

Do Musicians Have Better Short-Term Memory Than Nonmusicians? A Multilab Study



Advances in Methods and
Practices in Psychological Science
October-December 2025, Vol. 8, No. 4,
pp. 1–30
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25152459251379432
www.psychologicalscience.org/AMPPS



Massimo Grassi¹, Francesca Talamini², Gianmarco Altoè³,
Elvira Brattico^{4,5}, Anne Caclin⁶, Barbara Carretti¹,
Véronique Drai-Zerbib⁷, Laura Ferreri⁸, Filippo Gambarota³,
Jessica Grahn^{9,10}, Lucrezia Guiotto Nai Fovino¹, Marco Roccato¹,
Antoni Rodriguez-Fornells^{11,12}, Swathi Swaminathan^{9,10},
Barbara Tillmann⁷, Peter Vuust^{4,13}, Jonathan Wilbiks¹⁴,
Marcel Zentner², Karla Aguilar¹⁴, Christ B. Aryanto^{15,16},
Frederico C. Assis Leite¹⁷, Aíssa M. Baldé¹⁸, Deniz Başkent¹⁹,
Laura Bishop^{20,21}, Graziela Kalsi²², Fleur L. Bouwer²³,
Axelle Calcus²⁴, Giulio Carraturo⁵, Victor Cepero-Escribano^{25,12},
Antonia Čerič²⁶, Antonio Criscuolo²⁷, Léo Dairain²⁴,
Simone Dalla Bella^{28,29,30,31}, Oscar Daniel³², Anne Danielsen^{20,21},
Anne-Isabelle de Parcevaux⁷, Delphine Dellacherie³³, Tor Endestad^{34,20,35},
Juliana L. d. B. Fialho^{22,36}, Caitlin Fitzpatrick⁹, Anna Fiveash^{37,7},
Juliette Fortier²⁸, Noah R. Fram³⁸, Eleonora Fullone⁸,
Stefanie Gloggengießer²⁶, Lucia Gonzalez Sanchez¹², Reyna L. Gordon^{38,39},
Mathilde Groussard⁴⁰, Assal Habibi⁴¹, Heidi M. U. Hansen³⁴,
Eleanor E. Harding¹⁹, Kirsty Hawkins⁴², Steffen A. Herff^{43,44},
Veikka P. Holma⁴⁵, Kelly Jakubowski³², Maria G. Jol²³, Aarushi Kalsi³²,
Veronica Kandro⁴⁶, Rosaliina Kelo⁴⁷, Sonja A. Kotz^{27,48},
Gangothri S. Ladegam⁴⁹, Bruno Laeng^{34,20}, André Lee^{50,51}, Miriam Lense³⁸,
César F. Lima¹⁸, Simon P. Limmer², Chengran K. Liu⁴⁹,
Paulina d. C. Martín Sánchez⁵², Langley McEntyre³⁸, Jessica P. Michael⁴⁹,
Daniel Mirman⁴⁹, Daniel Müllensiefen⁴², Niloufar Najafi²³,
Jaakko Nokkala⁴⁷, Ndassi Nzonlang³³, Maria Gabriela M. Oliveira¹⁷,
Katie Overy^{53,54}, Andrew J. Oxenham⁵⁵, Edoardo Passarotto⁵⁰,
Marie-Elisabeth Plasse⁶, Herve Platel⁴⁰, Alice Poissonnier⁴⁰, Neha Rajappa⁵⁵,
Michaela Ritchie¹⁴, Italo Ramon Rodrigues Menezes¹⁷,
Rafael Román-Caballero^{52,56}, Paula Roncaglia⁵⁷, Farrah Y. Sa'adullah³⁷,
Suvi Saarikallio⁴⁵, Daniela Sammler^{26,48}, Séverine Samson^{33,58},
E. G. Schellenberg^{18,59}, Nora R. Serres^{34,20}, L. R. Slevc⁴⁶,
Ragnya-Noraso Souffiane⁶, Florian J. Strauch², Hannah Strauss²,
Nicholas Tantengco⁴⁶, Mari Tervaniemi^{47,60}, Rachel Thompson⁴⁶,
Renee Timmers⁶¹, Petri Toiviainen^{45,62}, Laurel J. Trainor^{63,64},
Clara Tuske⁴⁶, Jed Villanueva⁴¹, Claudia C. von Bastian¹⁵, Kelly L.
Whiteford⁵⁵, Emily A. Wood⁶³, Florian Worschech⁵⁰, and Ana Zappa¹²



¹Department of General Psychology, University of Padua, Padua, Italy; ²Department of Psychology, University of Innsbruck, Innsbruck, Austria; ³Department of Developmental Psychology and Socialization, University of Padua, Padua, Italy; ⁴Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University, Denmark; ⁵Department of Education, Psychology, Communication, University of Bari Aldo Moro, Bari, Italy; ⁶Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, F-69500, Bron, France; ⁷Université Bourgogne Europe, CNRS, LEAD UMR5022, 21000, Dijon, France; ⁸Department of Brain & Behavioural Science, Università degli Studi di Pavia, Pavia, Italy; ⁹Department of Psychology, Western University, London, Ontario, Canada; ¹⁰Brain and Mind Institute, Western University, London, Ontario, Canada; ¹¹Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain; ¹²Department of Cognition, Development and Education Psychology, University of Barcelona / Institut d'Investigació Biomèdica de Bellvitge, Barcelona, Spain; ¹³Royal Academy of Music Aarhus/Aalborg, Denmark; ¹⁴Department of Psychology, University of New Brunswick Saint John, 100 Tucker Park Road, Saint John, NB, Canada; ¹⁵Department of Psychology, University of Sheffield, Sheffield, United Kingdom; ¹⁶Faculty of Psychology, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia; ¹⁷Department of Psychobiology, Universidade Federal de São Paulo, São Paulo, Brazil; ¹⁸Centre for Psychological Research and Social Intervention (CIS-Iscte), Iscte - Instituto Universitário de Lisboa, Lisboa, Portugal; ¹⁹Department of Otorhinolaryngology / Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, Netherlands; ²⁰RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway; ²¹Department of Musicology, University of Oslo, Oslo, Norway; ²²Department of Music, Universidade Estadual Paulista "Júlio de Mesquita Filho", São Paulo, Brazil; ²³Cognitive Psychology Unit, Institute of Psychology & Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands; ²⁴Center for Research in Cognitive Neuroscience (CRCN), Université libre de Bruxelles (ULB), Brussels, Belgium; ²⁵Bridging Research in AI and Neuroscience (BRAIN), Computer Vision Center (CVC), Bellaterra, Spain; ²⁶Research Group Neurocognition of Music and Language, Max Planck Institute for Empirical Aesthetics, Frankfurt/M., Germany; ²⁷Department of Neuropsychology & Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands; ²⁸International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, Canada; ²⁹Department of Psychology, University of Montreal, Montreal, Canada; ³⁰Centre for Research on Brain, Language and Music (CRBLM), Montreal, Canada; ³¹Department of Cognitive Psychology, Vizja University, Warsaw, Poland; ³²Department of Music, Durham University, Durham, United Kingdom; ³³Département de Psychologie, Laboratoire PSITEC ULR 4072, Université de Lille, Lille, France; ³⁴Department of Psychology, Faculty of Social Sciences, University of Oslo, Oslo, Norway; ³⁵Department of Neuropsychology, Helgeland Hospital, Mosjøen, Norway; ³⁶Department of Psychology, Universidade Cruzeiro do Sul, São Paulo, Brazil; ³⁷The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia; ³⁸Department of Otolaryngology - Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, United States; ³⁹Vanderbilt Genetics Institute, Vanderbilt University, Nashville, United States; ⁴⁰Department of Psychology, Université de Caen Normandie, Inserm, EPHE-PSL, PSL Université Paris, CHU de Caen, GIP Cyceron, U1077, NIMH, 14000 Caen, France; ⁴¹Brain and Creativity Institute, Dornsife College of Letters Arts and Sciences, University of Southern California, United States; ⁴²Department of Psychology, Goldsmiths, University of London, London, United Kingdom; ⁴³Sydney Conservatorium of Music, The University of Sydney, Sydney, Australia; ⁴⁴The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia; ⁴⁵Centre of Excellence in Music, Mind, Body and Brain, University of Jyväskylä, Jyväskylä, Finland; ⁴⁶Department of Psychology, University of Maryland, College Park, United States; ⁴⁷Centre of Excellence in Music, Mind, Body and Brain; Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland; ⁴⁸Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; ⁴⁹University of Edinburgh, Edinburgh, United Kingdom; ⁵⁰Institute of Music Physiology and Musicians' Medicine, Hanover University of Music, Drama, and Media, Hanover, Germany; ⁵¹Department of Neurology, TUM University Clinic Rechts der Isar, Technical University Munich, Germany; ⁵²Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain; ⁵³Reid School of Music, ECA, University of Edinburgh, Edinburgh, United Kingdom; ⁵⁴Edinburgh Neuroscience, University of Edinburgh, Edinburgh, United Kingdom; ⁵⁵Department of Psychology, University of Minnesota, Minneapolis, United States; ⁵⁶Department of Psychology, Neuroscience & Behaviour, McMaster University; ⁵⁷Cognitive Science and Artificial Intelligence Department, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands; ⁵⁸Institut Pasteur, Université Paris Cité, AP-HP, INSERM, CNRS, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, F-75012 Paris, France; ⁵⁹Department of Psychology, University of Toronto Mississauga, Mississauga, Canada; ⁶⁰Cognitive Brain Research Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland; ⁶¹Department of Music, University of Sheffield, Sheffield, United Kingdom; ⁶²Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland; ⁶³Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada; and ⁶⁴McMaster Institute for Music and the Mind, McMaster University

Abstract

Musicians are often regarded as a positive example of brain plasticity and associated cognitive benefits. This emerges when experienced musicians (e.g., musicians with more than 10 years of music training and practice) are compared with nonmusicians. A frequently observed behavioral finding is a short-term memory advantage of the former over the latter. Although available meta-analysis reported that the effect size of this advantage is medium (Hedges's $g = 0.5$), no literature study was adequately powered to estimate reliably an effect of such size. This multilab study has been ideated, realized, and conducted in lab by several groups that have been working on this topic. Our ultimate goal was to provide a community-driven shared and reliable estimate of the musicians' short-term memory advantage (if any) and set a method and a standard for future studies in neuroscience and psychology comparing musicians and nonmusicians. Thirty-three research units recruited a total of 600 experienced musicians and 600 nonmusicians, a number that is sufficiently large to estimate a small effect size (Hedges's $g = 0.3$) with a high statistical power (i.e., 95%). Subsequently, we measured the difference in short-term memory for musical, verbal, and visuospatial stimuli. We also looked at cognitive, personality, and socioeconomic factors that might mediate the difference. Musicians had better short-term memory than nonmusicians for musical, verbal, and visuospatial stimuli with an effect size of, respectively, Hedges's $g_s = 1.08$ (95% confidence interval [CI] = [0.94, 1.22]; large), 0.16 (95% CI = [0.02, 0.30]; very small), and 0.28 (95% CI = [0.15, 0.41]; small). This work sets the basis for sound research practices in studies comparing musicians and nonmusicians and contributes to the ongoing debate on the possible cognitive benefits of musical training.

Keywords

musicians, music training, nonmusicians, multilab, cognitive abilities, open data, open materials, preregistration

Received 7/15/22; Revision accepted 7/23/25

Do experienced musicians have better short-term memory than nonmusicians? A recent meta-analysis (e.g., Talamini et al., 2017) said so, although none of the behavioral studies included in it had a sufficiently large number of participants to capture the size of the calculated advantage. Our study emerged from the shared need of the neuroscience and psychology-of-music community to shed light on this and similar issues, in which the possible effects of lifelong music training are investigated by comparing musicians and nonmusicians. With the present multilab study, conceptualized, realized, and conducted by several groups that worked on this topic in the recent past, we aimed to provide a sound answer to the above question by recruiting a number of participants that was large enough to provide a reasonable estimate of the advantage (if any). We also investigated possible interindividual differences in cognitive abilities, personality, and socioeconomic status (SES) of the family that may modulate this advantage.

In psychology and neuroscience, experts are studied to understand the effect of extensive, lifelong activity on behavior and brain plasticity (e.g., London taxi drivers, Maguire et al., 2000; Braille readers, Pascual-Leone & Torres, 1993; chess players, De Groot, 2014; athletes, for an overview, see Chang, 2014; Voss et al., 2010). A class of skilled individuals that is often studied is musicians. Musicians and their abilities have been studied since more than a century ago because the training to become

a musician is long, complex, and structured. It requires performing fine motor actions and developing fine auditory skills and multisensory integration. Noticeably, musicians are required to coordinate all these abilities in synchrony. Furthermore, music has its own language and theory, which varies across musical cultures: Musicians may be required to learn to read the music notation and the rules of melody, harmony, and temporal organization (Hannon & Trainor, 2007). In addition, music is taught in dedicated schools, and the study of music can end with a high education degree, such as a bachelor's or a master's.

In line with studies that have compared experts and nonexperts in other domains (De Groot, 2014; Maguire et al., 2000; Pascual-Leone & Torres, 1993), musicians outperform nonmusicians when tasks tap into the domain of their expertise. For example, musicians are better than nonmusicians in understanding whether a melody is presented in transposition (Halpern et al., 1995) or at a faster or slower tempo (Andrews et al.,

Corresponding Authors:

Massimo Grassi, Department of General Psychology, University of Padua, Padua, Italy
Email: massimo.grassi@unipd.it

Francesca Talamini, Department of Psychology, University of Innsbruck, Innsbruck, Tirol, Austria
Email: Francesca.Talamini@uibk.ac.at

1998) and better in detecting mistuned notes in chords and melodies (Koelsch et al., 1999; Schellenberg & Moreno, 2010). The expertise of musicians often transfers to skills and tasks that are similar to musical skills and tasks. Musicians perform better than nonmusicians in classic psychoacoustic tasks, such as frequency (Micheyl et al., 2006) and temporal discrimination (Rammsayer & Altenmüller, 2006). Better auditory ability in musicians is also reflected in speech-perception tasks because musicians seem better than nonmusicians at judging the prosody of a sentence (Deguchi et al., 2012; Jansen et al., 2023; Schön et al., 2004) and, to some extent, perceiving speech in noise (Başkent & Gaudrain, 2016; Hennessy et al., 2022; Parbery-Clark et al., 2009).

Musicians also seem to perform better than nonmusicians in tasks that tap into general cognitive skills. One of the cognitive skills often investigated is short-term/working memory, for which musicians were found to perform better than nonmusicians. Here, we adopt the same terminology used by Talamini et al. (2017) to distinguish between short-term-memory (assessed with tasks that require to store and recall the stimuli verbatim, e.g., the forward digit span) and working-memory tasks (tapped by tasks that require to store and manipulate information during encoding and/or recalling, e.g., the backward digit span). In short-term-memory and/or working-memory tasks, musicians perform better than nonmusicians in tasks that require to memorize musical stimuli (i.e., stimuli within the domain of expertise; see Pallesen et al., 2010; Schulze et al., 2012; Schulze, Mueller, & Koelsch, 2011; Schulze, Zysset, et al., 2011), but they also do so in tasks that require memorizing and manipulating verbal stimuli, such as the digit span, nonword span, and operation span (e.g., Fennell et al., 2021; George & Coch, 2011; Hansen et al., 2013; Lee et al., 2007; Ramachandra et al., 2012; Talamini et al., 2016). In contrast, findings with visuospatial short-term and working-memory tasks are mixed (Amer et al., 2013; Criscuolo et al., 2019; Grassi et al., 2017; Hansen et al., 2013; Lee et al., 2007).

In a recent meta-analysis, Talamini et al. (2017) revealed that the memory advantage of musicians unfolds along two main dimensions: the type of stimuli (i.e., musical, verbal, visual/spatial) and the memory system (i.e., long, short, working memory). Musicians show a large advantage over nonmusicians when the test stimuli are music stimuli (e.g., melodies), a moderate advantage when the stimuli are verbal (e.g., digits, words), and a small to null advantage when the stimuli are visual and/or spatial. The meta-analysis revealed also that the advantage is of moderate effect size for working memory and short-term memory and of small effect size for long-term memory.

Why should musicians be better than nonmusicians in memory tasks that involve nonmusical stimuli? There are two main positions: one supporting the hypothesis of a cause-effect relationship (i.e., music training improves memory) and one supporting the hypothesis of preexisting differences (i.e., people with better memory and/or better general cognitive abilities are more likely to engage and succeed in music training). For individuals who support the first position (e.g., Herholz & Zatorre, 2012; Hyde et al., 2009; Jäncke, 2009; Münte et al., 2002; Schlaug, 2001; Strait & Kraus, 2014), music training represents a good to ideal model of brain plasticity (for a review on neuroanatomical differences between musicians and nonmusicians, see Criscuolo et al., 2022), and the better performance observed in memory tasks (with different materials) is due to training-related neuroplasticity. In contrast, proponents of the second position (e.g., Schellenberg, 2020; Swaminathan & Schellenberg, 2018, 2019) argue that individuals with a combination of traits—including above-average cognitive skills—are more likely to choose and succeed in long-term training (e.g., the training required to become a musician). As a result, when these individuals become musicians, they tend to outperform nonmusicians in cognitive tasks.

Regardless of either position, it is also possible that other variables explain/mediate the relationship between better short-term/working memory and musical expertise. For example, the memory advantage could be a by-product of an advantage in a higher-order cognitive skill, such as intelligence (e.g., fluid intelligence, Schellenberg, 2020), or executive functions (Okada & Slevc, 2018), or both (Criscuolo et al., 2019). The advantage may be related to individual differences in personality or sensitivity to music (e.g., enjoyment or pleasure derived from musical-related activities), which might have a genetic basis (e.g., Hansen et al., 2024). The length of music training is positively associated with the open-mindedness trait (Corrigall et al., 2013; Corrigall & Schellenberg, 2015): Musically trained individuals may perform well on experimental tasks because, at least in part, they tend to be curious and particularly interested in learning new things. The memory advantage observed in musicians may also be related to music aptitude (instead of music training), that is, the ability to perceive, remember, and discriminate melodies and rhythms (e.g., Swaminathan et al., 2017; Swaminathan & Schellenberg, 2018). Music aptitude may be just another declination of the so-called positive manifold (i.e., the positive association between the various abilities tested by intelligence tests; see Spearman, 1904). Music aptitude and short-term/working memory could be correlated, and individuals with music aptitude could be more likely to

undergo and succeed in music training and later become musicians (Swaminathan et al., 2017; Swaminathan & Schellenberg, 2018).

Finally, an open issue emerges across literature: Who can be labeled as a musician, and who cannot? Music-aptitude tests (Law & Zentner, 2012; Wallentin et al., 2010) and sophistication inventories (Müllensiefen et al., 2014) suggest a continuum, but the literature often treats musical expertise as binary, dividing participants into musicians and nonmusicians. However, the criteria to define both groups vary from study to study. Recently, Zhang and colleagues (Zhang et al., 2020; Zhang & Schubert, 2019; for a discussion, see also Baker et al., 2020) investigated the issue. Zhang and Schubert (2019) suggested a parsimonious tripartition of the music-expertise continuum and divided individuals among people with no musical identity (fewer than 6 years of music training), people with musical identity (6–10 years of music training), and people with strong musical identity (SMI; music training longer than 10 years).

Although some studies have reported that musicians are better than nonmusicians in memory tasks, this result may not be as clear as it may seem. In the meta-analysis by Talamini et al. (2017), the largest and most robust memory advantage was that for short-term-memory tasks. For these types of tasks, the mean effect size of the advantage was Hedges's $g = 0.57$,¹ and the effect size varied depending on the stimuli presented in the memory tasks (i.e., musical stimuli: $g = 1.15$; verbal stimuli: $g = 0.54$; visual and spatial stimuli: $g = 0.28$). If one calculates the number of participants that is needed to estimate an effect as such (i.e., $g = 0.57$) via a t test that compares the memory performance of musicians and nonmusicians and adopt ordinary standards for controlling for Type II error (e.g., a statistical power equal to .8; Cohen, 1988), the number of participants needed is 100 (50 per group). If one adopts contemporary standards for power (e.g., power equal to .90 or .95; see Sawilowsky, 2009), the number of participants per group rises to, respectively, 130 (65 per group) and 160 (80 per group). Figure 1 represents the number of participants that are needed to estimate various possible effect sizes with a statistical power of .80, .90, and .95 via an independent-sample t test and assuming a classic approach: only one laboratory collecting the data.

How many participants were recruited in the studies on musicians' memory that are available in literature? Far fewer. To our knowledge, the largest laboratory study (Okhrei et al., 2016) recruited 64 participants (28 musicians). Noticeably, the effect size calculated by Talamini et al. (2017) for short-term memory shows a strong negative correlation between number of participants and effect size observed by a given study (i.e.,

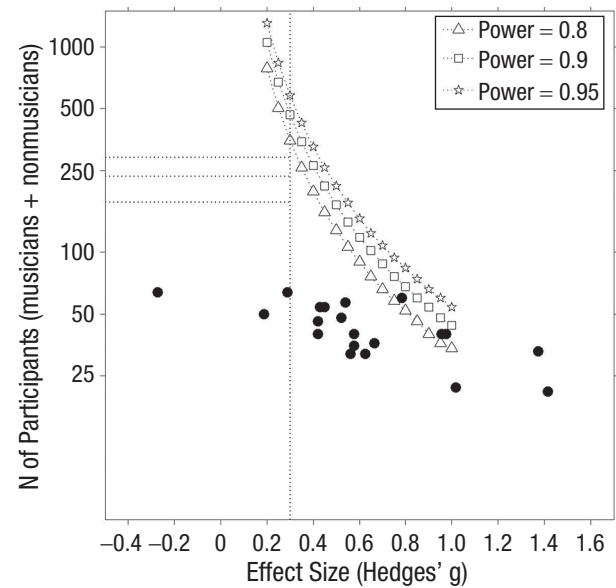


Fig. 1. Number of participants (musicians + nonmusicians) as a function of the effect size targeted by the study for three levels of statistical power. Note that the y-axis is logarithmic. The analysis hypothesizes groups of equal size and that the memory performance of the two groups is compared with a t test. It also hypothesizes that one single laboratory collects the data. The black circles represent the empirical studies included in the meta-analysis by Talamini et al. (2017): Each circle represents the number of participants and effect size observed by the study. Note how studies in the literature are underpowered to target the possible advantage of the musicians over nonmusicians. Note also how the size of the literature studies is negatively correlated with the effect size of the musicians' advantage ($r = -.70$): The smallest studies observed the largest effect size. The three horizontal dotted lines represent the number of participants one research unit needs to recruit to target, respectively, .80, .90, and .95 power with an effect size of 0.3, the effect size targeted in the current study.

$r = -.70$), a trend that is often observed in recent literature and that could hide statistical errors of both Type I and Type II (Cohen, 1988; Gelfand & Smith, 1990; Pastore et al., 2015).

In brief, before any theoretical discussion on the reasons for the short-term-memory advantage of musicians over nonmusicians, the studies conducted so far seem too small in terms of number of participants to assess a potential advantage. Moreover, there is a lack of systematic comparisons across stimuli, making it difficult to understand how the advantage might generalize for different types of materials. In the present study, we aim at revealing whether musicians have better short-term memory than nonmusicians with an adequately large sample size and how this advantage unfolds across different types of stimuli, while controlling for possible confounding variables. In addition, we aim at targeting a number of participants that is large enough to provide a good and reliable estimate of this advantage. Indeed,

Table 1. Expected Results and Effect Sizes of the Musicians' Short-Term-Memory Advantage

Question	Expected result	Interpretation given to different outcomes
1. Do musicians have better short-term memory for music stimuli than nonmusicians?	Musicians should outperform nonmusicians. The expected effect size is Hedges's $g \approx 1$.	The chances of a different outcome are unlikely.
2. Do musicians have better short-term memory for verbal stimuli than nonmusicians?	Musicians should perform better than nonmusicians. The expected effect size is Hedges's $g \approx 0.5$.	The advantage of musicians in verbal memory tasks is overestimated, negligible, or null.
3. Do musicians have better short-term memory for visuospatial stimuli than nonmusicians?	Musicians may perform better than nonmusicians. The expected effect size is Hedges's $g \approx 0.3$.	The advantage of musicians in visuospatial memory tasks is overestimated, negligible, or null.

Note: Because Question 3 requires the largest number of participants to be tested, in the current multilab study, we opted to recruit that number of participants to answer all three questions (see Power Analysis and Participants section).

any theoretical discussion about the possible beneficial effects of music training would be of little interest if the musicians' advantage is actually negligible in size.

Here, based on data and results in previous literature, we expected to observe an advantage of musicians in short-term-memory tasks. This advantage should be larger for music stimuli than for verbal stimuli. The smallest advantage (if any) should be observed for visuospatial stimuli. The effect size of this advantage could be smaller than that estimated by Talamini et al. (2017), a possibility supported by a recent study that compared effect sizes of meta-analyses with those of successive, empirical, large-N multilab registered reports replications investigating the same phenomena directly (see Kvarven et al., 2020). Because journals are more likely to publish false-positive than null results (e.g., Smaldino & McElreath, 2016) and because meta-analyses are calculated based on available literature (that might consequently be inflated by false positives), current meta-analyses might often overestimate the effect size of phenomena. Moreover, meta-analyses alone cannot overcome some of the limitations of the single studies included in them (e.g., heterogeneous inclusion/exclusion criteria, tools, stimuli). For a table showing this heterogeneity, see the supplemental material in the OSF repository. This table lists the main characteristics of the studies (e.g., characteristics of the musicians, the nonmusicians, the task) that compared the short-term memory of musicians and nonmusicians.

In the present study, we decided to target an effect size of Hedges's $g = 0.3$, which approximates the lowest effect size detected in the meta-analysis by Talamini et al. (2017) for visuospatial stimuli. For a summary of expected results see Table 1.

We also expected to replicate the positive associations observed in the literature, such as a positive

relationship between cognitive performance and music aptitude (e.g., Swaminathan & Schellenberg, 2018), cognitive performance and "open-mindedness" (e.g., Corrigan et al., 2013), and perhaps cognitive performance and the SES of the family (Osler et al., 2013). We also estimated the contribution of factors that might explain the memory advantage and that are suggested by the literature (e.g., intelligence, executive functions, and individual differences in sensitivity to musical experiences).

The present correlational study will not reveal whether music training may boost the short-term memory of individuals. However, current experimental research also seems unable to provide an answer to this question: Longitudinal studies that implement training of a sufficient duration to guarantee an SMI (i.e., 10 years or more of music training; Zhang & Schubert, 2019) are unlikely to ever be conducted. In the present study, nonetheless, we aimed to return a sound answer to the question whether musicians have better short-term memory than nonmusicians and, above all, the size of the advantage for different classes of stimuli (music, verbal, visuospatial).

Method

Ethics information

The research protocol complied with all relevant ethical regulations. The overarching ethical approval (No. 5305) was granted by the Ethics Committee (Comitato Etico della Ricerca Psicologica, Area 17) of the University of Padova, and the research was conducted in accordance with the Declaration of Helsinki. When the primary ethical approval could not be applied by a unit, the units obtained their own ethical approval for data collection.

The study protocol was approved by the Durham University Music Department Ethics Committee; the Research Ethics Board at the University of New Brunswick (File No. 2023-016); the local Ethics Committee of the Université libre de Bruxelles (1668/2024); the PPLS Research Ethics Committee of the University of Edinburgh (203-2324/2); the Brazilian National Ethical Committee at Plataforma Brasil (No. 70181123.8.0000.5411); the Ethics Review Committee Psychology and Neuroscience, Maastricht University (External Approval 76B8BA7B9078DF-27CF46CDC26FA9E5); the Ethics Council of the Max Planck Society (No. 2023_18); the Non-Medical Research Ethics Board of the University of Western Ontario; the Institutional Review Board of the University of Minnesota (0605S85872); the Institutional Review Board of Vanderbilt University (No. 162002); the Leiden University Psychology Research Ethics Committee (2023-11-24-F.L. Bouwer-V1-5105); the Ethics Committee of Research in Education and Psychology of the University of Montreal (Study Protocol No. 2023-4765); the Research Ethics Committee of the University of Sheffield (Alternative Ethics Approval No. 860); the Norwegian Agency for Shared Services in Education and Research SIKT (reference: 805630); the Central ethics Review Board non-WMO studies of the University Medical Center Groningen (Research Register No. 18645); and the Board for Ethical Questions in Science of the University of Innsbruck (28/2023).

Design

In the study, we compared a group of adult musicians with a group of adult nonmusicians. The experiment was carried out in person in the laboratories of the various units. All experimenters of the various units followed a shared written protocol to test the participant in the various tasks of the experiment. All the tasks were presented on a Google page that redirected to the various platforms used for the tasks (i.e., JsPsych, LimeSurvey). A clone of the page is accessible at <https://sites.google.com/view/the-memory-experiment-clone/home>.

Tasks

Each participant was asked to complete individually all of the following tasks. Participants could not skip any task, trial, or question of the following tools.

Memory tasks

Participants completed three short-term memory tasks: one testing the memory for musical material, one testing

verbal memory, and the third testing visuospatial memory. The verbal and the visuospatial tasks were recall tasks in which the participant had to reproduce the to-be-remembered stimuli. The music task was a recognition task in which the participant had to compare two melodies and judge whether they were identical or different. The reason for this difference (i.e., recall vs. recognition) is that music-recall tasks are difficult to conduct with nonmusicians: The participant should sing (or play with a music instrument or software) the to-be-remembered stimuli. Note also that the pace of the tasks was different: In verbal and visuospatial tasks, stimuli were presented at a pace of 1 per s. This same pace would produce extremely slow melodies in the music task, hence the pace for this task was one note every 500 ms. Participants received written feedback about their performance after each trial (i.e., correct/wrong) and the score they achieved at the end of each block (sum of correct responses over the total number of trials done). All the memory tasks were implemented in JSPSYCH (de Leeuw, 2015).

Music short-term memory. In each trial, the program presented a melody (called “standard”), followed after 2 s by a comparison melody of identical duration and number of notes. The notes of the melodies were piano tones with a duration of 500 ms created with the software Cubase 5.1 and Halion Sampler (Steinberg Media Technologies), presented in succession with no silent interval separating the notes. At the end of the comparison melody, the participant had to judge whether the comparison melody was identical (or not) to the standard melody. At the beginning of the task, the length of the melody was two notes. Each melody length was presented four times to the participant. Out of four trials for each melody length, for one, two, or three trials, the expected answer was “the two melodies are different,” whereas for the remaining trial(s), the expected answer was “the two melodies are identical.” This was implemented to avoid the participant adopting a strategy that balances the number of “identical” and “different” responses for each melody length. The task continued until the third incorrect response given by the participant. Melodies were fixed for all participants (i.e., melodies were not generated in real time). To prevent possible ceiling performances, the program could present melodies up to 40 notes long. Melodies were both preceded and followed by a signal (i.e., a visual cross) that informed the participant on the beginning/end of the standard comparison pair and a label (i.e., “standard melody,” “comparison melody”) that was presented simultaneously with the stimulus. Participants completed the task twice with different sets of melodies.

The notes of the melodies were taken from the C major diatonic scale (i.e., lower-do/C, re/D, mi/E, fa/F, sol/G, la/A, si/B, upper-do/C) and extended from do4/C4 ($f = 261.6\text{ Hz}$) to do5/C5 ($f = 523.3\text{ Hz}$). Melodies were pseudorandom melodies. The note succession did not follow any aesthetic tonal sense. Melodies longer than two notes had at least one rising and one falling pitch interval, and melodies up to the length of eight notes did not have repeated notes. In longer melodies, notes could be repeated (e.g., two mi, or E, notes were included in the melody), but the repeated notes were never adjacent. When the comparison melody was different from the standard melody, the difference was created by reversing the order of two consecutive notes of the standard melody (e.g., standard: do, fa, sol, re; comparison: do, sol, fa, re). Comparison melodies were constructed with the constraint that this reversal did not result in adjacent repeated notes, which would be very salient.

Verbal short-term memory. We implemented a forward digit span with custom sequences of digits presented visually to the participants. The digits were presented only visually because when stimuli are presented auditorily, musicians might exploit their better auditory processing, including a possible confound in the measure of verbal short-term memory (Caclin & Tillmann, 2018; Talamini et al., 2016, 2022). The participant was presented with a sequence of digits (1–9), and the task was to look at the sequence and when the sequence was over, type on the keyboard the digits of the sequence in the same order they were presented. Digits were presented one after the other at the center of the screen at a pace of 1 digit per second. Each digit remained on the screen for 500ms, followed by a 500ms blank. The digit span began with a sequence of two digits (sequence length = 2). Each sequence length was presented twice to the participant. When the participant reported correctly at least one out of two sequences, then they were presented with the successive sequence length that was one digit longer than the previous sequence. For sequences up to nine digits, digits were not repeated. For sequence lengths longer than nine digits (i.e., sequence length of 10 digits or more), digits could be repeated; however, the repeated digits were never concatenated. Sequences were also controlled for salient and “easy-to-chunk” patterns (e.g., “5-4-3-2,” “4-6-8”), which were eliminated and replaced. The task stopped when the participant gave two incorrect responses for a given sequence length. Sequences were fixed for all participants (i.e., sequences were not generated randomly in real time). Maximum sequence length was 18 digits (a number that the authors hypothesized was large enough to have no single ceiling performance). Sequences were preceded and followed by a signal (an auditory beep) that

informed the participant on the beginning/end of the sequence. Participants completed the task twice with different sets of sequences of digits.

Visuospatial short-term memory. We implemented a forward matrix span. The participant was presented with a four-by-four matrix grid. In each trial, a dot was switched on and off in the cells of the matrix grid highlighting—one after the other—several positions on the grid. The task of the participant was to look at the positions where the dot appeared and reproduce them in the same order. As soon as the last dot disappeared, the participant could click with the mouse on the matrix on all the positions occupied by the dot. The dots were presented at a pace of 1 dot per second. Dot onset-to-onset interval was 1 s. Each dot remained on the screen for 500ms, followed by a 500ms blank. In the experiment, the participant began the task with a sequence length of two positions. Each sequence length was presented twice to the participant. When the participant reproduced correctly at least one out of the two sequences presented, they moved to the next sequence length, and an additional spatial position was added to the sequence. The task ended when the participant made two errors for a given sequence length. Sequences were fixed for all participants (i.e., sequences were not generated randomly in real time). Up to a sequence length of 16, spatial positions were not repeated. In longer sequences, the spatial positions could be repeated, but repeated positions were never concatenated. Sequences were also controlled for salient and easy-to-chunk geometric patterns (e.g., the dot touching one after the other the four corners of the matrix grid), which were eliminated and replaced. Maximum sequence length was 32 (a number that the authors hypothesized was large enough to have no single ceiling performance). Sequences were preceded/followed by a signal (i.e., an auditory beep) that informed the participant about the beginning/end of the sequence. The visuospatial span was run twice by the participants.

Music aptitude, music sophistication, and music reward

We assessed music aptitude, music sophistication, and music reward with three tools: the Profile of Music Perception Skills - mini (Mini-PROMS; Zentner & Strauss, 2017), the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014), and the extended Barcelona Music Reward Questionnaire (eBMRQ; Cardona et al., 2022). The Mini-PROMS is a shorter version of PROMS (Law & Zentner, 2012), a performance test of music-perception skills. It includes four listening subtests (Melody, Tuning, Beat, and Speed) investigating different aspects of music perception and memory. In

each subtest, the participant listens twice to a standard stimulus, followed by a comparison stimulus. The participant has to judge whether the comparison stimulus is identical or different from the standard on a five-options scale (i.e., definitively same, probably same, I don't know, probably different, definitively different). The Melody subtest assesses the ability of recognizing whether two short melodies are identical. The Tuning subtest requires comparing chords. In the case of different trials, one of the middle notes of the comparison chord is different in frequency (i.e., it is mistuned). The Beat subtest requires comparing rhythmic patterns of clicks: The accent that defines the beat is produced by raising the intensity of a subset of the clicks. In the Speed subtest, the participant compares the speed (i.e., beats per minute) of both synthetic rhythmic structures and recorded samples of music. The Gold-MSI is a 38-item self-report questionnaire that aims to capture individual differences in musical sophistication. The Gold-MSI is a questionnaire developed to collect information about common and skilled behaviors related to music in the Western population, such as interest in music and music-related activities (Müllensiefen et al., 2014). The Gold-MSI includes five subscales of self-assessment on active engagement (e.g., resources spent on music), perceptual abilities (i.e., music-listening abilities), musical training (e.g., formal music training received), singing abilities, and emotional engagement with music (e.g., ability to talk about music-evoked emotions). The eBMRQ (Cardona et al., 2022) extends the former Barcelona Music Reward Questionnaire (Mas-Herrero et al., 2012) and is a 24-item questionnaire to quantify reward in music-related activities. The eBMRQ includes six subscales: Music Seeking (e.g., the tendency to engage in music-related activities), Emotion Evocation (e.g., music's capacity to induce relevant emotional responses), Mood Regulation (e.g., how music is employed to regulate mood), Sensory-Motor (e.g., the capacity of music to intuitively induce body movements), Social Reward (e.g., ability of music to promote and enhance social interaction), and Musical Absorption (e.g., willingness to be deeply drawn in by sensory stimuli, experiencing immersion without distraction). These three music assessment tools were implemented in LimeSurvey.

Measure of personality

The Big Five Inventory - 2 (BFI-2) questionnaire (Soto & John, 2017) was used in the study. The questionnaire includes 60 items and returns scores for the following personality traits: open-mindedness, conscientiousness, extraversion, agreeableness, and negative emotionality. The questionnaire was implemented in LimeSurvey.

Measures of intelligence

We used the Raven Advanced Progressive Matrices (Raven, 1965), which is a common test to measure fluid intelligence in a language-independent way. Participants were given 10 min to respond to as many items as possible. We concatenated the last 10 items of Set I and all the items of Set II (36 items) to have a single block of 46 items. Items of the two sets were concatenated to avoid ceiling effects. The first two items of Set I were used for two familiarization trials before the task. The test was implemented in LimeSurvey. We also used a measure of crystallized intelligence. Participants were asked to complete the vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008; maximum 30 items). In Portuguese-speaking countries, this version of the test was not available; therefore, we used the latest version available (WAIS-III; Wechsler, 1997; maximum 33 items). The WAIS vocabulary test was administered by the experimenter.

Measures of executive functions

We used a visual two-back task to measure the updating component of executive functions. This component seems the most strongly linked to music aptitude and sophistication, in particular when it is tapped with the n-back task (Okada & Slevc, 2018; Slevc et al., 2016). The participant saw a series of 22 single letters drawn from a set of eight capital letters ("C," "D," "G," "K," "P," "Q," "T," "V"). Each letter remained on the screen for 500 ms, followed by 1,500 ms blank screen (i.e., letter onset-to-onset interval: 2,000 ms). The participants' task was to press a button when the letter they were currently seeing matched the letter seen two letters before. The task was repeated five times, and it was preceded by one familiarization series. Each series included six targets. This task was implemented in JSPSYCH (de Leeuw, 2015).

Measures of SES

We controlled for the role of this factor with the Hollingshead four-factor index (Hollingshead, 1975). The four-factor index can be used to ask details about the family of origin of the participant, such as parents' education, occupational status, and relationship (e.g., married, separated). Furthermore, we asked about the education and the occupational status of the participant. These questions were implemented in LimeSurvey.

Further custom questions

We asked all participants further demographic questions: sex, age, type of nonmusical education (e.g., type of

bachelor's degree). The musicians were asked the number of years of music training in an official music school, conservatory, or private lessons; the number of hours of music practice per day at the moment of testing; which instrument they played; whether they sang; and whether the music activity in the last 12 months had been continuous with no interruptions. In the case the musicians played multiple instruments, they were asked the above questions for only the main instrument they played. All participants were further asked whether they had relative and absolute pitch and one question from the Ollen Musical Sophistication Index (Ollen, 2006; "Which title best describes you?"), which considered the individual's self-assessed level of musical identity. This question can be answered in six ways (1 = nonmusician, 2 = music-loving nonmusician, 3 = amateur musician, 4 = serious amateur musician, 5 = semiprofessional musician, 6 = professional musician). The custom questionnaire included also a question on expertise, that is, whether the participant self-rated to be expert (on a scale from 1 to 6, similar to the question of the Ollen Sophistication Index) in one or more of the following domains: art (including music but not played), sport, game, and others.

Translations of the various tools into the different languages

Table S1 (in the Supplemental Material available online) includes the various tools that were used in the study and (if available) the existence of the official translation of the tool. For the remaining tools, units were asked to provide a translation based on two independent translations followed by back-translation.

Procedure

Participants were recruited in different ways (e.g., word of mouth, advertisement). Inclusion criteria were assessed at the recruitment stage. The experiment took place in person in the laboratory in one session. At the beginning of the experiment, the participant was asked to read and sign the informed consent. Then, to ensure that the participant was in an unaltered psychophysiological state, a few questions were asked (i.e., about drug/alcohol intake, amount of sleep in the past 24 hr, and current sickness). If the participant reported to have used any drug and/or taken an excessive amount of alcohol in the 24 hr preceding the experiment or reported insufficient sleep (i.e., <6 hr) or sickness, they were not allowed to continue with the experiment. Subsequently, the participant completed the memory tasks: the music-memory task, the verbal-memory task, and the visuo-spatial-memory task. The order of the three tasks was

counterbalanced within each participant group and data-collection unit in a Latin square way. After the memory tasks, the participant took the Raven test and the WAIS vocabulary test. Next, the participant was asked to take the n-back task and the PROMS test and complete all the self-report questionnaires (custom questions, Gold-MSI, eBMRQ, BFI-2, and Holligshead). Participants were allowed to take short breaks in between any of the tasks of the experiment. The approximate duration of the whole experiment was 2 hr.

Power analysis and participants

Given the multilevel data structure of the study (i.e., a multilab study), the required sample size and number of units was calculated with a power analysis in a meta-analysis framework. We used the approach by Hedges and Pigott (2001) and Borenstein et al. (2009). We fixed the α level to .05 and power to 90%. The average effect size was fixed to Hedges's $g = 0.3$, which approximates the estimated effect size observed by Talamini et al. (2017) for visuospatial stimuli. Although smaller effect sizes could be targeted, differences in memory smaller than $g = 0.3$ would reveal a nonsubstantial memory advantage for musicians such as the individuals recruited in the study (see the criteria below). To account for the possible heterogeneity across data-collection units, we fixed τ (the index that reflects the true heterogeneity across research units) to .11. This value is half of the estimated heterogeneity by Talamini et al. and was selected given the identical experimental protocol shared by all units participating in the current study. This present power analysis requires a minimum of 21 participants per group (i.e., 21 musicians and 21 nonmusicians) collected by a minimum of 13 research units.

As inclusion criteria, participants had to be healthy young adults ranging from 18 to 30 years of age. They had to have normal hearing and normal or corrected-to-normal vision (both assessed via self-report). As far as the classification of musicians and nonmusicians is concerned, we adopted stricter criteria than those suggested by Zhang and Schubert (2019). Musicians had to have 10 years (or more) of music lessons (voice or instrument) in music schools or with a private teacher (SMI; Zhang & Schubert, 2019) and had to be musically active at the moment of the experiment with no interruption in the music activity in the year preceding the experiment. Nonmusicians should not have received more than 2 years of music lessons, except for compulsory school activities (a criterion that is stricter than that suggested by Zhang & Schubert, 2019, for the "no musical identity"), and had to be musically inactive for at least the 5 years before the day of the experiment.² Musicians and

nonmusicians were individually matched for gender, age, and years of education with a tolerance of ± 1 year.³ For example, a male 22-year-old musician with 16 years of education could be approximately matched with a male 21-year-old nonmusician with 17 years of education. Years of education included music training. If the music training was parallel to another type of school (e.g., in some countries, students can be enrolled at the same time in the conservatory of music and at a different high school), each overlapping year of education between music and nonmusic school counted as 1 year of education. Participants were compensated for their participation according to the standards of the recruiting unit (e.g., money, university credits).

At the end of data collection, 1,357 participants were recruited. Of these, 157 were excluded because the unit could not find a corresponding matched musician (or nonmusician) or the participant did not complete all the tasks. This resulted in a final sample of 1,200 participants, comprising 600 adult musicians and 600 adult nonmusicians. A total of 33 units collected the data: one in Australia, one in Austria, one in Belgium, one in Brazil, four in Canada, two in Germany, two in Spain, one in Finland, four in France, three in Italy, three in the Netherlands, one in Norway, one in Portugal, four in the United Kingdom, and four in the United States. For the number of participants collected by each unit, see Table 2. This number of units and participants enabled us to target an effect size of Hedges's $g = 0.3$ with over 95% power at a α level to .05. For smaller effect sizes (Hedges's $g = 0.2$), the statistical power we achieved is above 70%.

For a synthesis of the demographic variables of both groups, see Table 3.

Analyses and Results

From raw data to data for statistical analysis

The following values were calculated from the raw data, separately for each participant. For each memory task, we calculated the average of the sums of the correct responses of each block of trials.⁴ Subsequently, we calculated the PROMS score in two ways by including all subtests (i.e., sum of the raw scores of each of the subtests divided by 2; range = 0–36) or excluding the Melody subtest from the total score (range = 0–26). Hereafter, these two scores are referred to as “PROMS” and “PROMS-noMelody.” These two scores were calculated because of the similarity between the Melody subtest and the music-memory test, both requiring memorizing and comparing different melodies. Then, we calculated the general Gold-MSI score (range = 18–126),

Table 2. Units That Participated in the Data Collection and the Number of Musicians and Nonmusicians Included in the Analysis for Each Unit

Unit	<i>n</i> musicians	<i>n</i> nonmusicians
IT-Padova	22	22
UK-Edinburgh	22	22
AU-Sydney	21	21
BE-Bruxelles	21	21
BR-Sao Paulo	21	21
CA-Montreal	21	21
DE-Frankfurt	21	21
DE-Hannover	21	21
ES-Granada	21	21
FI-Helsinki	21	21
FR-Dijon	21	21
FR-Lyon	21	21
IT-Pavia	21	21
NL-Groningen	21	21
NO-Oslo	21	21
PT-Lisbon	21	21
US-Minnesota	21	21
US-Southern California	21	21
AT-Innsbruck	20	20
CA-New Brunswick	20	20
CA-Western Ontario	20	20
ES-Barcelona	20	20
UK-Durham	20	20
IT-Bari	19	19
FR-Caen	18	18
CA-McMaster	16	16
NL-Leiden	12	12
UK-Sheffield	12	12
US-Vanderbilt	11	11
US-Maryland	11	11
FR-Lille	8	8
NL-Maastricht	7	7
UK-Goldsmiths	6	6
Total	600	600

Note: AU = Australia; AT = Austria; BE = Belgium; BR = Brazil; CA = Canada; DE = Germany; ES = Spain; FI = Finland; FR = France; IT = Italy; NL = the Netherlands; NO = Norway; PT = Portugal; UK = United Kingdom; US = United States.

Raven score (sum of the number of the correct responses; range = 0–48), two-back score (the sum of hits minus the sum of false alarms; range = –75 to +30), BFI-2 open-mindedness score (range = 12–60), eBMRQ score (range = 24–120), and Hollingshead Index for the SES of the family (range = 8–66; for the calculation algorithm, see Hollingshead, 1975).⁵ Concerning the WAIS vocabulary score, because WAIS-III and WAIS-IV have different maximum scores (respectively, 66 and 57), the individual scores were transformed into proportion of the

Table 3. Demographic Data of the Participants of the Study

	Gender, <i>n</i>	Age (years)	Education (years)	Music training (years)
Musicians	F = 353	F = 21.7 (2.9)	F = 15.7 (2.1)	F = 12.5 (3.5)
	M = 245	M = 22.8 (3.5)	M = 16.1 (2.3)	M = 12.0 (3.9)
	NB = 2	NB = 19.5 (2.1)	NB = 14.5 (2.1)	NB = 13.0 (1.4)
Nonmusicians	F = 353	F = 21.7 (2.8)	F = 15.7 (2.0)	F = 0.6 (1.3)
	M = 245	M = 22.8 (3.4)	M = 16.1 (2.3)	M = 0.5 (0.9)
	NB = 2	NB = 20.5 (2.1)	NB = 15.5 (2.1)	NB = 1.5 (2.1)

Note: Shown are means with standard deviations in parentheses. Number of years is truncated after the first decimal. F = females; M = males; NB = nonbinary.

maximum score (range = 0–1). The remaining subscores of questionnaire and tests (e.g., the Melody, Tuning, Beat, and Speed subscales of PROMS; the conscientiousness, extraversion, agreeableness, and negative-emotionality scores of the BFI-2⁶) were analyzed exploratorily. All the variables collected are reported in the final data set (when the variable is analyzed here or in the supplemental materials) or provided as raw data. All data files are available on OSF.

Preregistered analyses and results

We compared the memory performance of musicians and nonmusicians with three separate random-effects meta-analyses (i.e., for musical, verbal, and visuospatial stimuli) using the *Metafor* R Package (Viechtbauer, 2010). This analysis takes into account the multilevel data structure of the current study. First, we calculated the standardized effect (i.e., Hedges's *g*) and the sampling variance (Borenstein et al., 2009; Hedges, 1981) for each research unit. We interpreted the Hedges's *g* as follows: very small (<0.2), small (0.2–0.5), medium (0.5–0.8), large (0.8–1.2), very large (1.2–2), and huge (>2; see Sawilowsky, 2009).

The average effect size for each meta-analysis was tested for significance using two-tailed Wald-type *z* tests. The α value was set to .05 for all the analyses. For each meta-analysis, we also report the *Q* test for the heterogeneity (Berkey et al., 1995) and the I^2 statistic that represents the percentage of total variation because of real heterogeneity (i.e., τ).

Musicians performed better than nonmusicians in music short-term memory, with a large effect size (Hedges's *g* = 1.08, *SE* = 0.07, 95% confidence interval [CI] = [0.94, 1.22], *z* = 15.11, *p* < .001); they also performed better in verbal short-term memory, with a very small effect size (Hedges's *g* = 0.16, *SE* = 0.07, 95% CI = [0.02, 0.30], *z* = 2.21, *p* = .027), and in visuospatial short-term memory, with a small effect size (Hedges's *g* = 0.28, *SE* = 0.06, 95% CI = [0.15, 0.41], *z* = 4.33, *p* < .001). For music and visuospatial short-term memory, the heterogeneity

was not significantly different from zero according to the *Q* statistics, and the I^2 suggests a small proportion of true heterogeneity (τ^2) out of the total variability. For the verbal short-term memory, the heterogeneity was significantly different from zero, and both τ^2 and I^2 were higher compared with the other tasks. Figure 2 depicts the three forest plots of the meta-analysis results. Table 4 represents the detailed summary of the meta-analysis models, and the memory scores are represented in the plot depicted in Figure 3.

Next, Pearson correlations were calculated for the following variables, separately for musicians and nonmusicians: music, verbal, and visuospatial short-term memory scores; PROMS score; PROMS-noMelody score; Gold-MSI score; Raven and WAIS vocabulary scores; n-back score; BFI-2 open-mindedness score; eBMRQ score; SES of the family of the participant; age; and years of education. Correlations are reported in Figure 4.

Subsequently, we calculated three separate multilevel regression models for each memory task including the research unit as random effect and a random slope for the group (musicians vs. nonmusicians) effect. As predictors, we included the main effects of Raven, WAIS vocabulary, n-back, BFI-2 open-mindedness, eBMRQ, SES, age, and years of education and their interaction with the group (musicians and nonmusicians). Using a model-comparison approach, we selected the best model predicting each memory outcome (see Tables 5, 6, and 7). PROMS and Gold-MSI were not included in the regression models because both are strongly related to musical expertise (possibly dependent on it). In fact, music-perception skills are highly correlated with music expertise (Zentner & Strauss, 2017), and the Gold-MSI includes items exploring the relationship with music, for example, music-training history, self-reported perceptual skills, and active engagement, leading to much higher scores in musicians (Müllensiefen et al., 2014). Numerical predictors have been centered on the grand mean, and the group factor has been coded using sum-to-zero

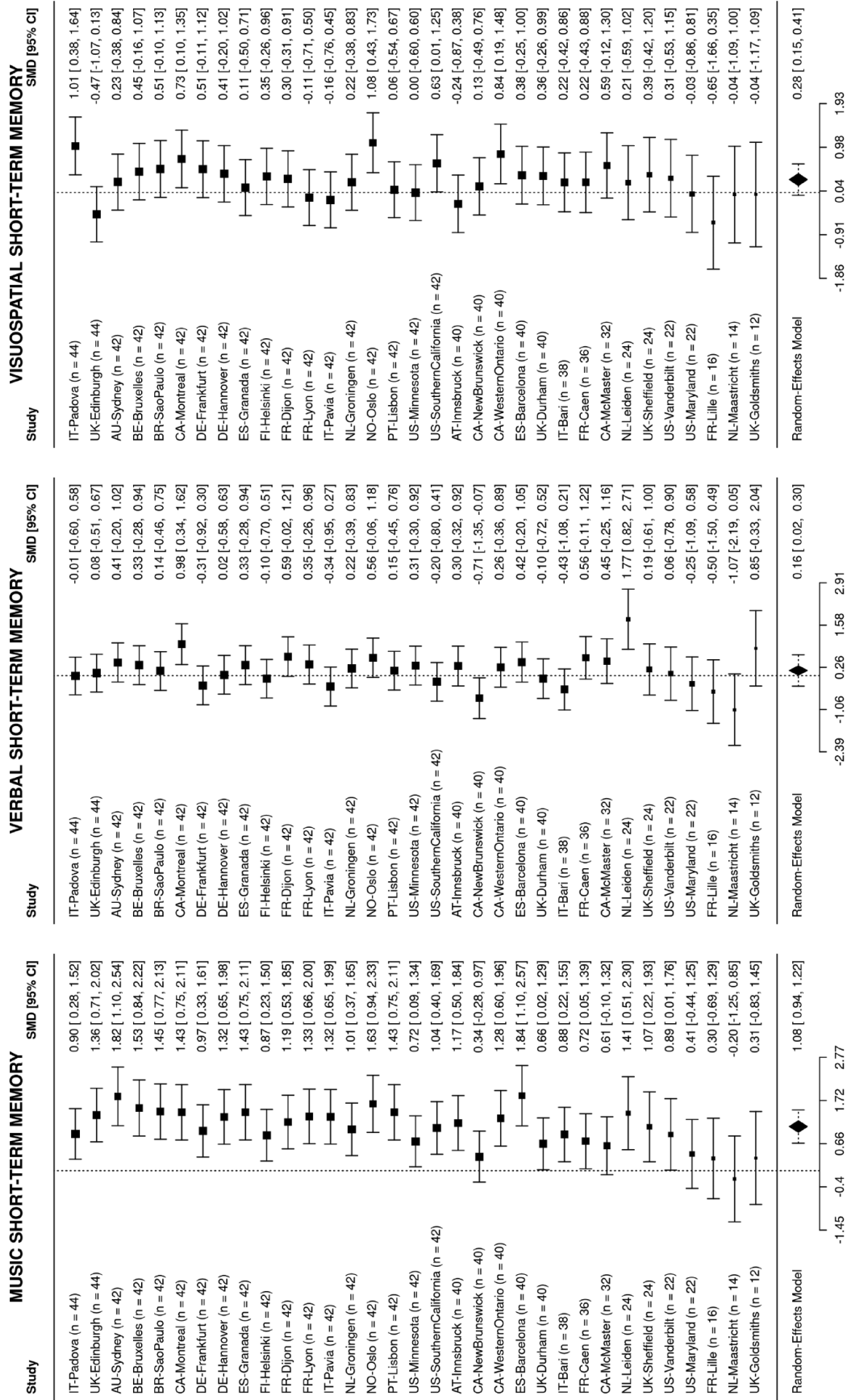


Fig. 2. Forest plots of the random-effect meta-analysis for each memory task. For each research unit, the plot reports the average effect with the 95% confidence interval. The diamond represents the meta-analytic effect and the 95% confidence interval. The error bar of the diamond represents the 95% prediction interval. For each unit, the forest plot reports the number of participants recruited by the unit. Within each unit, there was an equal number of musicians and nonmusicians.

Table 4. Meta-Analysis Models Results

	<i>g</i>	<i>SE</i>	<i>z</i>	<i>p</i>	95% CI
Music short-term memory	1.08	0.07	15.11	<.001	[0.94, 1.22]
$Q_{32} = 44.49, p = .07, \tau^2 = 0.04$ ($\tau = 0.19$), $I^2 = 22.87\%$					
Verbal short-term memory	0.16	0.07	2.21	.027	[0.02, 0.30]
$Q_{32} = 53.04, p = .011, \tau^2 = 0.06$ ($\tau = 0.24$), $I^2 = 33.84\%$					
Visuospatial short-term memory	0.28	0.06	4.32	<.001	[0.15, 0.41]
$Q_{32} = 38.55, p = .197, \tau^2 = 0.03$ ($\tau = 0.16$), $I^2 = 18.56\%$					

Note: For each memory task, we report the effect size (*g*), the standard error, the *z* statistics, the two-sided *p* value, and the 95% CI. In terms of heterogeneity, we report the *Q* statistics with the associated *p* value. The *Q* test for heterogeneity tests the null hypothesis of no heterogeneity ($\tau^2 = 0$). τ^2 is the estimated variance of the true effects, and I^2 is the relative measure of heterogeneity, interpreted as the percentage of the total variance because of true heterogeneity (Higgins & Thompson, 2002). CI = confidence interval.

contrasts for better interpreting main effects and interactions.

For all calculated models, we selected the best model with a backward model selection starting from the full model containing all the main effects, all the two-way interactions between the group factor and the other predictors, and the random slope for the group factor

and random intercept for units. The procedure was implemented in the “stepO” function from the *ImerTest* R package (Kuznetsova et al., 2017). The “stepO” function implements a backward elimination based on the likelihood ratio test between nested models. For the random-effects stepwise selection, the final music-memory model contained the random intercept for the

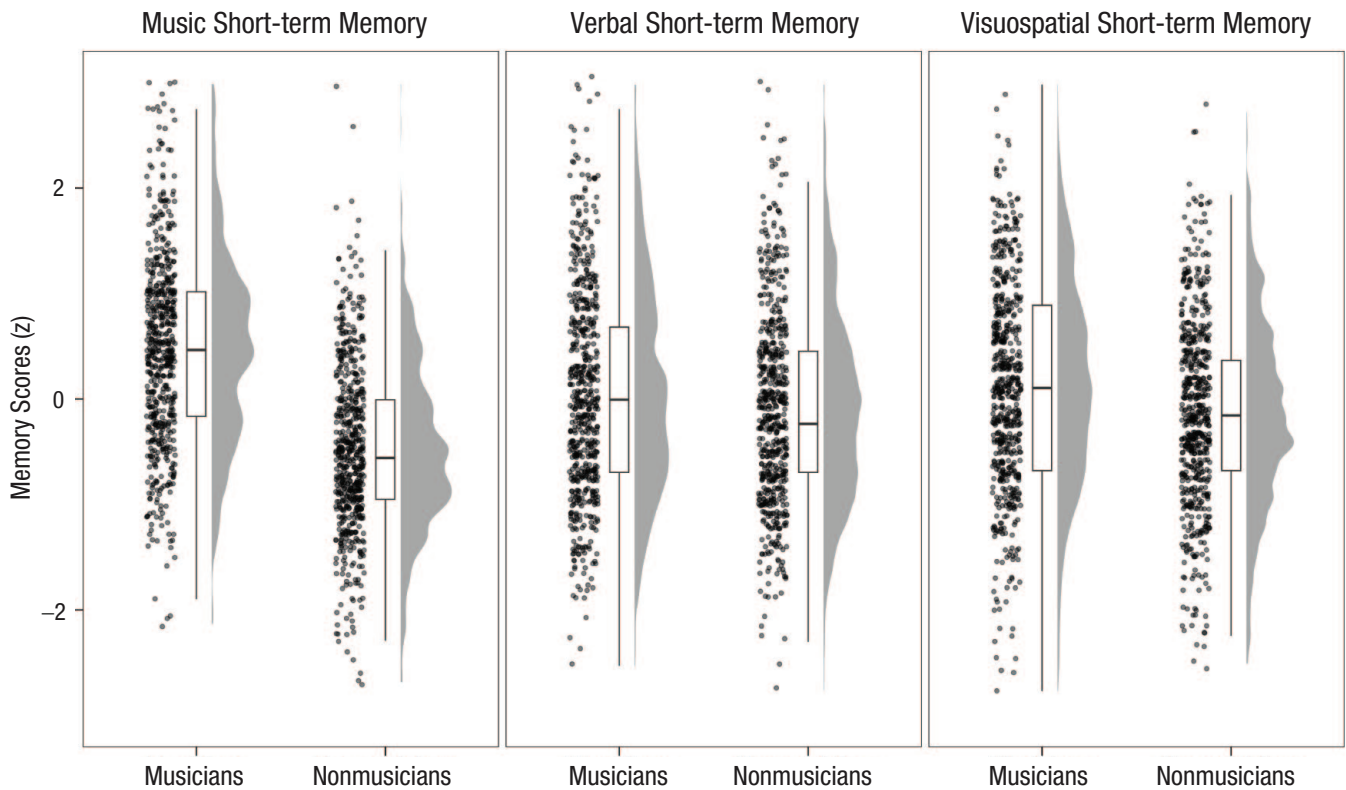


Fig. 3. Short-term memory for music, verbal, and visuospatial stimuli of musicians and nonmusicians. Scores were converted into *z* scores for easier comparisons across stimulus types. In each graph, from left to right, dots represent individual observations; box plots represent the median (i.e., the horizontal lines inside the box), the 25th and 75th percentiles (i.e., the edges of the box), and the interquartile range Q1-Q3 augmented by 50% (i.e., the whiskers); and the gray area represents density plots. For graphical reasons, the plots do not represent 16 outlier performances ($|z|$ score >3; music: $N = 6$; verbal: $N = 5$; visuospatial: $N = 5$).

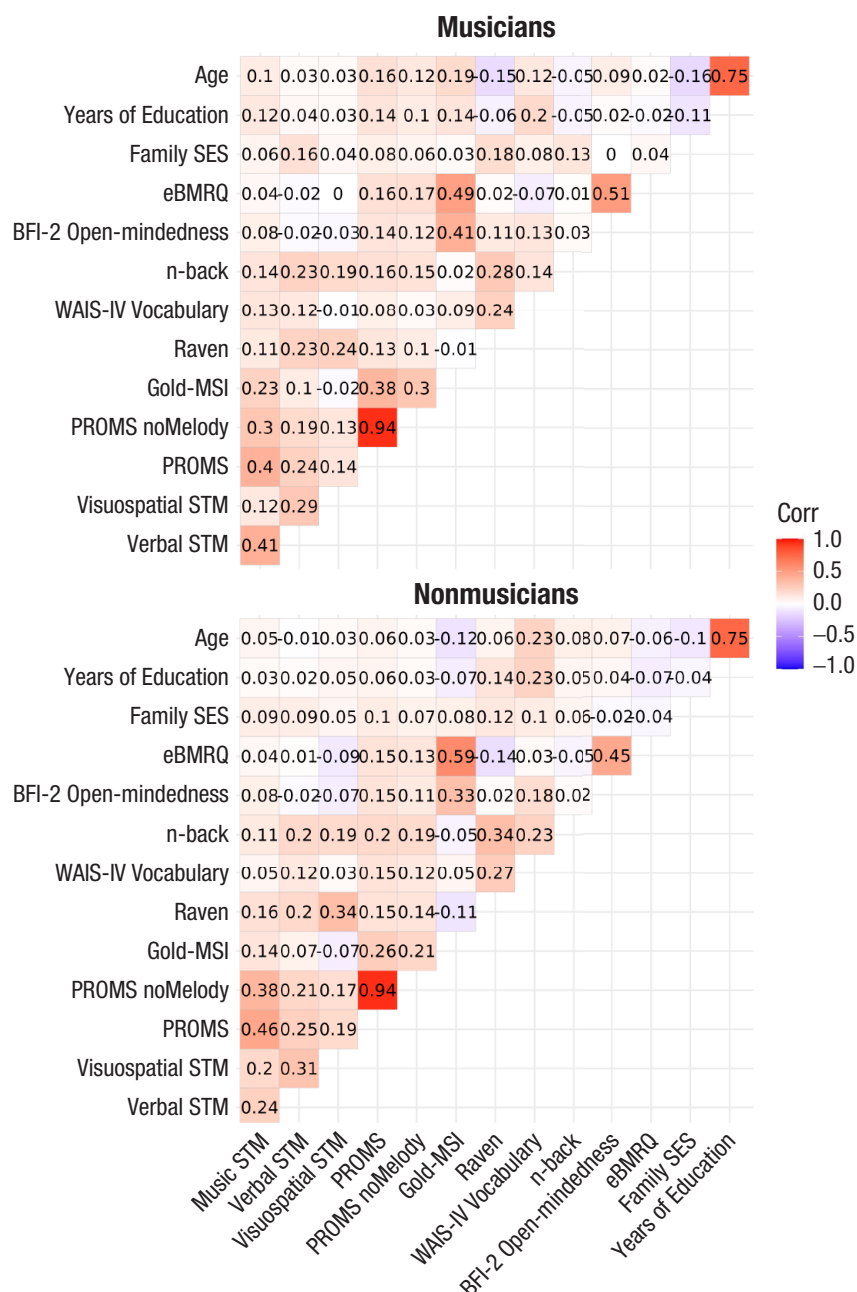


Fig. 4. Pearson correlations among the main measures collected in the study. STM = short-term memory.

unit and a random slope for the group effect. Because the correlation between slopes and intercepts was not properly estimated (i.e., singular fit), we fixed the correlation to zero (intercepts and slopes are independent). Tables 5 to 7 report the results of the best models for music short-term memory, verbal short-term memory, and visuospatial short-term memory, respectively.

The best model for music short-term memory included the following predictors: group, fluid intelligence (Raven), crystallized intelligence (WAIS vocabulary),

executive functions (n-back), BFI-2 open-mindedness, SES of the family, years of education, and the interaction between group and crystallized intelligence. The model explains 26.2% of the variance through fixed effects and 29.3% when both fixed and random effects are considered. The best verbal-short-term-memory model included the following predictors: fluid intelligence (Raven), executive functions (n-back), the SES of the family, and age. The model explains 9.3% of the variance through fixed effects and 14.6% when both fixed and random effects

Table 5. Summary of the Best Regression Model for Music Short-Term Memory, Selected Using the Backward Stepwise Procedure

Parameter	β	SE	t	df	p
(Intercept)	24.48	0.21	113.93	27.22	<.001
Group (musicians, nonmusicians)	5.22	0.44	11.85	35.60	<.001
Raven Advanced Progressive Matrices	0.08	0.03	2.82	1,187.41	.005
WAIS vocabulary	2.20	1.28	1.72	296.61	.087
n-back	0.05	0.01	2.83	1,181.13	.005
BFI-2 open-mindedness	0.05	0.02	2.01	1,175.98	.045
SES of the family	0.02	0.01	2.28	1,057.27	.023
Years of education	0.18	0.08	2.33	858.02	.020
Group $\times$ WAIS Vocabulary	6.36	2.34	2.71	330.87	.007
σ_{Unit}	0.82				
σ_{Group}	1.53				
σ_{ϵ}	5.35				
Marginal $R^2 = .26$					
Conditional $R^2 = .29$					

Note: β is the parameter value, t is the Student's t statistics, and df is the degrees of freedom calculated using the Satterthwaite method (Kuznetsova et al., 2017). σ is the random-intercepts standard deviation, σ_{Group} (if present) is the standard deviation of the random slopes for the group effect, and σ_{ϵ} is the residual standard deviation. We also report the marginal (variance explained by the fixed effects) and conditional (variance explained by the fixed plus random effects) R^2 values (Nakagawa et al., 2017). WAIS = Wechsler Adult Intelligence Scale; BFI-2 = Big Five Inventory - 2; SES = socioeconomic status.

are considered. As far as the visuospatial short-term memory is concerned, the best model included the following predictors: group, fluid intelligence (Raven), executive functions (n-back), and BFI-2 open-mindedness. The model explains 11.7% of the variance through fixed effects and 14.8% when both fixed and random effects are considered. For the graphical representations of the effects of these predictors on the three short-term memory performances, see Figures S1 through S3 in the Supplemental Material.

Exploratory analyses and results

We compared musicians and nonmusicians along the various measures we collected. For each measure, we calculated a random-effect meta-analysis using the *Meta-for R* Package (Viechtbauer, 2010). Figures 5 and 6 report the differences between musicians and nonmusicians for the musical measures (i.e., PROMS and its subscales, Gold-MSI and its subscales, eBMRQ and its subscales) and for the remaining measures (i.e., Raven and WAIS

Table 6. Summary of the Best Regression Model for Verbal Short-Term Memory, Selected Using the Backward Stepwise Procedure

Parameter	β	SE	t	df	p
(Intercept)	10.02	0.11	93.67	30.35	<.001
Raven Advanced Progressive Matrices	0.07	0.01	5.97	1,190.27	<.001
n-back	0.04	0.01	5.57	1,194.30	<.001
SES of the family	0.01	0.004	2.90	1,187.47	.004
Age	0.06	0.02	2.97	1,041.16	.003
σ_{Unit}	0.51				
σ_{ϵ}	2.03				
Marginal $R^2 = .09$					
Conditional $R^2 = .15$					

Note: β is the parameter value, t is the Student's t statistics, and df is the degrees of freedom calculated using the Satterthwaite method (Kuznetsova et al., 2017). σ is the random-intercepts standard deviation, and σ_{ϵ} is the residual standard deviation. We also report the marginal (variance explained by the fixed effects) and conditional (variance explained by the fixed plus random effects) R^2 values (Nakagawa et al., 2017). SES = socioeconomic status.

Table 7. Summary of the Best Regression Model for Visuospatial Short-Term Memory, Selected Using the Backward Stepwise Procedure

Parameter	β	SE	t	df	p
(Intercept)	8.80	0.08	112.14	32.09	<.001
Group (musicians, nonmusicians)	0.46	0.11	4.05	1,173.84	<.001
Raven Advanced Progressive Matrices	0.09	0.01	8.97	1,194.11	<.001
n-back	0.02	0.005	3.93	1,193.00	<.001
BFI-2 open-mindedness	-0.02	0.01	-2.42	1,194.94	.016
σ_{Unit}	0.34				
σ_{ϵ}	1.77				
Marginal $R^2 = .12$					
Conditional $R^2 = .15$					

Note: β is the parameter value, t is the Student's t statistics, and df is the degrees of freedom calculated using the Satterthwaite method (Kuznetsova et al., 2017). σ is the random-intercepts standard deviation, and σ_{ϵ} is the residual standard deviation. We also report the marginal (variance explained by the fixed effects) and conditional (variance explained by the fixed plus random effects) R^2 values (Nakagawa et al., 2017). BFI-2 = Big Five Inventory - 2.

vocabulary, n-back, the five subscales of the BFI-2, and the SES of the family). We highlight here the major differences between musicians and nonmusicians emerging from these meta-analyses. For the full statistics for these meta-analyses, see Tables S2 through S7 in the Supplemental Material. Musicians were “more musical” (as emerged from the results of the Gold-MSI and the PROMS) than nonmusicians. The effect sizes (g) of the difference between musicians and nonmusicians in the music-related measures (including questionnaires' subscales) ranged from 0.38 to 5.83, and the majority were large or more than large (see the Supplemental Material). Musicians performed better than nonmusicians in the PROMS test, with a very large effect size (Hedges's $g = 1.69$, 95% CI = [1.50, 1.88]; see Fig. 5, left); were

more musically sophisticated than nonmusicians, with a huge effect size (Gold-MSI; Hedges's $g = 3.28$, 95% CI = [3.02, 3.55]; see Fig. 5, right); and were more sensitive to music reward, with a large effect size (eBMRQ; Hedges's $g = 1.07$, 95% CI = [0.93, 1.20]; see Fig. 5, center). These differences could be observed also in almost all subscales of the three tools.

As far as the cognitive measures are concerned, musicians performed better than nonmusicians in the Raven test (Hedges's $g = 0.26$, 95% CI = [0.11, 0.40]), the WAIS-V vocabulary (Hedges's $g = 0.40$, 95% CI = [0.24, 0.56]), and the n-back (Hedges's $g = 0.33$, 95% CI = [0.20, 0.47]), and all three effect sizes were small (see Fig. 6, left). Various differences could be observed in the measure of the five dimensions of personality (see Fig. 6, right).

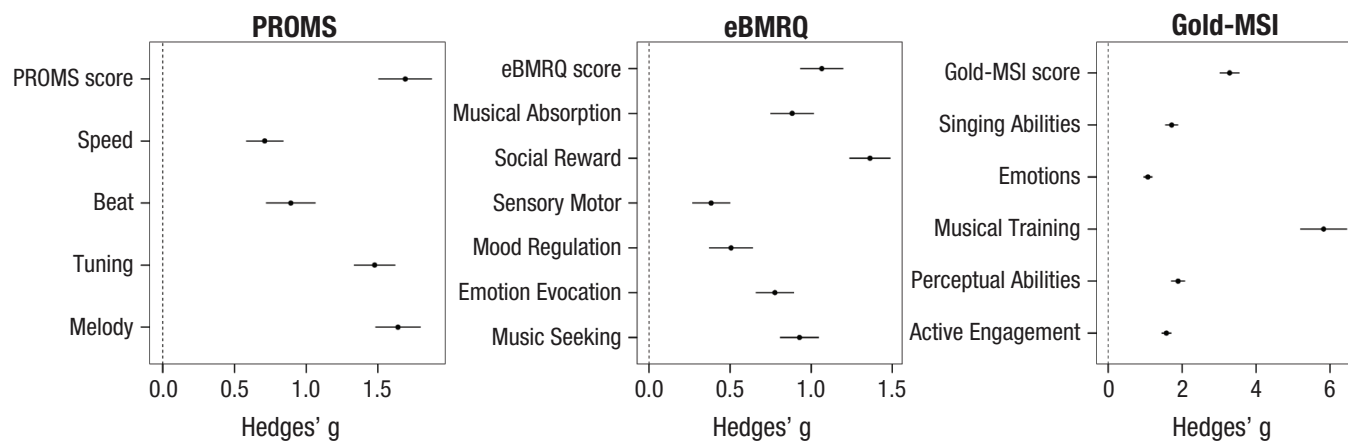


Fig. 5. (Left to right) Differences between musicians and nonmusicians in music-perception skills (PROMS), music reward (eBMRQ), and music sophistication (Gold-MSI). For each test, the total score (on the top of each plot) and the subscales scores are represented. The dots represent the mean effect sizes (Hedges's g), and the bars represent the 95% confidence interval. PROMS = Profile of Music Perception Skills; eBMRQ = extended Barcelona Music Reward Questionnaire; Gold-MSI = Goldsmiths Musical Sophistication Index.

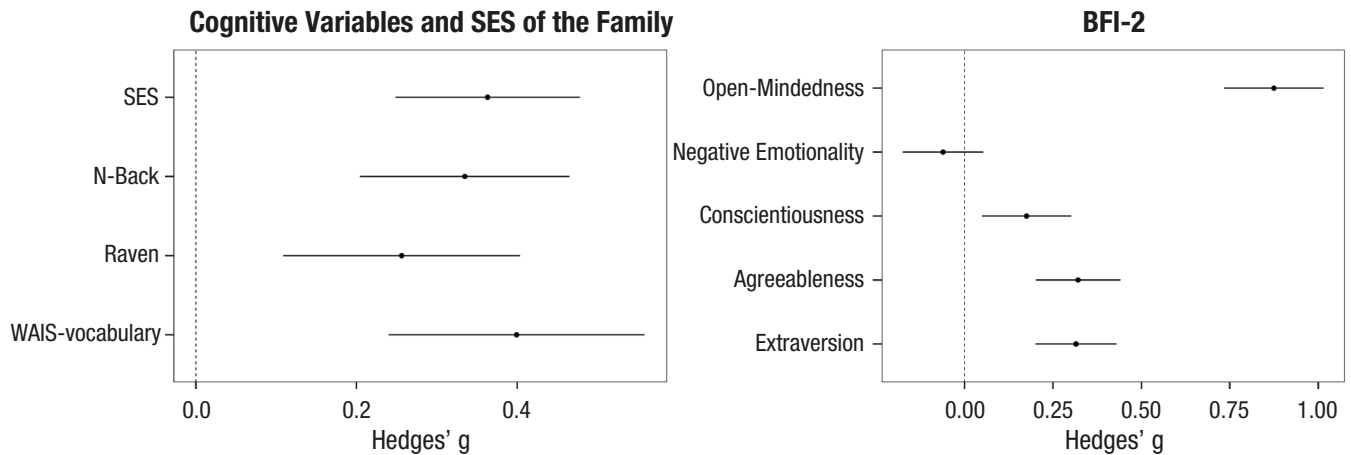


Fig. 6. (Left) Differences between musicians and nonmusicians in the SES of the family, executive functions (n-back), fluid intelligence (Raven Advanced Progressive Matrices), and crystallized intelligence (WAIS vocabulary). (Right) Differences between musicians and nonmusicians in personality (the five dimensions of the BFI-2). In each graph, the dots represent the mean effect sizes (Hedges's g), and the bars represent the 95% confidence interval. SES = socioeconomic status; WAIS = Wechsler Adult Intelligence Scale; BFI-2 = Big Five Inventory - 2.

Musicians reported higher scores than nonmusicians in open-mindedness, with a large effect size (Hedges's $g = 0.87$, 95% CI = [0.73, 1.02]). They also reported higher scores in both extraversion and agreeableness, with a small effect size: Hedges's $g = 0.32$, 95% CI = [0.20, 0.43] for extraversion, and Hedges's $g = 0.32$, 95% CI = [0.20, 0.44] for agreeableness. In conscientiousness, too, musicians reported higher scores, with a very small effect size (Hedges's $g = 0.17$, 95% CI = [0.05, 0.30]), and they reported lower scores in negative emotionality, with a very small effect size (Hedges's $g = -0.06$, 95% CI = [-0.17, 0.05]). Finally, musicians reported coming from families of higher SES than nonmusicians, with a small effect size (Hedges's $g = 0.36$, 95% CI = [0.25, 0.48]).

In addition, multilevel regression models were calculated separately for each memory task and separately for musicians and nonmusicians with the research unit included as a random effect. These separate models provide a clearer view of the potential effects of all predictors by avoiding issues arising from the association between the group factor and musical variables (i.e., PROMS and Gold-MSI). Specifically, including these predictors in separate models for musicians and nonmusicians allows for a better understanding of whether and how a person's musicality explains performance in musical, verbal, and visuospatial short-term-memory tasks. The predictors included the main effects of Raven, WAIS vocabulary, n-back, BFI-2 open-mindedness, Gold-MSI, eBMRQ, SES, age, and years of education. For music short-term memory, the model was calculated in two ways: using the overall PROMS score (including all subscales) and using the PROMS score excluding the Melody subscale. Using a model-selection approach, as described above, we selected the best model for predicting each

memory outcome. Below, we report only betas, p values, and marginal and conditional R^2 s. For the full results, see the Supplemental Material.

For musicians, when the full PROMS was included in the analysis, the multilevel regression analysis for music short-term memory identified three significant predictors. WAIS vocabulary ($\beta = 6.01$, $p = .001$), PROMS ($\beta = 0.57$, $p < .001$), and Gold-MSI ($\beta = 0.05$, $p = .029$) were all positively predicting music short-term-memory performance. The model explained 17.6% of the variance through fixed effects and 21.1% when both fixed and random effects are considered. When the model included the PROMS without the Melody subscale, five predictors emerged: WAIS vocabulary ($\beta = 6.05$, $p = .002$), n-back ($\beta = 0.06$, $p = .025$), eBMRQ ($\beta = -0.05$, $p = .04$), PROMS-noMelody ($\beta = 0.47$, $p < .001$), and Gold-MSI ($\beta = 0.10$, $p < .001$). The model explained 14.1% of the variance through fixed effects and 18.8% when both fixed and random effects are considered. The multilevel regression analysis for verbal short-term memory identified five significant predictors: Raven ($\beta = 0.06$, $p < .001$), n-back ($\beta = 0.04$, $p < .001$), SES ($\beta = 0.02$, $p = .017$), age ($\beta = 0.07$, $p = .014$), and PROMS ($\beta = 0.12$, $p < .001$) were all positive predictors of verbal short-term-memory performance. The model explained 13.5% of the variance through fixed effects and 20.5% when both fixed and random effects are considered. The multilevel regression analysis for visuospatial short-term memory identified four significant predictors: Raven ($\beta = 0.07$, $p < .001$), n-back ($\beta = 0.03$, $p = .001$), and PROMS ($\beta = 0.07$, $p = .003$) positively predicted visuospatial short-term-memory performance. In contrast, Gold-MSI ($\beta = -0.016$, $p = .035$) was negatively associated. The model explained 9.1% of the variance

through fixed effects and 13.2% when both fixed and random effects are considered.

For nonmusicians, the multilevel regression analysis for music short-term memory using the overall PROMS score identified two significant predictors: Raven ($\beta = 0.09, p = .009$) and PROMS ($\beta = 0.53, p < .001$) both positively predicted music short-term-memory performance. The model explained 21.9% of the variance through fixed effects and 22% when both fixed and random effects are considered. The multilevel regression analysis for music short-term memory using the PROMS score excluding the Melody subscale identified three significant predictors: Raven ($\beta = 0.11, p = .001$) and PROMS-noMelody ($\beta = 0.52, p < .001$) were positive predictors, as was Gold-MSI ($\beta = 0.03, p = .031$). The model explained 16.2% of the variance through fixed effects and 16.5% when both fixed and random effects are considered. The multilevel regression analysis for verbal short-term memory identified three significant predictors: Raven ($\beta = 0.05, p < .001$), n-back ($\beta = 0.02, p = .006$), and PROMS ($\beta = 0.09, p < .001$) were all positive predictors of verbal short-term-memory performance. The model explained 10.1% of the variance through fixed effects and 15% when both fixed and random effects are considered. The multilevel regression analysis for visuospatial short-term memory identified three significant predictors: Raven ($\beta = 0.10, p < .001$) and PROMS ($\beta = 0.60, p < .001$) both positively predicted visuospatial short-term-memory performance, and BFI-2 open-mindedness was a negative predictor ($\beta = -0.02, p = .016$). The model explained 14.8% of the variance through fixed effects and 19.4% when both fixed and random effects are considered. For the summaries of these models, see the Supplemental Material.

Discussion

In the present multilab study, we collected evidence on whether a domain-specific training and expertise may be associated with superior domain-general abilities such as memory, a classic question of psychology (e.g., Thurstone, 1934). In particular, we investigated whether being an experienced musician may be in a relationship with a better short-term-memory performance for domain-specific stimuli (e.g., music) but also for stimuli that are far or not explicitly connected with the domain of expertise, such as verbal stimuli or visuospatial stimuli. To respond to this question, in the present project, we tried to set a standard for studies in this field. In many cases, past researchers have compared musicians and nonmusicians in studies that recruited too few participants to reliably capture the phenomena they were trying to investigate. Moreover, there has been no consensus on the selection of the possible mediators of the memory advantage or the definition of “musician” and

“nonmusician.” Here, in particular, we implemented a comprehensive set of measures, and above all, we pre-registered all actions taken before the collection of the data. Thirty-three different units from 15 different countries contributed to the data collection, for a total of 1,200 participants. Measures of musical, verbal, and visuospatial short-term memory were administered together with some questionnaires and tests assessing both possible confounding variables (i.e., measures of crystallized and fluid intelligence, executive functions, SES of the participant’s family, and personality) and music-related variables (i.e., music perception skills, music sophistication, and music reward).

Preregistered analyses: meta-analyses

The study revealed three main results. The first is that the short-term-memory performance of musicians was better than that of nonmusicians for musical stimuli and that the size of this difference was large (Hedges’s $g = 1.08$). The second is that the short-term-memory performance of musicians was better than that of nonmusicians for verbal stimuli and that the size of this difference was very small (Hedges’s $g = 0.16$). The third is that the short-term-memory performance of musicians was better than that of nonmusicians for visuospatial stimuli and that the size of this difference was small (Hedges’s $g = 0.28$). The large short-term-memory advantage of musicians over nonmusicians with musical stimuli is consistent with what was observed in the meta-analysis by Talamini et al. (2017). It is plausible to hypothesize that the musicians’ advantage in music short-term memory is the result of a near-transfer effect of music training (for a recent debate, see Neves et al., 2022; Román-Caballero & Lupiáñez, 2022) given that prior studies have shown that music training enhances auditory processing (e.g., Kraus & Chandrasekaran, 2010). In contrast, the advantage of musicians over nonmusicians in verbal short-term memory was very small, compared with the medium effect size reported in the meta-analysis by Talamini et al. Note that the studies included in the meta-analysis used relatively heterogeneous procedures. Most verbal short-term-memory tasks were presented auditorily, which is the standard modality for this type of task, except for two studies that used visual stimuli (i.e., Okhrei et al., 2016; Talamini et al., 2016). Concerning visuospatial short-term memory, results align with those observed in the meta-analysis, showing a small effect size. However, this result is substantially more robust and reliable in the present study because it is based on a much larger data set and a shared protocol. The meta-analysis included only six studies, and among these, only three involved tasks requiring participants to memorize sequences of spatial positions (Okhrei et al., 2016; Rodrigues et al., 2014; Suárez et al., 2016).

Preregistered analyses: multilevel models

Subsequent analyses revealed that the model that best predicted music short-term memory included the Raven test (i.e., a measure of fluid intelligence), WAIS vocabulary (i.e., a measure of crystallized intelligence), n-back (i.e., a measure of the “updating” component of executive functions), personality trait open-mindedness, SES of the family, years of education, and interaction between group and the vocabulary score: In musicians, music short-term memory was positively related to crystallized intelligence, whereas for nonmusicians, it was negatively related. The model thus suggests that music training explains not only the variance in the music short-term-memory task, as expected, but also several different individual characteristics. For instance, general cognitive abilities (i.e., intelligence, executive functions) explained part of the variance in the music short-term-memory tasks, and this is in line with prior literature showing a connection between short-term memory and general cognitive abilities (e.g., Colom et al., 2005; Miller & Vernon, 1992). Open-mindedness was also found in the past to be linked with higher cognitive abilities and musicianship status (e.g., Corrigall et al., 2013; DeYoung et al., 2013; Harris, 2004; Vincenzi et al., 2024). A high SES can foster access to stimulating environments that are important for cognitive development during critical phases, and in contrast, a low SES can limit access to the same environments (e.g., Osler et al., 2013; Schellenberg, 2020). This could explain its significant role in the best-fitting model.

Different results were observed for verbal and visuospatial short-term memory. For verbal memory, the best model included the Raven test, n-back, SES of the family, and age of the participant. These results support again that higher cognitive abilities may explain a better verbal short-term-memory performance (e.g., Colom et al., 2005). Counterintuitively, age predicted positively the verbal memory performance. However, we note that in our sample, age correlated strongly with the years of education, although this latter factor was not part of the best model. The music group factor was not part of the best model, and this also reflects the very small difference in performance between the two groups observed with the meta-analysis.

The very small difference between groups in verbal short-term memory and the lack of contribution of the group factor in the predictive model contrast with previous literature (e.g., M. Hansen et al., 2013; Talamini et al., 2017). One possible explanation is that unlike most studies, the digit-span task here was presented visually. The previously observed advantages of musicians in verbal short-term memory may be a by-product of superior auditory processing (e.g., Talamini et al., 2016), therefore not necessarily reflecting a general short-term-memory verbal advantage. Understanding the

possibility of a (auditory) verbal advantage is, however, challenging: Although in the past the modality of presentation was found to influence verbal short-term-memory performance (e.g., for serial position effects, suggesting that auditory and visual stimuli are represented differently; Macken et al., 2016; Talamini et al., 2022), classic models of working memory (e.g., Baddeley 1992) do not distinguish stimuli based on the sensory channel but, rather, on the type of the stimulus (i.e., verbal or visuospatial). More recent models seem to give importance to the sensory modality and/or discuss possible different representations of auditory versus visual stimuli but are still too recent to change the classic paradigms (e.g., Christophel et al., 2017; Nozari & Martin, 2024). In any case, the comparison between the present results and the results of the small-sized studies of the literature should be made cautiously: Results provided by small-sized studies may be very variable, making the comparison between the present multilab study and those studies extremely inconclusive because of the uncertainty associated in the estimates provided by these latter (see below for an extended explanation).

For visuospatial short-term memory, the best model included group, Raven test, n-back test, and open-mindedness. Here, for all numerical predictors except for open-mindedness, the contribution of the predictor was positive: The higher the value of the predictor, the better the short-term-memory performance was. This result stresses once more the role of higher cognitive functions in short-term-memory performance. Open-mindedness contributed negatively to visuospatial performance. This result is unexpected given that previous literature has observed that open-mindedness is positively associated with cognitive performance (e.g., Harris, 2004) and could be a factor explaining the advantage of the musicians over the nonmusicians in cognitive tasks (Corrigall et al., 2013). Further research is needed to understand the role of open-mindedness in cognitive performance.

Overall, musicians performed better than nonmusicians in all short-term-memory tasks. Only two predictors emerged in all models explaining the difference in performance between groups: a measure of fluid intelligence and a measure of executive functions. Musical expertise contributes to explain differences in music and visuospatial short-term memory but not in verbal short-term memory.

Exploratory analyses: meta-analyses on control variables

In further exploratory analyses, we investigated whether musicians and nonmusicians differed along the various measures collected in the study. Although the two groups were recruited with the aim of being directly comparable

(musicians and nonmusicians were matched for gender, age, and education), several differences were observed between musicians and nonmusicians. Musicians were much more musical than nonmusicians in both performance tests and self-assessment tools. Small differences were also observed in higher cognitive faculties, such as fluid and crystallized intelligence, and executive functions (e.g., Arndt et al., 2023; Clayton et al., 2016; Criscuolo et al., 2019). Altogether with the advantages in music, verbal, and visuospatial short-term-memory tasks, these results may remind one of a classic observation in psychology that dates back to Spearman's "positive manifold" (Spearman, 1904): Individuals performing well in one cognitive task tend to perform well also in other cognitive tasks. In other words, it is possible that the small and constant musicians' advantage observed throughout the various cognitive measures reflects a general advantage in g-factor (e.g., Floyd et al., 2021), which, in turn, influences performances in the different cognitive tasks. Here, large differences in cognitive performance between musicians and nonmusicians emerged only when the tasks tapped into the musicians' expertise (e.g., music memory), whereas all other cognitive tasks had small and comparable effect sizes.

Many differences could be observed in personality with musicians that were much more "open to experiences" but also slightly more "extroverted," "agreeable," and "conscientious." The difference in open-mindedness aligns with previous studies that found a link between being a musician and this personality trait (Corrigall et al., 2013; Gjermunds et al., 2020). The effect size of this difference was large, suggesting that this trait is indeed an important dimension that needs to be considered when studying the cognitive differences in musicians and nonmusicians and more generally, the correlates of music training. Finally, musicians came from families of higher SES than nonmusicians. This finding aligns with previous research, such as Müllensiefen et al. (2014), who observed that the SES of an individual's family correlates with the individual's musical sophistication.

The multilevel regression models calculated separately for musicians and nonmusicians revealed both commonalities and differences in the predictors of memory performance across tasks. Fluid intelligence emerged as a positive predictor for the majority of memory tasks in musicians and for all memory tasks in nonmusicians. The PROMS was a positive predictor for all tasks in both groups. For musicians, we observed the additional role of crystallized intelligence and general musical sophistication in predicting musical memory performance. For nonmusicians, we also observed a negative association between open-mindedness and visuospatial memory performance, as in the main model on the overall sample. A possible reason explaining why PROMS emerged

as a predictor in all models, in both musicians and nonmusicians, is that PROMS implements a recognition paradigm, thus leveraging short-term-memory skills. It is also possible that musical abilities are part of the positive manifold mentioned above. Previous studies have observed, indeed, a correlation between music-perception skills and other cognitive abilities and general intelligence (Swaminathan & Schellenberg, 2018; for a review, see Schellenberg & Weiss, 2013). When comparing musicians and nonmusicians, though, it is challenging to understand to what extent music-perception skills are improved by (and thus dependent on) training or are genetically influenced skills that influence who takes music lessons. Future studies should aim at collecting participants that are equally distributed on a continuum of music-perception skills (note that in the present study, we had instead only the two "extremes" of this continuum) to further investigate the relationship between these abilities, other cognitive abilities, and music training.

Although the variance explained by these models was moderate, the separate models provide a more detailed view of how individual cognitive and personal characteristics contribute to memory performance within each group, avoiding the confounding effects of musical variables inherently linked to the group factor.

General comments on the results

We would like to draw some general comments on the results. The first is that musicians do have a short-term-memory advantage compared with nonmusicians but that existing literature seems to overestimate this advantage for verbal short-term memory. The most comprehensive data on the memory performance of musicians and nonmusicians come from the meta-analysis by Talamini et al. (2017). They observed an advantage for musical stimuli of Hedges's $g = 1.15$ and for visuospatial stimuli of $g = 0.28$, values that are similar to those observed here. However, the meta-analysis observed an advantage of $g = 0.57$ for verbal stimuli, whereas here, the advantage observed is only $g = 0.18$, out of the CI estimated by the meta-analysis. In contemporary literature, meta-analyses are often used to get an estimate of the temperature of a given phenomenon. Our results suggest that this practice should be taken with some caution. Empirical evidence collected directly by multi-lab studies often contrasts with the meta-analyses of the literature, and in general, meta-analyses tend to observe larger effect sizes than multilabs (Kvarven et al., 2020). Journals often favor publishing positive results, and studies comparing musicians and nonmusicians typically involve a limited number of participants because experienced musicians are rare in the population. When a

phenomenon has a small effect size, smaller experiments are likely to report positive outcomes only if they observe an inflated effect size—essentially, a large but inaccurate difference between groups (e.g., Button et al., 2013). One further positive aspect of the multilab approach is that it uses a shared research protocol, whereas meta-analyses are often forced to compare studies that may differ substantially. Shared protocols reduce heterogeneity, likely producing clearer results.

Another important point that we would like to raise is why we observe many contrasting results in the literature of psychology/neuroscience of music, in particular, when studies contrast musicians and nonmusicians along nonmusical dimensions. Contrasting results likely emerge because researchers collect too few participants, and if participants are few, studies cannot detect subtle effects and phenomena and, above all, cannot provide a reliable estimate of a phenomenon. Contrasting results are rare when musicians and nonmusicians are confronted in musical tasks and even in many auditory tasks, in which the difference in performance between the groups is likely to be large. In these cases, studies are likely to observe musicians' advantage even with a small number of participants (although the estimate of the exact size of this advantage may be uncertain). For example, detecting a large effect size, such as $g = 1.12$ (as observed here for music short-term memory), with substantial statistical power (e.g., 80%) requires recruiting as few as 14 musicians and 14 nonmusicians. In contrast, detecting a smaller effect size, such as $g = 0.28$ (as observed here for visuospatial short-term memory), with the same statistical power would require a much larger sample: 201 musicians and 201 nonmusicians.⁷ In these cases, at the level of the single study, one may observe contrasting results. For example, one study may observe a musician advantage and another study observing a nonmusician advantage. This is clearly evident here. If one observes results at a unit level for verbal and visuospatial short-term memory, the outcome of single units ranges from a nonmusician advantage (occasionally large) to a musician advantage (occasionally large). If literature faces the challenge of understanding whether musicians are cognitively more performing than nonmusicians, studies should aim to have sufficiently large sample sizes.

Another issue in the music psychology/neuroscience literature is the lack of shared standards and protocols. For example, studies in this field often classify participants into two groups: musicians and nonmusicians. However, there is no standard, universally accepted definition of "musician" or "nonmusician," and studies differ significantly in the inclusion and exclusion criteria used to recruit participants. One could also argue that this binary distinction is overly simplistic because many individuals fall between the two categories (e.g., amateur

musicians, music enthusiasts). These individuals are often excluded from studies even though existing tools now allow for the modeling of a "musicianship continuum" (e.g., Gold-MSI). Moreover, it is unclear whether musicians should be studied as a single, homogeneous group or whether specific research questions should target distinct subgroups of musicians. For instance, some musicians play music by manipulating a physical object (e.g., an instrument), whereas others, such as singers, do not. Likewise, some musicians perform melodies, whereas others, such as percussionists, typically do not. Musicians also differ in their training background: Some receive formal music education, whereas others are self-taught (e.g., Jimi Hendrix). These distinctions underscore the diversity within the category of "musician" (e.g., Tervaniemi, 2009). In addition, experimental protocols and measures (e.g., tools for assessing memory performance) vary widely across studies. This heterogeneity often impedes direct comparisons between studies and their results.

Finally, we want to address the possible reasons why musicians perform better than nonmusicians in short-term-memory tasks, as observed here and reported in the literature. We acknowledge that because of the correlational nature of the present study, the current results cannot provide a definitive answer on whether training to become a musician is the cause for this memory advantage. However, this study, involving the largest sample size to date, provides evidence of musicians outperforming nonmusicians, and in some cases, the advantage emerges even after controlling for potential confounding factors, such as SES, personality, age, gender, and education level. Several possibilities might explain the advantage in short-term memory. First, music training might directly improve short-term-memory performance. This improvement may be stronger for closely related abilities (e.g., short-term memory for musical stimuli) and weaker for abilities that are less related to the domain of expertise (e.g., short-term memory for visuospatial or verbal stimuli). The advantage in music short-term memory is substantial, whereas the advantage for visuospatial and verbal short-term memory is small or very small. A limitation of this hypothesis is that the current literature does not provide clear predictions about the magnitude of the advantage that should result from extensive music training (e.g., that of the participants in this study) or whether this advantage should be greater for visuospatial or verbal stimuli. Another possibility is that music training enhances higher-order cognitive functions, which, in turn, influence other cognitive abilities, such as short-term memory. For example, music training might enhance executive functions (see Frischen et al., 2021), which could mediate the relationship between music training and intelligence (Degé et al., 2011). In the current study, musicians exhibited

small advantages in tasks assessing fluid and crystallized intelligence, and executive functions, such as small advantages for musicians in the Raven test (fluid intelligence), WAIS vocabulary scores (crystallized intelligence), and n-back task, which is thought to measure an updating component of working memory. This could have downstream effects on short-term-memory performance. However, the results of our multilevel models suggest that beyond a general cognitive advantage, being a musician predicts superior performance in tasks of music and visuospatial short-term memory (although in music short-term memory, the group was in interaction with crystallized intelligence). Although previous meta-analyses with randomized controlled trials have shown small cognitive benefits of music training (e.g., Jamey et al., 2024; Román-Caballero et al., 2022), the question of whether music training can cause far-transfer (either direct or indirect) effects remains debated (Sala & Gobet, 2020; Schellenberg & Lima, 2024; although the analytical decisions of the latter study were criticized by Bigand & Tillmann, 2022).

A further possible causal explanation, specific to the current experiment and methodology, is that musicians performed better in music tasks because of their training but also outperformed nonmusicians in other tasks (albeit to a lesser degree) because they had received more education overall. Although musicians and non-musicians were matched for general education, this matching excluded musical education. In practice, many musicians in this study had more education overall because they often received extensive music training alongside regular schooling. Research shows that cognitive performance is associated with education (e.g., Ritchie & Tucker-Drob, 2018). Likewise, other groups of experts, such as master chess players, who receive additional education outside regular schooling tend to perform better on cognitive tasks than individuals without such education (e.g., Sala et al., 2017).

Finally, an alternative hypothesis is that the observed musicians' advantage is due to preexisting cognitive abilities, that is, cognitively high-performing individuals may be more likely to become musicians, in particular, individuals of higher SES families (who support their musical careers) and individuals with personality traits that are known to be positively associated with the length of music training (Schellenberg & Lima, 2024; Swaminathan & Schellenberg, 2018). These factors were anyway accounted for in some of the present analyses.

As stated above, the present correlational study cannot resolve the "chicken and egg" problem, and the present discussion has been the result of a vibrant debate among the authors, who as far as interpretations are concerned, did not always share the same views. Nonetheless, the present work does provide robust estimates of the differences between experienced musicians

and nonmusicians and underlines the importance of the open-science practice fully embraced by this study: Although interpretations might diverge, the data set and analysis pipeline are fully accessible, allowing the reader to form an opinion and interpretation starting from a direct and personal analysis.

Final remarks

We would like to explicitly acknowledge some of the limitations of the present study. Above all, this is a correlational study: Any relationship observed here does not imply causation. In other words, we cannot conclude that the music training (or the lack of it) causes any of the relationships observed here: It is impossible to disentangle the effect of training (or the lack of it) from the specific selection process that made one individual a musician. However, the question targeted by the present study (whether music expertise is in a relationship with short-term-memory performance) seems too challenging to be investigated experimentally. One becomes a skilled musician after long practice. If we assume that the musicians of the current study were studying music only 1 hr per day, the number of hours of practice per musician becomes well over 3,000. It is difficult to imagine an experimental study implementing such a long training. A second limitation is that results observed here cannot be extended to other types of memory, such as working memory or long-term memory. The decision to study short-term memory was driven by available data. In the meta-analysis by Talamini et al. (2017), this advantage was the largest and the least variable, suggesting the existence of a true relationship. This is the reason this relationship was investigated. Another limitation is the specific music culture: the Western one. The sample we recruited underwent training that is typical of Western countries in terms of type of music, type of instruments, type of schools, and so on. We cannot conclude that results observed here could be observed by recruiting musicians of non-Western cultures. Finally, the results of the current experiment are valid for a specific age range, that is, young adults, thus making them not generalizable to younger and older ages.

To conclude, in the present study, we showed that musicians have a short-term-memory advantage compared with nonmusicians, an advantage that is large for music stimuli, small for visuospatial stimuli, and very small for verbal stimuli. Musicians were also different from nonmusicians in many other aspects. They performed better in music-perception tasks, were more musically sophisticated, and got more reward from music. But they also had small advantages in tasks tapping intelligence and executive functions, and they came from families of higher SES. Finally, they differed in personality. Above all, they were substantially more

“open to experiences.” The present study sheds light on a topic (i.e., whether music training is associated with enhanced cognitive performance), which revealed weaknesses in the past literature (e.g., heterogeneous designs, small sample sizes, large variation in results). Moreover, the present study can set a standard for future research on the topic, highlighting the importance of collaborative and preregistered studies as a tool to explore controversial but also new phenomena.

Transparency

Action Editor: Katie Corker

Editor: David A. Sbarra

Author Contributions

M. Grassi and F. Talamini share first authorship.

Massimo Grassi: Conceptualization; Data curation; Formal analysis; Methodology; Project administration; Resources; Software; Validation; Visualization; Writing – original draft; Writing – review & editing.

Francesca Talamini: Conceptualization; Data curation; Formal analysis; Methodology; Project administration; Resources; Software; Validation; Visualization; Writing – original draft; Writing – review & editing.

Gianmarco Altoè: Data curation; Formal analysis; Methodology; Writing – review & editing.

Elvira Brattico: Methodology; Resources; Writing – original draft; Writing – review & editing.

Anne Caclin: Methodology; Resources; Software; Writing – original draft; Writing – review & editing.

Barbara Carretti: Conceptualization; Methodology; Writing – original draft; Writing – review & editing.

Véronique Drai-Zerbib: Methodology; Resources; Writing – original draft; Writing – review & editing.

Laura Ferreri: Methodology; Resources; Writing – original draft; Writing – review & editing.

Filippo Gambarota: Data curation; Formal analysis; Methodology; Validation; Writing – review & editing.

Jessica Grahn: Methodology; Resources; Writing – original draft; Writing – review & editing.

Lucrezia Guiotto Nai Fovino: Investigation; Software; Visualization; Writing – review & editing.

Marco Roccato: Software; Visualization; Writing – review & editing.

Antoni Rodriguez-Fornells: Methodology; Resources; Writing – original draft; Writing – review & editing.

Swathi Swaminathan: Methodology; Resources; Software; Writing – original draft; Writing – review & editing.

Barbara Tillmann: Methodology; Resources; Writing – original draft; Writing – review & editing.

Peter Vuust: Methodology; Resources; Writing – original draft; Writing – review & editing.

Jonathan Wilbiks: Methodology; Resources; Writing – original draft; Writing – review & editing.

Marcel Zentner: Methodology; Resources; Writing – original draft; Writing – review & editing.

Karla Aguilar: Investigation.

Christ B. Aryanto: Investigation; Resources; Writing – review & editing.

Frederico C. Assis Leite: Project administration; Resources.

Aïssa M. Baldé: Formal analysis; Investigation; Resources.

Deniz Başkent: Supervision; Writing – review & editing.

Laura Bishop: Formal analysis; Investigation; Writing – review & editing.

Graziela Kalsi: Investigation; Resources; Supervision.

Fleur L. Bouwer: Investigation; Resources; Supervision; Writing – review & editing.

Axelle Calcus: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.

Giulio Carraturo: Investigation.

Victor Cepero-Escribano: Investigation.

Antonia Cerić: Investigation.

Antonio Criscuolo: Formal analysis; Investigation; Resources; Writing – review & editing.

Léo Dairain: Investigation.

Simone Dalla Bella: Investigation; Resources; Supervision; Writing – review & editing.

Oscar Daniel: Investigation.

Anne Danielsen: Formal analysis; Resources; Writing – review & editing.

Anne-Isabelle de Parcevaux: Investigation.

Delphine Dellacherie: Formal analysis; Resources; Supervision; Writing – review & editing.

Tor Endestad: Formal analysis; Resources; Supervision; Writing – review & editing.

Juliana L. d. B. Fialho: Investigation.

Caitlin Fitzpatrick: Investigation.

Anna Fiveash: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.

Juliette Fortier: Investigation.

Noah R. Fram: Formal analysis; Investigation; Writing – review & editing.

Eleonora Fullone: Investigation; Writing – review & editing.

Stefanie Gloggenieß: Investigation.

Lucia Gonzalez Sanchez: Investigation.

Reyna L. Gordon: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.

Mathilde Groussard: Investigation; Resources; Supervision; Writing – review & editing.

Assal Habibi: Formal analysis; Resources; Writing – review & editing.

Heidi M. U. Hansen: Formal analysis.

Eleanor E. Harding: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.

Kirsty Hawkins: Investigation; Resources; Writing – review & editing.

Steffen A. Herff: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.

Veikka P. Holma: Formal analysis.

Kelly Jakubowski: Investigation; Resources; Supervision; Writing – review & editing.

Maria G. Jol: Investigation.

Aarushi Kalsi: Investigation.

Veronica Kandro: Investigation.

Rosaliina Kelo: Formal analysis; Investigation; Writing – review & editing.

Sonja A. Kotz: Investigation; Resources; Supervision.

Gangothri S. Ladegam: Investigation.
Bruno Laeng: Investigation; Resources; Supervision; Writing – review & editing.
André Lee: Investigation; Resources; Supervision.
Miriam Lense: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
César F. Lima: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
Simon P. Limmer: Investigation.
Chengran K. Liu: Investigation.
Paulina d. C. Martín Sánchez: Investigation; Writing – review & editing.
Langley McEntyre: Investigation.
Jessica P. Michael: Investigation.
Daniel Mirman: Resources; Supervision.
Daniel Müllensiefen: Formal analysis; Resources; Supervision; Writing – review & editing.
Nilou Najafi: Investigation.
Jaakko Nokkala: Investigation.
Ndassi Nzonlang: Investigation.
Maria Gabriela M. Oliveira: Supervision.
Katie Overy: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
Andrew J. Oxenham: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
Edoardo Passarotto: Investigation; Resources.
Marie-Elisabeth Plasse: Investigation.
Herve Platel: Formal analysis; Investigation; Writing – review & editing.
Alice Poissonnier: Investigation; Resources.
Neha Rajappa: Investigation.
Michaela Ritchie: Investigation.
Italo Ramon Rodrigues Menezes: Investigation.
Rafael Román-Caballero: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
Paula Roncaglia: Investigation; Resources; Writing – review & editing.
Farrah Y. Sa’adullah: Investigation.
Suvi Saarikallio: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
Daniela Sammler: Formal analysis; Resources; Supervision.
Séverine Samson: Formal analysis; Resources; Supervision; Writing – review & editing.
E. G. Schellenberg: Formal analysis; Supervision; Writing – review & editing.
Nora R. Serres: Investigation; Writing – review & editing.
L. R. Slevc: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
Ragnya-Norasoa Souffiane: Investigation.
Florian J. Strauch: Investigation.
Hannah Strauss: Software; Writing – review & editing.
Nicholas Tantengco: Investigation.
Mari Tervaniemi: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.
Rachel Thompson: Investigation.
Renee Timmers: Investigation; Resources; Supervision; Writing – review & editing.
Petri Toiviainen: Resources; Supervision; Writing – review & editing.

Laurel J. Trainor: Investigation; Resources; Supervision; Writing – review & editing.

Clara Tuske: Investigation.

Jed Villanueva: Investigation; Writing – review & editing.

Claudia C. von Bastian: Investigation; Resources; Writing – review & editing.

Kelly L. Whiteford: Investigation; Writing – review & editing.

Emily A. Wood: Investigation.

Florian Worschech: Formal analysis; Investigation; Resources; Supervision; Writing – review & editing.

Ana Zappa: Resources; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Massimo Grassi  <https://orcid.org/0000-0002-3784-2784>
 Francesca Talamini  <https://orcid.org/0000-0001-6083-0282>
 Gianmarco Altoè  <https://orcid.org/0000-0003-1154-9528>
 Filippo Gambarota  <https://orcid.org/0000-0002-6666-1747>
 Barbara Tillmann  <https://orcid.org/0000-0001-9676-5822>
 Marcel Zentner  <https://orcid.org/0000-0001-8580-8030>
 Victor Cepero-Escribano  <https://orcid.org/0009-0003-6394-8034>
 Andrew J. Oxenham  <https://orcid.org/0000-0002-9365-1157>
 Rafael Román-Caballero  <https://orcid.org/0000-0003-0943-6217>
 E. G. Schellenberg  <https://orcid.org/0000-0003-3681-6020>
 Mari Tervaniemi  <https://orcid.org/0000-0002-9651-2929>
 Claudia C. von Bastian  <https://orcid.org/0000-0002-0667-2460>

Acknowledgments

We commit to share all our raw data upon acceptance for publication of the Stage 2 manuscript. Data, analyses scripts, and materials are stored on OSF, <https://osf.io/y97t3/>. Repository of all the relevant materials of the project (e.g., data, analysis scripts, digital materials, instruction for units and experimenters) is available at https://osf.io/y97t3/?view_only=f115add461524b8f8e62af9cd360c33d.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/25152459251379432>.

Notes

1. Talamini et al. (2017) used the Hedges's g as an index for the effect size. The formula to calculate the g index is almost identical to Cohen's d but adjusted for small sample sizes. Therefore, the two indexes can be interpreted in the same way: $g = 0.2$ (small effect), $g = 0.5$ (medium effect), $g = 0.8$ (large effect; see Sawilowsky, 2009).
2. Inclusion/exclusion criteria were checked at recruitment by the experimenters, but post hoc verification revealed inconsistencies: 137 musicians self-assessed insufficient training, and 85 nonmusicians self-assessed excessive training. To assess their impact, we excluded these participants and their matched counterparts (202 pairs) and recalculated the analyses. The exclusions had negligible effects on the results (see the Supplemental Material available online). Given the robustness of the findings and the lack of an objective method to resolve discrepancies, we retained all participants in the main analyses.
3. At the end of data collection, nine pairs (with respect to age matching) and six pairs (with respect to education matching) violated the matching criteria but were nevertheless included in the final data set. In these 15 pairs, the differences in age or education were within ± 2 years.
4. After data collection began, an error was found in Block 2 of the melody span: The comparison melody differed from the standard in multiple notes instead of two consecutive ones (see Method section). The block was corrected in mid-September 2023. Of the 1,200 participants, 67 completed the task with the incorrect stimuli, and their melody-span score was based solely on Block 1 responses.
5. Forty-six participants reported "no income" for both parents. The Hollingshead Four Factor Index of Social Status (Hollingshead, 1975) cannot be calculated under these circumstances. Therefore, an SES score of 8, the minimum possible value of the Hollingshead index, was assigned to these participants.
6. After data collection began, we discovered that two items of the BFI-2 were missing (i.e., Item 1, extraversion: "I am someone who is outgoing, sociable" and Item 33, conscientiousness: "I am someone who keeps things neat and tidy"). These items were added in mid-September 2023. Participants who completed the incomplete version were individually contacted to provide responses, and all but 12 responded. For these 12 participants, the missing responses were imputed with the neutral option, corresponding to the midpoint of the BFI-2 Likert scale.
7. These calculations hypothesize that the performance of musicians and nonmusicians is compared with an independent-samples t test.

References

- Amer, T., Kalender, B., Hasher, L., Trehub, S. E., & Wong, Y. (2013). Do older professional musicians have cognitive advantages? *PLOS ONE*, 8(8), Article e71630. <https://doi.org/10.1371/journal.pone.0071630>
- Andrews, M. W., Dowling, W. J., Bartlett, J. C., & Halpern, A. R. (1998). Identification of speeded and slowed familiar melodies by younger, middle-aged, and older musicians and nonmusicians. *Psychology and Aging*, 13(3), 462–471. <https://doi.org/10.1037/0882-7974.13.3.462>
- Arndt, C., Schlemmer, K., & Van der Meer, E. (2023). The relationship of musical expertise, working memory, and intelligence. *Music Perception: An Interdisciplinary Journal*, 40(4), 334–346.
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559.
- Baker, D. J., Ventura, J., Calamia, M., Shanahan, D., & Elliott, E. M. (2020). Examining musical sophistication: A replication and theoretical commentary on the Goldsmiths Musical Sophistication Index. *Musicae Scientiae*, 24(4), 411–429. <https://doi.org/10.1177/1029864918811879>
- Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, 139(3), EL51–EL56.
- Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine*, 14(4), 395–411. <https://doi.org/10.1002/sim.4780140406>
- Bigand, E., & Tillmann, B. (2022). Near and far transfer: Is music special? *Memory & Cognition*, 50(2), 339–347. <https://doi.org/10.3758/s13421-021-01226-6>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470743386>
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.
- Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory: Insights from neurodevelopmental and neurological disorders. *Annals of the New York Academy of Sciences*, 1423(1), 155–165. <https://doi.org/10.1111/nyas.13733>
- Cardona, G., Ferreri, L., Lorenzo-Seva, U., Russo, F. A., & Rodriguez-Fornells, A. (2022). The forgotten role of absorption in music reward. *Annals of the New York Academy of Sciences*, 1514(1), 142–154. <https://doi.org/10.1111/nyas.14790>
- Chang, Y. (2014). Reorganization and plastic changes of the human brain associated with skill learning and expertise. *Frontiers in Human Neuroscience*, 8, Article 35. <https://doi.org/10.3389/fnhum.2014.00035>
- Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, 21(2), 111–124. <https://doi.org/10.1016/j.tics.2016.12.007>
- Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G., Jr. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLOS ONE*, 11(7), Article e0157638. <https://doi.org/10.1371/journal.pone.0157638>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences*, 39(5), 1005–1014.
- Corrigall, K. A., & Schellenberg, E. G. (2015). Predicting who takes music lessons: Parent and child characteristics.

- Frontiers in Psychology*, 6, Article 282. <https://doi.org/10.3389/fpsyg.2015.00282>
- Corrigall, K. A., Schellenberg, E. G., & Misura, N. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, 4, Article 222. <https://doi.org/10.3389/fpsyg.2013.00222>
- Criscuolo, A., Bonetti, L., Särkämö, T., Kliuchko, M., & Brattico, E. (2019). On the association between musical training, intelligence and executive functions in adulthood. *Frontiers in Psychology*, 10, Article 1704. <https://doi.org/10.3389/fpsyg.2019.01704>
- Criscuolo, A., Pando-Naude, V., Bonetti, L., Vuust, P., & Brattico, E. (2022). An ALE meta-analytic review of musical expertise. *Scientific Reports*, 12(1), Article 11726. <https://doi.org/10.1038/s41598-022-14959-4>
- De Groot, A. D. (2014). *Thought and choice in chess*. De Gruyter Mouton.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: A relation mediated by executive functions. *Music Perception*, 29(2), 195–201.
- Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research*, 1455, 75–89. <https://doi.org/10.1016/j.brainres.2012.03.034>
- DeYoung, C. G., Quilty, L. C., Peterson, J. B., & Gray, J. R. (2013). Openness to experience, intellect, and cognitive ability. *Journal of Personality Assessment*, 96(1), 46–52. <https://doi.org/10.1080/00223891.2013.806327>
- Fennell, A. M., Bugos, J. A., Payne, B. R., & Schotter, E. R. (2021). Music is similar to language in terms of working memory interference. *Psychonomic Bulletin & Review*, 28(2), 512–525. <https://doi.org/10.3758/s13423-020-01833-5>
- Floyd, R. G., Farmer, R. L., Schneider, W. J., & McGrew, K. S. (2021). Theories and measurement of intelligence. In L. M. Glidden, L. Abbeduto, L. L. McIntyre, & M. J. Tassé (Eds.), *APA handbook of intellectual and developmental disabilities: Foundations* (pp. 385–424). American Psychological Association. <https://doi.org/10.1037/0000194-015>
- Frischen, U., Schwarzer, G., & Degé, F. (2021). Music lessons enhance executive functions in 6-to 7-year-old children. *Learning and Instruction*, 74, 101442. <https://doi.org/10.1016/j.learninstruc.2021.101442>
- Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410), 398–409. <https://doi.org/10.1080/01621459.1990.10476213>
- George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, 49(5), 1083–1094. <https://doi.org/10.1016/j.neuropsychologia.2011.02.001>
- Gjermunds, N., Brechan, I., Johnsen, S., & Watten, R. G. (2020). Personality traits in musicians. *Current Issues in Personality Psychology*, 8(2), 100–107.
- Grassi, M., Meneghetti, C., Toffalini, E., & Borella, E. (2017). Auditory and cognitive performance in elderly musicians and nonmusicians. *PLOS ONE*, 12(11), Article e0187881. <https://doi.org/10.1371/journal.pone.0187881>
- Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1995). Aging and experience in the recognition of musical transpositions. *Psychology and Aging*, 10(3), 325–342. <https://doi.org/10.1037/0882-7974.10.3.325>
- Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466–472. <https://doi.org/10.1016/j.tics.2007.08.008>
- Hansen, H. M. U., Røysamb, E., Vassend, O. M., Czajkowski, N. O., Endestad, T., Danielsen, A., & Laeng, B. (2024). The underlying architecture of musical sensibility: One general factor, four subdimensions, and strong genetic effects. *Annals of the New York Academy of Sciences*, 1540(1), 291–306. <https://doi.org/10.1111/nyas.15227>
- Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, 41(6), 779–793. <https://doi.org/10.1177/0305735612452186>
- Harris, J. A. (2004). Measured intelligence, achievement, openness to experience, and creativity. *Personality and Individual Differences*, 36(4), 913–929.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics: A Quarterly Publication Sponsored by the American Educational Research Association and the American Statistical Association*, 6(2), 107–128. <https://doi.org/10.3102/10769986006002107>
- Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6(3), 203–217. <https://www.ncbi.nlm.nih.gov/pubmed/11570228>
- Hennessy, S., Mack, W. J., & Habibi, A. (2022). Speech-in-noise perception in musicians and non-musicians: A multi-level meta-analysis. *Hearing Research*, 416, Article 108442. <https://doi.org/10.1016/j.heares.2022.108442>
- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, 76(3), 486–502. <https://doi.org/10.1016/j.neuron.2012.10.011>
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Hollingshead, A. B. (1975). *Four factor index of social status*. Yale University. https://sociology.yale.edu/sites/default/files/files/yjs_fall_2011.pdf#page=21
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience*, 29(10), 3019–3025. <https://doi.org/10.1523/JNEUROSCI.5118-08.2009>
- Jamey, K., Foster, N. E., Hyde, K. L., & Dalla Bella, S. (2024). Does music training improve inhibition control in children? A systematic review and meta-analysis. *Cognition*, 252, Article 105913. <https://doi.org/10.1016/j.cognition.2024.105913>

- Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biology Reports*, 1, Article 78. <https://doi.org/10.3410/B1-78>
- Jansen, N., Harding, E. E., Loerts, H., Başkent, D., & Lowie, W. (2023). The relation between musical abilities and speech prosody perception: A meta-analysis. *Journal of Phonetics*, 101, Article 101278. <https://doi.org/10.1016/j.wocn.2023.101278>
- Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport*, 10(6), 1309–1313. <https://doi.org/10.1097/00001756-199904260-00029>
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Review Neuroscience*, 11, 599–605. <https://doi.org/10.1038/nrn2882>
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Kvarven, A., Strømmland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, 4(4), 423–434. <https://doi.org/10.1038/s41562-019-0787-z>
- Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLOS ONE*, 7(12), Article e52508. <https://doi.org/10.1371/journal.pone.0052508>
- Lee, Y., Lu, M., & Ko, H. (2007). Effects of skill training on working memory capacity. *Learning and Instruction*, 17(3), 336–344. <https://doi.org/10.1016/j.learninstruc.2007.02.010>
- Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition*, 155, 113–124. <https://doi.org/10.1016/j.cognition.2016.06.013>
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97(8), 4398–4403. <https://doi.org/10.1073/pnas.070039597>
- Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodríguez-Fornells, A. (2012). Individual differences in music reward experiences. *Music Perception*, 31(2), 118–138.
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1–2), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Miller, L. T., & Vernon, P. A. (1992). The general factor in short-term memory, intelligence, and reaction time. *Intelligence*, 16(1), 5–29.
- 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), Article e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Münste, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3(6), 473–478. <https://doi.org/10.1038/nrn843>
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134), Article 20170213. <https://doi.org/10.1098/rsif.2017.0213>
- Neves, L., Correia, A. I., Castro, S. L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience & Biobehavioral Reviews*, 140, Article 104777. <https://doi.org/10.1016/j.neubiorev.2022.104777>
- Nozari, N., & Martin, R. C. (2024). Is working memory domain-general or domain-specific? *Trends in Cognitive Sciences*, 28(11), 1023–1036. <https://doi.org/10.1016/j.tics.2024.06.006>
- Okada, B. M., & Slevc, L. R. (2018). Individual differences in musical training and executive functions: A latent variable approach. *Memory & Cognition*, 46(7), 1076–1092. <https://doi.org/10.3758/s13421-018-0822-8>
- Okhrei, A., Kutsenko, T., & Makarchuk, M. (2016). Performance of working memory of musicians and non-musicians in tests with letters, digits, and geometrical shapes. *Biologija*, 62(4), Article 4. <https://doi.org/10.6001/biologija.v62i4.3408>
- Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* [Doctoral dissertation, The Ohio State University]. Ohio LINK. http://rave.ohiolink.edu/etdc/view?acc_num=osu1161705351
- Osler, M., Avlund, K., & Mortensen, E. L. (2013). Socio-economic position early in life, cognitive development and cognitive change from young adulthood to middle age. *The European Journal of Public Health*, 23(6), 974–980. <https://doi.org/10.1093/eurpub/cks140>
- 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), Article e11120. <https://doi.org/10.1371/journal.pone.0011120>
- Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience*, 29(45), 14100–14107. <https://doi.org/10.1523/JNEUROSCI.3256-09.2009>
- Pascual-Leone, A., & Torres, F. (1993). Plasticity of the sensorimotor cortex representation of the reading finger in Braille readers. *Brain*, 116(1), 39–52. <https://doi.org/10.1093/brain/116.1.39>
- Pastore, M., Nucci, M., & Bobbio, A. (2015). Vita di p: 16 anni di statistiche sul GIP. *Giornale italiano di psicologia (Italian Journal of Psychology)*, 42(1–2), 303–328. <https://doi.org/10.1421/79851>
- Ramachandra, V., Meighan, C., & Gradzki, J. (2012). The impact of musical training on the phonological memory and the central executive: A brief report. *North American Journal of Psychology*, 14(3), 541–541.
- Rammseyer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians.

- Music Perception*, 24(1), 37–48. <https://doi.org/10.1525/mp.2006.24.1.37>
- Raven, J. C., Raven, J. C., & Court, J. H. (1965). *Advanced progressive Sets I and II*. London: H. K. Lewis.
- Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, 29(8), 1358–1369.
- Rodrigues, A. C., Loureiro, M., & Caramelli, P. (2014). Visual memory in musicians and non-musicians. *Frontiers in Human Neuroscience*, 8, Article 424. <https://doi.org/10.3389/fnhum.2014.00424>
- Román-Caballero, R., & Lupiáñez, J. (2022). Suggestive but not conclusive: An independent meta-analysis on the auditory benefits of learning to play a musical instrument. Commentary on Neves et al. (2022). *Neuroscience & Biobehavioral Reviews*, 142, Article 104916. <https://doi.org/10.1016/j.neubiorev.2022.104916>
- Román-Caballero, R., Vadillo, M. A., Trainor, L. J., & Lupiáñez, J. (2022). Please don't stop the music: A meta-analysis of the cognitive and academic benefits of instrumental musical training in childhood and adolescence. *Educational Research Review*, 35, Article 100436. <https://doi.org/10.1016/j.edurev.2022.100436>
- Sala, G., Burgoyne, A. P., Macnamara, B. N., Hambrick, D. Z., Campitelli, G., & Gobet, F. (2017). Checking the “Academic Selection” argument: Chess players outperform non-chess players in cognitive skills related to intelligence: A meta-analysis. *Intelligence*, 61, 130–139.
- Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, 48(8), 1429–1441. <https://doi.org/10.3758/s13421-020-01060-2>
- Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), Article 26. <https://doi.org/10.22237/jmasm/1257035100>
- Schellenberg, E. G. (2020). Music training, individual differences, and plasticity. In M. S. C. Thomas, D. Mareschal, & I. Dumontheil (Eds.), *Educational neuroscience* (pp. 415–441). Routledge.
- Schellenberg, E. G., & Lima, C. F. (2024). Music training and nonmusical abilities. *Annual Review of Psychology*, 75(1), 87–128.
- Schellenberg, E. G., & Moreno, S. (2010). Music lessons, pitch processing, and g. *Psychology of Music*, 38(2), 209–221. <https://doi.org/10.1177/0305735609339473>
- Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 499–550). Elsevier.
- Schlaug, G. (2001). The brain of musicians. *Annals of the New York Academy of Sciences*, 930(1), 281–299. <https://doi.org/10.1111/j.1749-6632.2001.tb05739.x>
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language: Music and prosody: An ERP study. *Psychophysiology*, 41(3), 341–349. <https://doi.org/10.1111/1469-8986.00172.x>
- Schulze, K., Jay Dowling, W., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception*, 29(3), 255–267. <https://doi.org/10.1525/mp.2012.29.3.255>
- Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians: Neural correlates of strategy use. *European Journal of Neuroscience*, 33(1), 189–196. <https://doi.org/10.1111/j.1460-9568.2010.07470.x>
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, 32(5), 771–783. <https://doi.org/10.1002/hbm.21060>
- Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, 152, 199–211. <https://doi.org/10.1016/j.cognition.2016.03.017>
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), Article 160384. <https://doi.org/10.1098/rsos.160384>
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113, 117–143.
- Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology*, 15, 201–292.
- Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. *Hearing Research*, 308, 109–121. <https://doi.org/10.1016/j.heares.2013.08.004>
- Suárez, L., Elangovan, S., & Au, A. (2016). Cross-sectional study on the relationship between music training and working memory in adults. *Australian Journal of Psychology*, 68(1), 38–46.
- Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Reports*, 8(1), Article 9223. <https://doi.org/10.1038/s41598-018-27571-2>
- Swaminathan, S., & Schellenberg, E. G. (2019). Music training and cognitive abilities: Associations, causes, and consequences. In M. H. Thaut, & D. A. Hodges (Eds.), *The Oxford handbook of music and the brain* (pp. 644–670). Oxford University Press. <https://doi.org/10.1093/oxfordhpb/9780198804123.013.26>
- Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence*, 62, 119–124. <https://doi.org/10.1016/j.intell.2017.03.005>
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLOS ONE*, 12(10), Article e0186773. <https://doi.org/10.1371/journal.pone.0186773>
- Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2022). Auditory and visual short-term memory: Influence of material type, contour, and musical expertise. *Psychological Research*, 86(2), 421–442. <https://doi.org/10.1007/s00426-021-01519-0>
- Talamini, F., Carretti, B., & Grassi, M. (2016). The working memory of musicians and nonmusicians. *Music Perception*, 34(2), 183–191. <https://doi.org/10.1525/mp.2016.34.2.183>

- Tervaniemi, M. (2009). Musicians—Same or different? *Annals of the New York Academy of Sciences*, 1169(1), 151–156.
- Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, 41(1), 1–32. <https://doi.org/10.1037/h0075959>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Vincenzi, M., Correia, A. I., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2024). Associations between music training and cognitive abilities: The special case of professional musicians. *Psychology of Aesthetics, Creativity, and the Arts*, 18(6), 895–903. <https://doi.org/10.1037/aca0000481>
- Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes ‘expert’ in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied Cognitive Psychology*, 24(6), 812–826. <https://doi.org/10.1002/acp.1588>
- Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20(3), 188–196. <https://doi.org/10.1016/j.lindif.2010.02.004>
- Wechsler, D. (1997). *WAIS-III. Administration and scoring manual*. The Psychological Corporation.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—fourth edition (WAIS-IV)*. Pearson.
- Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, 1400(1), 33–45. <https://doi.org/10.1111/nyas.13410>
- Zhang, J. D., & Schubert, E. (2019). A single item measure for identifying musician and nonmusician categories based on measures of musical sophistication. *Music Perception*, 36(5), 457–467. <https://doi.org/10.1525/mp.2019.36.5.457>
- Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, 48(3), 389–409. <https://doi.org/10.1177/030573561880403>