasion. *Harvard Business Review*, 79(9), 72-81.

- Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic bulletin & review*, *18*(3), 586-591.
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, *4*(1), 5-13.
- Curtis, M. E., & Bharucha, J. J. (2009). Memory and musical expectation for tones in cultural context. *Music Perception: An Interdisciplinary Journal*, *26*(4), 365-375. DOI: 10.1525/mp.2009.26.4.365
- Davis, D., & Loftus, E. F. (2007). Internal and external sources of misinformation in adult witness memory. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. Lindsay (Eds.), *The handbook of eyewitness psychology, Vol. 1. Memory for events* (pp. 195-237). Mahwah, NJ: Lawrence Erlbaum Associates.
- Delhasse, P. (1985). *Le Concours Reine Elisabeth des Origines Aujord'hui*. Bruxelles: Vander.
- Eisen, M. L., Winograd, E., & Qin, J. (2002). Individual differences in adults' suggestibility and memory performance. In M. L. Eisen, J. A. Quas, & G. S. Goodman (Eds.), *Memory and suggestibility in the forensic interview*. New York, NY: Routledge.
- Eren, P. E., Temizel, T. T., & Unal, P. (2014, May). An Exploratory Study on the Outcomes of Influence Strategies in Mobile Application Recommendations. In *Proceedings of the Second International Workshop on Behavior Change Support Systems (BCSS2014), Padova, Italy*.

- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.
- Flôres Jr, R. G., & Ginsburgh, V. A. (1996). The Queen Elisabeth musical competition: how fair is the final ranking?. *The Statistician*, 97-104. DOI: 10.2307/2348415
- Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graces, S., Heiberger, R., Monette, G., Murdoch, D., Nilsson, H., Ogle, D., Ripley, B., Venable, W., Winsemius, D., Zeileis (2017). The car package. R Package Version 2.1-6. Retrieved from <http://ftp.uni-bayreuth.de/math/statlib/R/CRAN/doc/packages/car.pdf>
- Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science*, *20*(1), 20-23.
- Fruehwald, E. S. (1992). Copyright infringement of musical compositions: A systematic approach. *Akron Law Review*, *26*(1), 15-44.
- Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology*, *17*(5), 533-543. DOI: 10.1002/acp.885
- Graham, D. (2008, December 04). Ex-Thin Lizzy guitarist loses German plagiarism case. Retrieved from <https://www.reuters.com/article/us-moore/ex-thin-lizzy-guitarist-loses-german-plagiarism-case-idUSTRE4B28HQ20081203>
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26. DOI: 10.1037/0003-066X.48.1.26
- Gordon, E. E. (1989). *Advanced measures of music audiation*. Chicago, IL: GIA Publications.

- Halpern, A. R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *The Quarterly Journal of Experimental Psychology*, *61*(9), 1371-1384. DOI: 10.1080/17470210701508038
- Harrison, P. M., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: item response theory, computerised adaptive testing, and automatic item generation. *Scientific reports*, *7*(1), 3618.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Hierarchical clustering. In T. Hastie, E. Tibshirani, & J. Friedman (Eds.), *The elements of statistical learning: Data mining, inference and prediction* (2nd ed., pp. 520-528). New York: Springer.
- Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, *7*, 355-373. DOI: 10.1093/biostatistics/kxj011
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*, 651-674. DOI: 10.1198/106186006X133933
- Hyman, I. E., Husband, F., & Billings, J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology*, *9*, 181-197. DOI:10.1002/acp.2350090302
- Hyman, I. E. Jr., & Kleinknecht, E. (1999). False childhood memories: Research, theory, and applications. In L. M. Williams & V. L. Banyard (Eds.), *Trauma and memory* (pp. 175-188). Thousand Oaks, CA: Sage.
- Janitza, S., Strobl, C., & Boulesteix, A.-L. (2013). An AUC- based permutation variable importance measure for random forests. *BMC Bioinformatics*, *14*, 119. DOI: 10.1186/1471-2105-14-119

- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, 2, 102-138.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin*, 114(1), 3-28.
- Johnson, M. K., & Raye, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain and beliefs* (pp. 35-86). Cambridge, MA, USA: Harvard University Press.
- Kaptein, M., De Ruyter, B., Markopoulos, P., & Aarts, E. (2012). Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(2), 10.
- Kuhn, M. (2008). Caret package. *Journal of statistical software*, 28(5), 1-26. Retrieved from <http://www.download.nextag.com/cran/web/packages/caret/caret.pdf>
- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, 56(4), 414-423. DOI: 10.3758/BF03206733
- Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 58(6), 927-935. DOI: 10.3758/BF03205494
- Liebman, J. I., McKinley-Pace, M. J., Leonard, A. M., Sheesley, L. A., Gallant, C. L., Renkey, M. E., & Lehman, E. B. (2002). Cognitive and psychosocial correlates of adults' eyewitness accuracy and suggestibility. *Personality and Individual Differences*, 33(1), 49-66. DOI: 10.1016/S0191-8869(01)00135-0

- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361–366. DOI: 10.1101/lm.94705
- Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118*(1), 100–104. DOI: 10.1037/0096-3445.118.1.100
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*(1), 19–31. DOI: 10.1037/0278-7393.4.1.19
- Loftus, E. F., Schooler, J. W., & Wagenaar, W. A. (1985). The fate of memory: Comment on McCloskey and Zaragoza. *Journal of Experimental Psychology: General, 114*(3), 375–380. DOI: 10.1037/0096-3445.114.3.375
- McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114*(1), 1–16.
- Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics, 41*(6), 576–600.
- 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. DOI: 10.1371/journal.pone.0089642
- Müllensiefen, D., & Pendzich, M. (2009). Court decisions on music plagiarism and the predictive value of similarity algorithms. *Musicae Scientiae, 13*, 257–295.
- Neuschatz, J. S., Lampinen, J. M., Togliola, M. P., Payne, D. G., & Cisneros, E. P. (2017). False memory research: History, theory, and applied implications. In

M. P. Toglia, J. D., Read, D. F. Ross, & R. C. Lindsay (Eds.), *The Handbook of Eyewitness Psychology: Volume I: Memory for Events* (pp. 239-260).

Mahwah, NJ: Lawrence Erlbaum Associates.

Paulhus D.L. (1986) Self-Deception and Impression Management in Test Responses.

In: Angleitner A., Wiggins J.S. (eds) *Personality Assessment via Questionnaires*. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-70751-3_8

Pickrell, J. E., McDonald, D., Bernstein, D. M., & Loftus, E. F. (2016). Misinformation effect. In R. F. Pohl (Ed.) *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking and Memory* (pp. 406-423). London: Psychology Press.

Porter, S., Birt, A. R., Yuille, J. C., & Lehman, D. R. (2000). Negotiating false memories: Interviewer and rememberer characteristics relate to memory distortion. *Psychological Science*, 11(6), 507-510.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803-814. DOI: 10.1037/0278-7393.21.4.803

Stuart, E. A., King, G., Imai, K., & Ho, D. E. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8).

Qualtrics, L. L. C. (2014). Qualtrics: Online Survey Software & Insight Platform. Retrieved from <http://qualtrics.com>.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PloS one*, 5(6), e11120.

- Peretz, I., Gaudreau, D., & Bonnel, A. M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26*(5), 884-902.
- Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89-114. DOI: 10.1146/annurev.psych.56.091103.070225.
- Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, 3-56.
- Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., & McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Perception: An Interdisciplinary Journal*, *22*(2), 239-263.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1-75.
- Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & cognition*, *43*(7), 1021-1031.
- Schellenberg, E. G., Stalinski, S. M., & Marks, B. M. (2014). Memory for surface features of unfamiliar melodies: Independent effects of changes in pitch and tempo. *Psychological research*, *78*(1), 84-95.
- Schiavio, A., & Timmers, R. (2016). Motor and audiovisual learning consolidate auditory memory of tonally ambiguous melodies. *Music Perception: An Interdisciplinary Journal*, *34*(1), 21-32.
- Schooler, J. W., & Loftus, E. F. (1993). Multiple mechanisms mediate individual differences in eyewitness accuracy and suggestibility. *Mechanisms of everyday cognition*, 177-203.

- Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, *33*(1), 189-196.
- Scoboria, A., Wade, K. A., Lindsay, D. S., Azad, T., Strange, D., Ost, J., & Hyman, I. E. (2017). A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies. *Memory*, *25*(2), 146–163. DOI: 10.1080/09658211.2016.1260747
- Shaw, J. (2016). *The Memory Illusion: Remembering, Forgetting, and the Science of False Memory*. New York, NY: Random House.
- Smelser, N. J., & Baltes, P. B. (Eds.). (2001). *International encyclopedia of the social & behavioral sciences* (Vol. 11). Amsterdam: Elsevier.
- Stöber, J. (1999). Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität [The Social Desirability Scale-17 (SDS-17): Development and first results on reliability and validity]. *Diagnostica*, *45*, 173-177.
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, *17*, 222-232. DOI: 10.1027//1015-5759.17.3.222
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 307. DOI: 10.1186/1471-2105-9-307
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*, 323–348. DOI: 10.1037/a0016973

- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLoS ONE*, *12*(10), e0186773. DOI: 10.1371/journal.pone.0186773
- Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*(3), 289-296.
- Vuvan, D. T., Podolak, O. M., & Schmuckler, M. A. (2014). Memory for musical tones: the impact of tonality and the creation of false memories. *Frontiers in Psychology*, *5*. DOI: 10.3389/fpsyg.2014.00582
- Wade, K. A., Garry, M., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin & Review*, *9*, 597–603. DOI: 10.3758/BF03196318
- Ward, R. A., & Loftus, E. F. (1985). Eyewitness performance in different psychological types. *The Journal of General Psychology*, *112*(2), 191-200.
- Weingardt, K. R., Toland, H. K., & Loftus, E. F. (1994). Reports of suggested memories: Do people truly believe them? In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 3-26). Cambridge: Cambridge University Press.
- Weiss, M. W., Vanzella, P., Schellenberg, E. G., & Trehub, S. E. (2015). Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *The Quarterly Journal of Experimental Psychology*, *68*(5), 866-877.
- Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & Cognition*, *38*(2), 163-175.

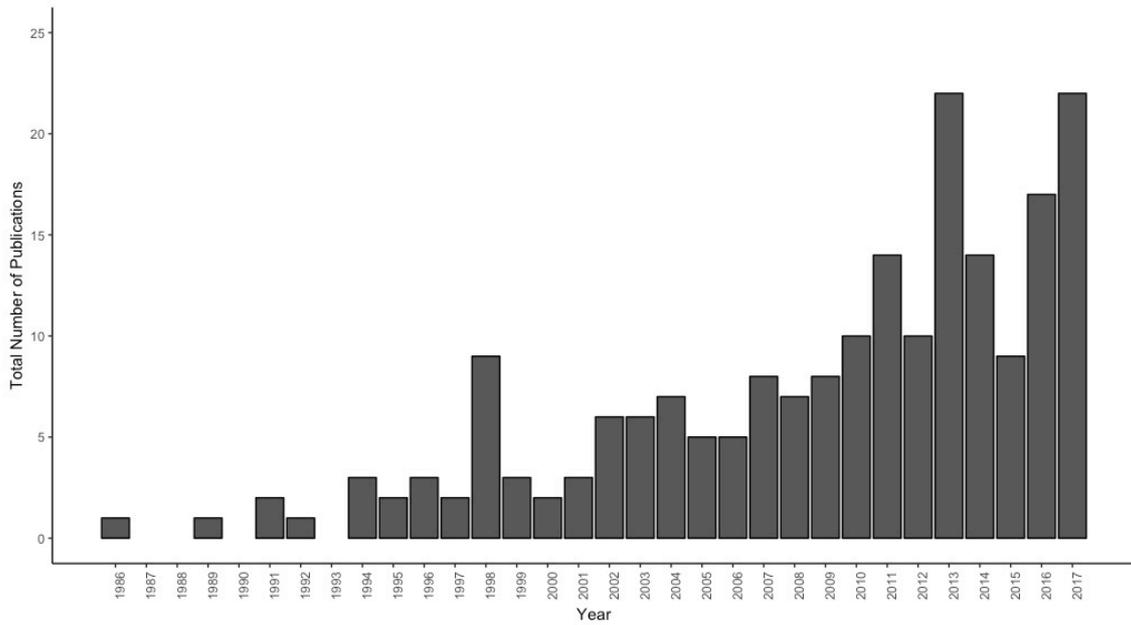
- Wylie, L. E., Patihis, L., McCuller, L. L., Davis, D., Brank, E., Loftus, E. F., & Bornstein, B. (2014). Misinformation effect in older versus younger adults: A meta-analysis and review. In M. P. Toglia, D. F. Ross, J. Pozzulo, & E. Pica (Eds.), *The Elderly Eyewitness in Court* (pp. 38-66). London: Psychology Press.
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., & Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory*, *18*(5), 543-555.
- Ziegler, M., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality*, *46*(2), 173-183. DOI: 10.1016/j.jrp.2012.01.002

Table 1. Summary of the logistic regression analyses before and after the matching

	β	SE β	Wald's χ^2	df	<i>p</i>
<i>Model before matching</i>					
Constant	1.92	1.18	1.62	1	.10
Age	-.11	.03	-3.09	1	.002
Musical Training	.03	.02	1.89	1	.06
<i>Model after matching</i>					
Constant	.81	1.26	.64	1	.52
Age	-.06	.04	-1.71	1	.09
Musical Training	.03	.02	1.97	1	.05

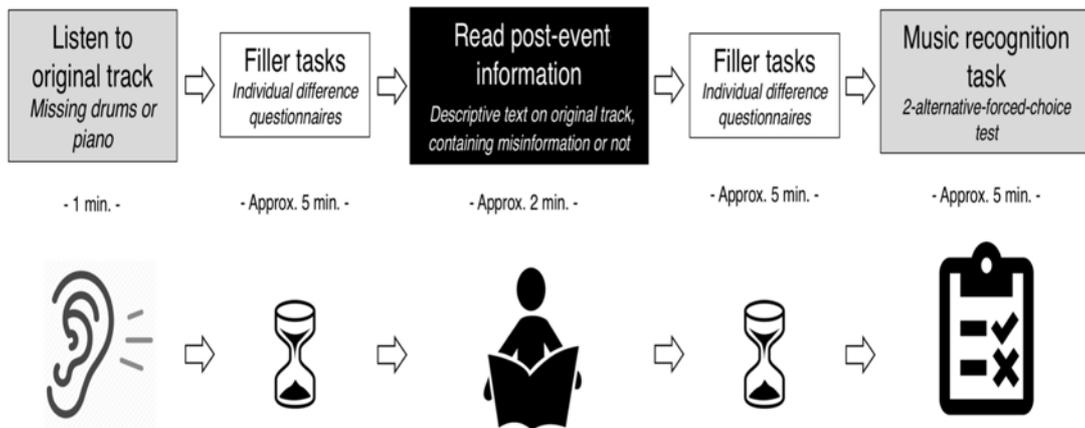
procedure.

Figure 1. Total number of publications on the misinformation effect from 1986 to 2017.



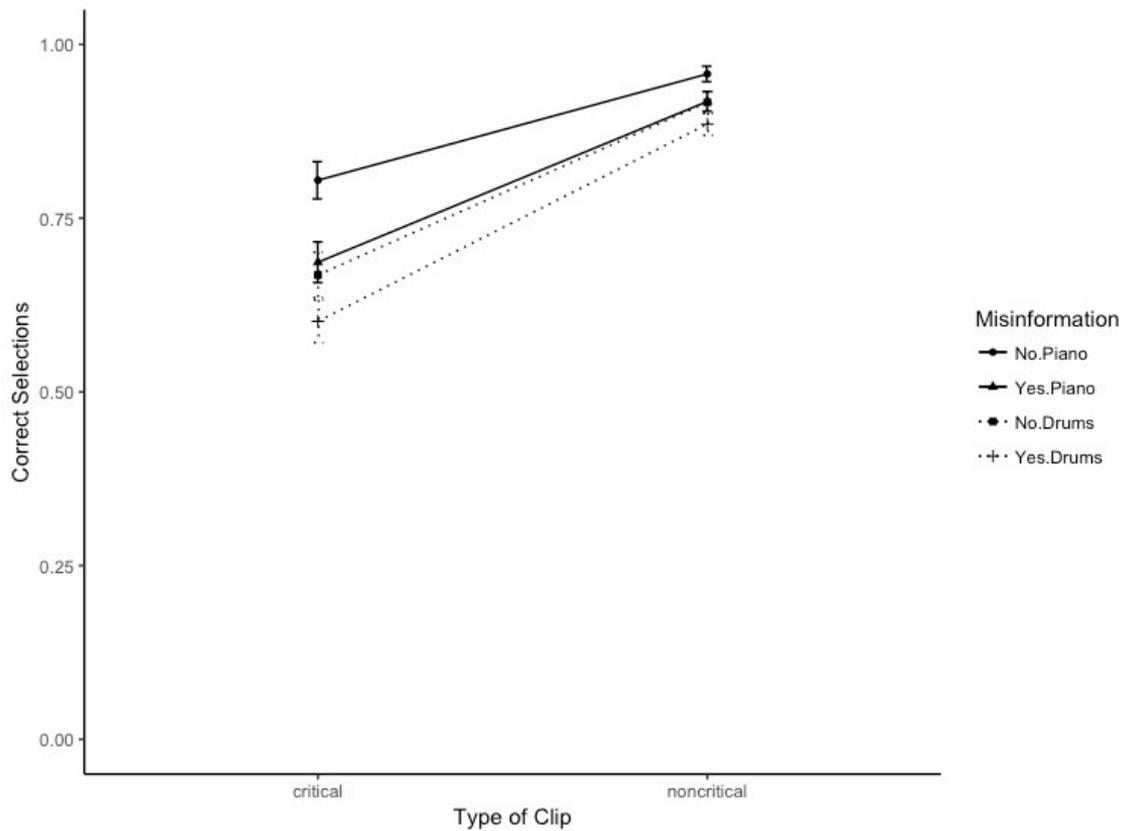
**Note. Data retrieved from Scopus on the 4th of July 2018. We searched for all available literature containing the keyword “misinformation effect*” in the title, abstract, or author keywords. Only peer-reviewed articles published in English from 1986 to 2017 were included, resulting in a total of 203 publications.*

Figure 2. Diagram of the misinformation paradigm procedure used in the experiment.



**Note. The entire procedure was repeated once after a brief break, using a new track as well as a different instrumental manipulation (drums or piano).*

Figure 3. Proportion of correct selections in the presence and absence of misinformation (yes vs. no) in the two types of clips (critical vs. noncritical) and the two timbral manipulations (piano vs. drums).



**Note. Error Bars Represent the Standard Error of proportion.*

Appendix A. Descriptive texts used in all music conditions in the two groups.

Descriptive Text Misinformation Group – Bebop Jazz/ Drums Manipulation.

The musical piece was a contemporary performance simply entitled “Bebop Jazz”, from the "Music Delta” collection. The song begins quickly, with staccato stabs and drums accompaniment, before the track moves into a rolling bassline groove. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1)".

(1) Please see <http://www.musicdelta.com/> for more information

To confirm you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins quickly, with _____ stabs and drums accompaniment"

Descriptive Text Misinformation Group – BeBop Jazz/ Piano Manipulation.

The musical piece was a contemporary performance simply entitled “Bebop Jazz”, from the "Music Delta” collection. The song begins quickly, with staccato stabs and piano accompaniment, before the track moves into a rolling bassline groove. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1).

(1) Please see <http://www.musicdelta.com/> for more information

To confirm you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins quickly, with _____ stabs and piano accompaniment"

Descriptive Text Misinformation Group – Cool Jazz/ Drums Manipulation.

The musical piece was a contemporary performance simply entitled “Cool Jazz”, from the "Music Delta" collection. The song begins slowly, with a gentle groove and drums accompaniment, before moving into a partly-improvised jazz performance. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1).

(1) Please see <http://www.musicdelta.com/> for more information

To confirm, you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins slowly, with a _____ groove and drums accompaniment"

Descriptive Text Misinformation Group – Cool Jazz/ Piano Manipulation.

The musical piece was a contemporary performance simply entitled “Cool Jazz”, from the "Music Delta" collection. The song begins slowly, with a gentle groove and piano accompaniment, before moving into a partly-improvised jazz performance. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1).

(1) Please see <http://www.musicdelta.com/> for more information

To confirm, you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins slowly, with a _____ groove and piano accompaniment"

Descriptive Text Control Group – Bebop Jazz/ Drums Condition.

The musical piece was a contemporary performance simply entitled “Bebop Jazz”, from the “Music Delta” collection. The song begins quickly, with staccato stabs, before the track moves into a rolling bassline groove. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1).

(1) Please see <http://www.musicdelta.com/> for more information

To confirm you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins quickly, with _____ stabs"

Descriptive Text Control Group - Bebop Jazz/ Piano Manipulation.

The musical piece was a contemporary performance simply entitled “Bebop Jazz”, from the “Music Delta” collection. The song begins quickly, with staccato stabs, before the track moves into a rolling bassline groove. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1).

(1) Please see <http://www.musicdelta.com/> for more information

To confirm you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins quickly, with _____ stabs"

Descriptive Text Control Group – Cool Jazz/ Drums Manipulation.

The musical piece was a contemporary performance simply entitled “Cool Jazz”, from the "Music Delta" collection. The song begins slowly, with a gentle groove, before moving into a partly-improvised jazz performance. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1).

(1) Please see <http://www.musicdelta.com/> for more information

To confirm, you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins slowly, with a _____ groove"

Descriptive Text Control Group – Cool Jazz - Piano Condition.

The musical piece was a contemporary performance simply entitled “Cool Jazz”, from the "Music Delta" collection. The song begins slowly, with a gentle groove, before moving into a partly-improvised jazz performance. The excerpt listened to was 60 seconds long. More information regarding Music Delta and the range of musical resources they have available may be found at their website (1).

(1) Please see <http://www.musicdelta.com/> for more information

To confirm, you have read the overview, please type the word that should appear in the gap below, as per the text above. You may then proceed with the experiment.

"The song begins slowly, with a _____ groove"

