Part 3: Guidelines for Reporting Human-Subjects Research

Claire Arthur, Nat Condit-Schultz,
David R.W. Sears, John Ashley Burgoyne, Joshua Albrecht

International Society of Music Information Retrieval (ISMIR)
San Francisco, CA

November 10, 2024

Audience

(Shared) Motivations

- Design/analyze/publish research involving human subjects.
- Adhere to APA guidelines with regard to ethics, transparency, and openness.

ISMIR/TISMIR/IEEE

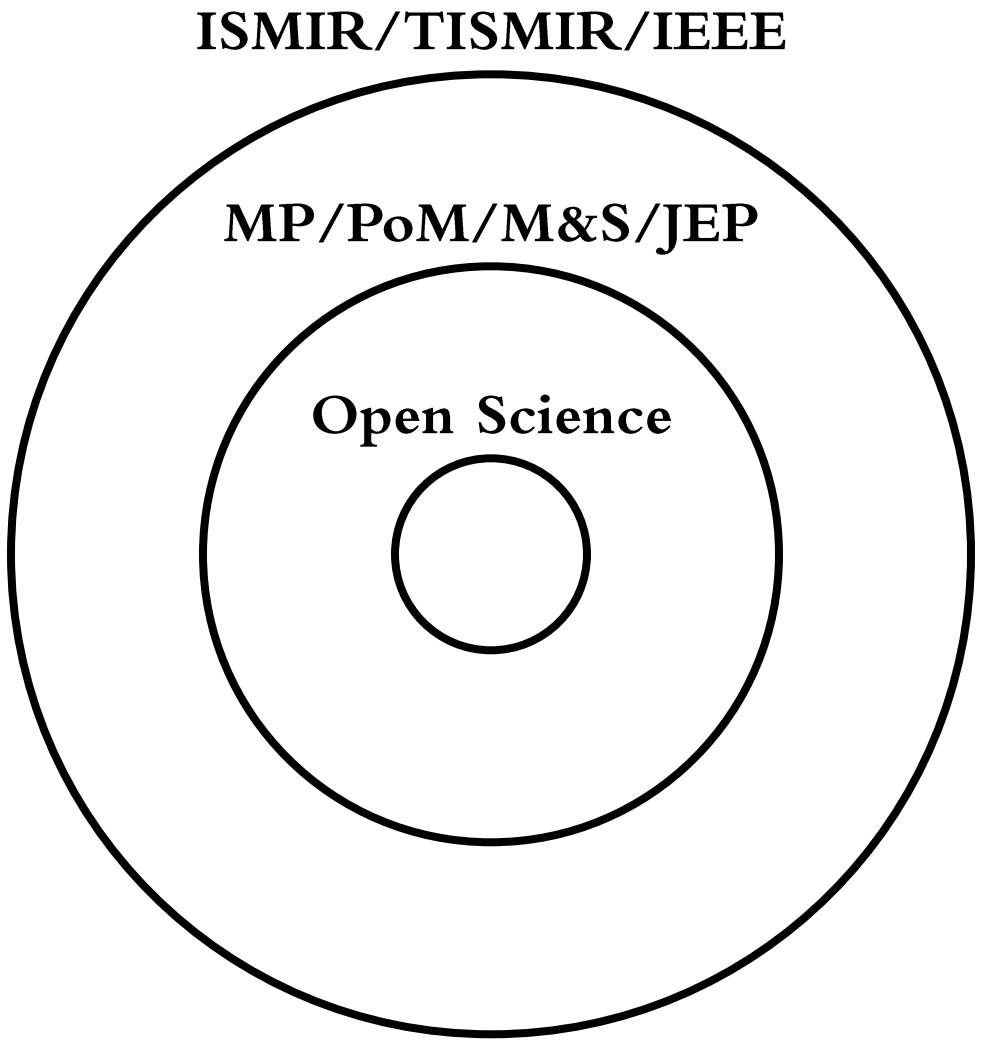
- Targeting venues that don't typically publish behavioral research.

MP/PoM/M&S/JEP

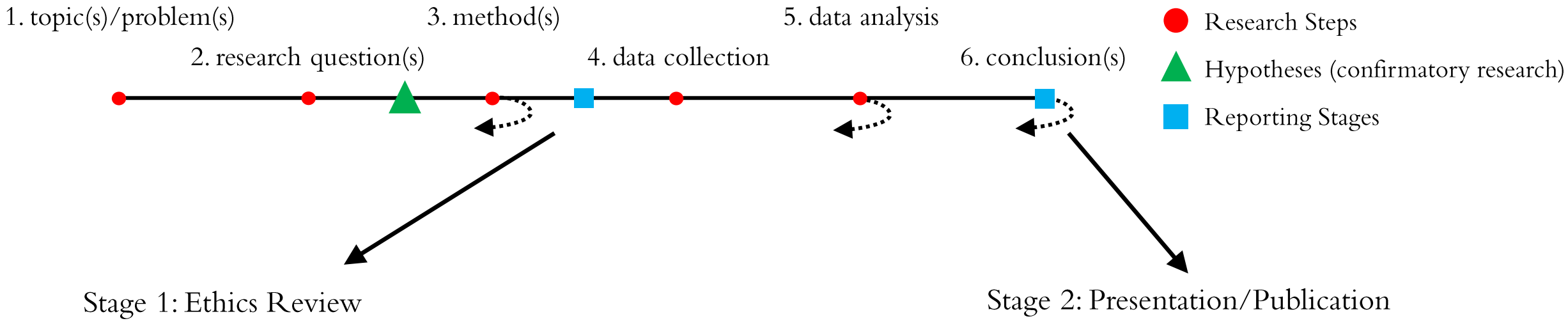
- Targeting venues that typically publish behavioral research.

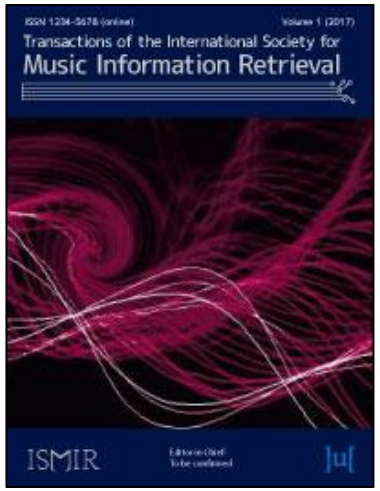
Open Science

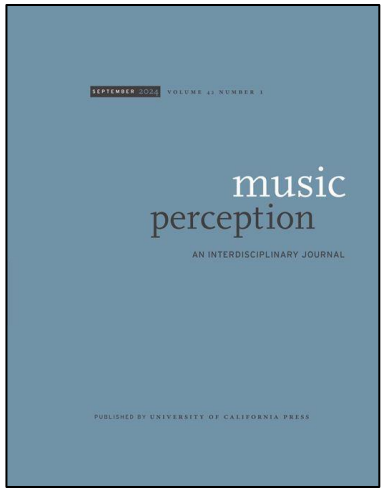
- Developing projects that adhere to new publication standards (e.g., pre-registration).

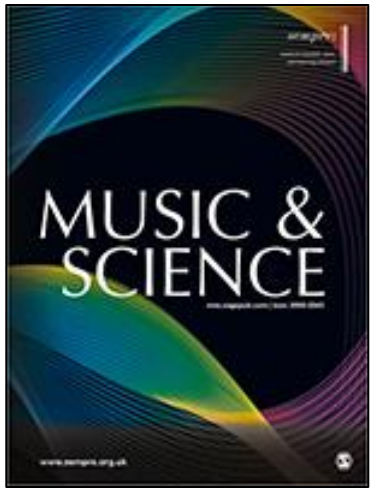


Research Life Cycle (with Human Subjects)

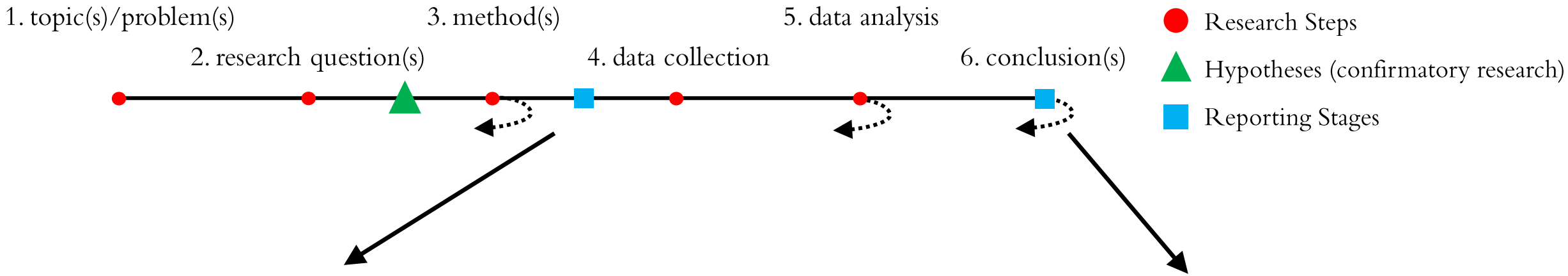




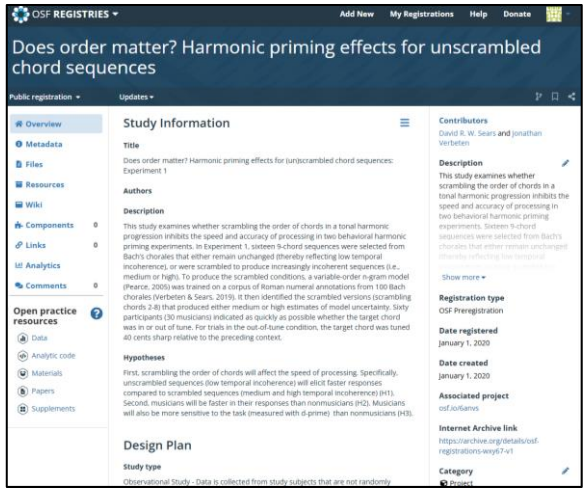




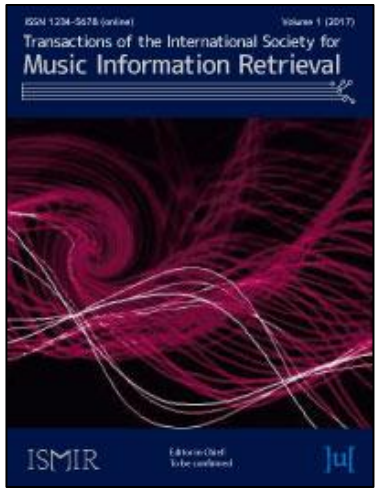
Research Life Cycle (with Human Subjects)

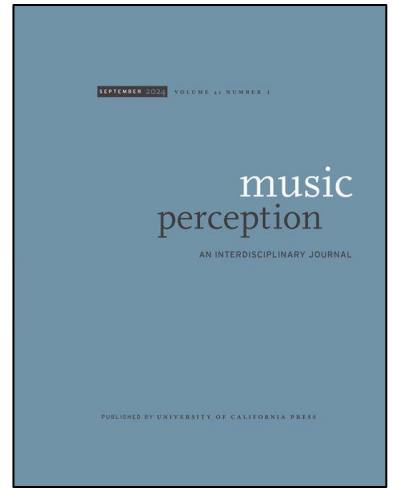


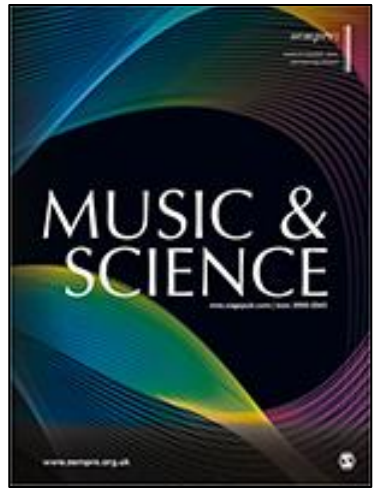
Stage 1: Ethics Review & Pre-Registration



Stage 2: Presentation/Publication







Outline

- **Stage 2: Journal Article Reporting Standards (JARS)**
  - Resources (Quant, REC)
  - APA Manuscript (Layout)
  - Method Section (Participants, Materials, Procedure, etc.)
  - Results Section (Statistics, Visualizations, etc.)

- **Stage 1: Research Ethics and Approval**
  - Drafting Proposals
  - Cross-Cultural Research Guidelines

- **Stages 1-2: New Publication Standards**
  - Supplemental Material (Open Resources)
  - Pre-registrations
  - Registered Reports

Outline

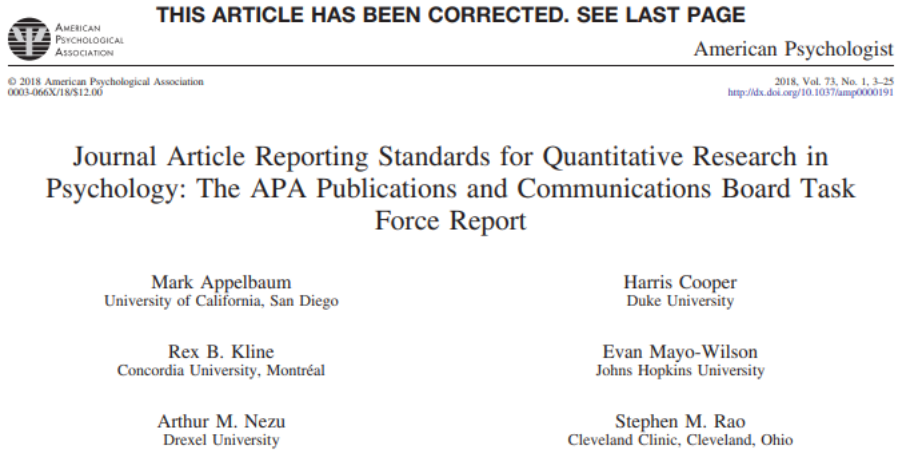
- **Stage 2: Journal Article Reporting Standards (JARS)**
  - Resources (Quant, REC)
  - APA Manuscript (Layout)
  - Method Section (Participants, Materials, Procedure, etc.)
  - Results Section (Statistics, Visualizations, etc.)

- **Stage 1: Research Ethics and Approval**
  - Drafting Proposals
  - Cross-Cultural Research Guidelines

- **Stages 1-2: New Publication Standards**
  - Supplemental Material (Open Resources)
  - Pre-registrations
  - Registered Reports

Stage 2: Journal Article Reporting Standards (JARS)

Resources


THIS ARTICLE HAS BEEN CORRECTED. SEE LAST PAGE

American Psychological Association

American Psychologist

0003-066X/18/$12.00

2018, Vol. 73, No. 1, 3–25
http://dx.doi.org/10.1037/amp0000191

Journal Article Reporting Standards for Quantitative Research in Psychology: The APA Publications and Communications Board Task Force Report

Mark Appelbaum
University of California, San Diego

Harris Cooper
Duke University

Rex B. Kline
Concordia University, Montréal

Evan Mayo-Wilson
Johns Hopkins University

Arthur M. Nezu
Drexel University

Stephen M. Rao
Cleveland Clinic, Cleveland, Ohio

JARS Basics

- https://apastyle.apa.org/jars
- Applebaum et al. (2018; repo)

Quantitative Designs

- https://apastyle.apa.org/jars/quant-table-1.pdf (repo)
- https://apastyle.apa.org/jars/quant-table-2.pdf (repo)
- https://apastyle.apa.org/jars/jars-quant-decision-flowchart.pdf (repo)
- https://apastyle.apa.org/jars/jars-quant-participant-flowchart.pdf (repo)

Race, Ethnicity, and Culture

- https://apastyle.apa.org/jars/rec-table-1.pdf (repo)

APA Author Resource Center

- https://www.apa.org/pubs/journals/resources/authors-resources

Stage 2: Journal Article Reporting Standards (JARS)

Resources

JARS Basics

- https://apastyle.apa.org/jars
- Applebaum et al. (2018; repo)

Quantitative Designs

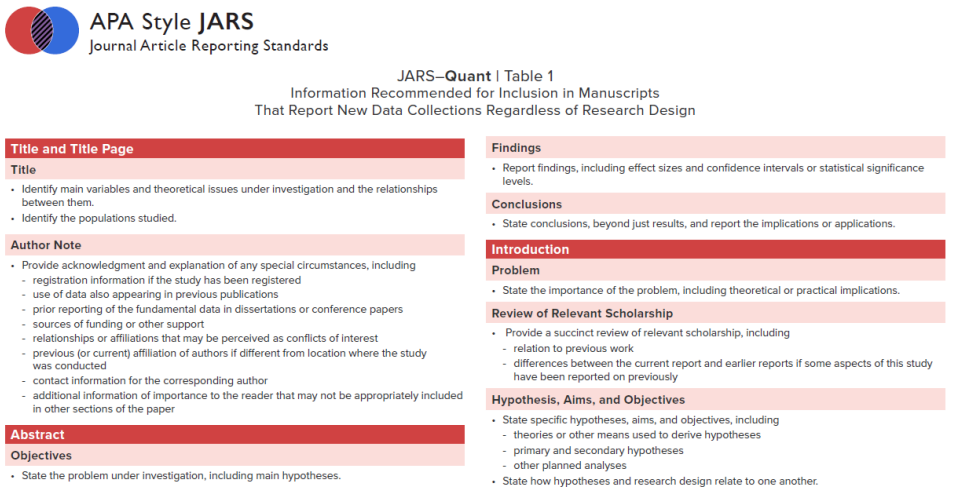
- https://apastyle.apa.org/jars/quant-table-1.pdf (repo)
- https://apastyle.apa.org/jars/quant-table-2.pdf (repo)
- https://apastyle.apa.org/jars/jars-quant-decision-flowchart.pdf (repo)
- https://apastyle.apa.org/jars/jars-quant-participant-flowchart.pdf (repo)

Race, Ethnicity, and Culture

- https://apastyle.apa.org/jars/rec-table-1.pdf (repo)

APA Author Resource Center

- https://www.apa.org/pubs/journals/resources/authors-resources



APA Style JARS
Journal Article Reporting Standards

JARS–**Quant** | Table 1
Information Recommended for Inclusion in Manuscripts
That Report New Data Collections Regardless of Research Design

Title and Title Page

Title

- Identify main variables and theoretical issues under investigation and the relationships between them.
- Identify the populations studied.

Author Note

- Provide acknowledgment and explanation of any special circumstances, including
  - registration information if the study has been registered
  - use of data also appearing in previous publications
  - prior reporting of the fundamental data in dissertations or conference papers
  - sources of funding or other support
  - relationships or affiliations that may be perceived as conflicts of interest
  - previous (or current) affiliation of authors if different from location where the study was conducted
  - contact information for the corresponding author
  - additional information of importance to the reader that may not be appropriately included in other sections of the paper

Abstract

Objectives

- State the problem under investigation, including main hypotheses.

Findings

- Report findings, including effect sizes and confidence intervals or statistical significance levels.

Conclusions

- State conclusions, beyond just results, and report the implications or applications.

Introduction

Problem

- State the importance of the problem, including theoretical or practical implications.

Review of Relevant Scholarship

- Provide a succinct review of relevant scholarship, including
  - relation to previous work
  - differences between the current report and earlier reports if some aspects of this study have been reported on previously

Hypothesis, Aims, and Objectives

- State specific hypotheses, aims, and objectives, including
  - theories or other means used to derive hypotheses
  - primary and secondary hypotheses
  - other planned analyses
- State how hypotheses and research design relate to one another.

Stage 2: Journal Article Reporting Standards (JARS)

APA Manuscript

Resources

- English-Language Services + Sample Papers
  - https://www.apa.org/pubs/journals/resources/editing-services
  - https://apastyle.apa.org/style-grammar-guidelines/paper-format/professional-annotated.pdf (repo)

Anatomy of a Journal Article

- https://apastyle.apa.org/style-grammar-guidelines/paper-format
  - Title Page
  - Abstract
  - Text (Introduction, Method, Results, Discussion, Conclusion)
  - References
  - Footnotes
  - Tables
  - Figures
  - Appendices

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**
- Sampling & Recruitment
- Sample Size
- Participant Characteristics
- Inventories/Instruments
- Ethics

★ **Materials & Design**
- Experimental Conditions

★ **Procedure**
- Data Collection

Apparatus
- Equipment

Analysis
- Analysis Strategy

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**

- Sampling & Recruitment – study design (e.g., between-subjects), site selection
- Sample Size – power ($1 - \beta$ in NHST, where β refers to the Type II error)
- Participant Characteristics – interindividual differences (descriptive/inferential)
- Inventories/Instruments – interindividual differences (descriptive/inferential)
- Ethics – ethics approval (e.g., IRB2023-85)

★ **Materials & Design**

- Experimental Conditions

★ **Procedure**

- Data Collection

Apparatus

- Equipment

Analysis

- Analysis Strategy

Stage 2: Journal Article Reporting Standards (JARS)

Method Section – Eerola & Vuoskoski (2011)

Research Question

- How do the discrete and dimensional models account for ratings of expressed emotion during music listening?

Task

- Rate discrete and dimensional emotions using likert scales for 110 film soundtrack excerpts.

Dependent Variables

- Emotion ratings (×8; 9-point Likert)

Independent Variables

- Block (*order 1*, *order 2*)
- Dimensional categories (arousal, valence, …)
- Discrete categories (happy, sad,…)

Psychology of Music
39(1) 18–49

DOI: 10.1177/0305735610362821
http://pom.sagepub.com
SAGE

A comparison of the discrete and dimensional models of emotion in music

Tuomas Eerola and Jonna K. Vuoskoski
University of Jyväskylä, Finland

Abstract

The primary aim of the present study was to systematically compare perceived emotions in music using two different theoretical frameworks: the discrete emotion model, and the dimensional model of affect. A secondary aim was to introduce a new, improved set of stimuli for the study of music-mediated emotions. A large pilot study established a set of 110 film music excerpts, half were moderately and highly representative examples of five discrete emotions (anger, fear, sadness, happiness and tenderness), and the other half moderate and high examples of the six extremes of three bipolar dimensions (valence, energy arousal and tension arousal). These excerpts were rated in a listening experiment by 116 non-musicians. All target emotions of highly representative examples in both conceptual sets were discriminated by self-ratings. Linear mapping techniques between the discrete and dimensional models revealed a high correspondence along two central dimensions that can be labelled as valence and arousal, and the three dimensions could be reduced to two without significantly reducing the goodness of fit. The major difference between the discrete and categorical models concerned the poorer resolution of the discrete model in characterizing emotionally ambiguous examples. The study offers systematically structured and rich stimulus material for exploring emotional processing.

Key words

battery, dimensional, music, discrete, emotion, three-dimensional

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Participants. The participants for the main experiment were 116 university students aged 18–42 years (mean 24.7, SD 3.75, 68% females and 32% males). Forty-eight percent of the participants did not play any instrument and were not musically trained, 41% had experience of playing an instrument or some level of musical training, and 11% were in between with music as a hobby for less than three years. The participants received cinema tickets in return for their participation, and a number of individual variables were collected from each of them. First, a survey of their musical taste was made using a localized version of the STOMP questionnaire (Rentfrow & Gosling, 2003). Second, their current mood at the time of taking the test was evaluated using the POMS-A questionnaire (Terry, Lane, Lane, & Keohane, 1999). Third, an assessment was made of the participants' personality traits using a 44-item personality measure known as 'The Big Five Inventory' (John & Srivastava, 1999). And finally a short questionnaire was given out to gather information about the participants' film genre preferences, musical training, and any hearing problems.

– Eerola & Vuoskoski (2011)

Sampling & Recruitment | Participant Characteristics | Sample Size (Power) | Inventories/Instruments

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Participants. The participants for the main experiment were 116 university students aged 18–42 years (mean 24.7, SD 3.75, 68% females and 32% males). Forty-eight percent of the participants did not play any instrument and were not musically trained, 41% had experience of playing an instrument or some level of musical training, and 11% were in between with music as a hobby for less than three years. The participants received cinema tickets in return for their participation, and a number of individual variables were collected from each of them. First, a survey of their musical taste was made using a localized version of the STOMP questionnaire (Rentfrow & Gosling, 2003). Second, their current mood at the time of taking the test was evaluated using the POMS-A questionnaire (Terry, Lane, Lane, & Keohane, 1999). Third, an assessment was made of the participants' personality traits using a 44-item personality measure known as 'The Big Five Inventory' (John & Srivastava, 1999). And finally a short questionnaire was given out to gather information about the participants' film genre preferences, musical training, and any hearing problems.

– Eerola & Vuoskoski (2011)

Sampling & Recruitment | Participant Characteristics | Sample Size (Power) | Inventories/Instruments

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Gender – free response†

Power? – G*Power, `simr`

Participants. The participants for the main experiment were 116 university students aged 18–42 years (mean 24.7, SD 3.75, 68% females and 32% males). Forty-eight percent of the participants did not play any instrument and were not musically trained, 41% had experience of playing an instrument or some level of musical training, and 11% were in between with music as a hobby for less than three years. The participants received cinema tickets in return for their participation, and a number of individual variables were collected from each of them. First, a survey of their musical taste was made using a localized version of the STOMP questionnaire (Rentfrow & Gosling, 2003). Second, their current mood at the time of taking the test was evaluated using the POMS-A questionnaire (Terry, Lane, Lane, & Keohane, 1999). Third, an assessment was made of the participants' personality traits using a 44-item personality measure known as 'The Big Five Inventory' (John & Srivastava, 1999). And finally a short questionnaire was given out to gather information about the participants' film genre preferences, musical training, and any hearing problems.

Site?

Ethics?

– Eerola & Vuoskoski (2011)

Sampling & Recruitment | Participant Characteristics | Sample Size (Power) | Inventories/Instruments

† Reporting Gender – https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/gender

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**

- Sampling & Recruitment – study design (e.g., between-subjects), site selection
- Sample Size – power ($1 - \beta$ in NHST, where β refers to the Type II error)
- Participant Characteristics – interindividual differences (descriptive/inferential)
- Inventories/Instruments – interindividual differences (descriptive/inferential)
- Ethics – ethics approval (e.g., IRB2023-85)

★ **Materials & Design**

- Experimental Conditions – conditions/manipulations, visualization(s), access/repo

★ **Procedure**

- Data Collection – session walkthrough, task(s), trials and study design

Apparatus

- Equipment – data collection site(s), equipment/tools

Analysis

- Analysis Strategy – measures, inferential statistics, experiment-wise error correction

Stage 2: Journal Article Reporting Standards (JARS)

Method Section – Sears, Verbeten, & Percival (2023)

Research Question

- Does scrambling the order of chords in a tonal harmonic progression slow the speed of processing?

Task

- Speeded Intonation Discrimination (*in*, *out*)

Dependent Variables

- Discrimination (*d'*; all trials)
- Bias (*c*; all trials)
- Reaction Time (RT; correct in-tune trials)

Independent Variables

- Temporal Coherence (*high*, *medium*, *low*)
- Musical Expertise (*musician*, *general*)

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Human Perception and Performance

ISSN: 0096-1523

https://doi.org/10.1037/xhp0001103

Does Order Matter? Harmonic Priming Effects for Scrambled Tonal Chord Sequences

David R. W. Sears, Jonathan E. Verbeten, and Hannah M. Percival
Performing Arts Research Lab (PeARL), School of Music, Texas Tech University

This study examines whether scrambling the order of events in a tonal chord sequence inhibits the speed and accuracy of processing in two behavioral harmonic priming experiments. Sixteen 9-chord sequences were adapted from Bach's chorales that either remained unchanged (thereby reflecting high temporal coherence) or were scrambled to produce increasingly incoherent sequences (i.e., medium or low). To produce the scrambled conditions, a finite-context (or *n*-gram) model trained on a corpus of chord annotations and then identified the scrambled versions (scrambling chords 2–8) that produced high estimates of model surprisal. In both experiments, 60 participants (30 musicians) indicated as quickly as possible whether the target chord was in or out of tune, where out-of-tune trials were either fixed at a tuning level of 40 cents sharp relative to the preceding context (Experiment 1), or at a tuning level representing the intonation discrimination threshold of each participant, which was estimated using an adaptive staircasing procedure before the main session began (Experiment 2). Correct response times and sensitivity measures replicated the high-to-low staircase found in the model estimates, suggesting harmonic priming effects reflect the order of chords in a sequence. Implications for topological and temporal models of tonal-harmonic structure are discussed.

Public Significance Statement
This study supports the view that harmonic priming effects reflect the order of chords in a sequence. Thus, listeners with exposure to tonal music may possess schematic knowledge about the statistical dependencies between contiguous events such that a stimulus context primes certain targets over others.

Keywords: musical priming, harmonic expectations, tonal harmony, temporal coherence, order

Stage 2: Journal Article Reporting Standards (JARS)

Method Section – Sears, Verbeten, & Percival (2023)

Research Question

- Does scrambling the order of chords in a tonal harmonic progression slow the speed of processing?

Task

- Speeded Intonation Discrimination (*in*, *out*)

Dependent Variables

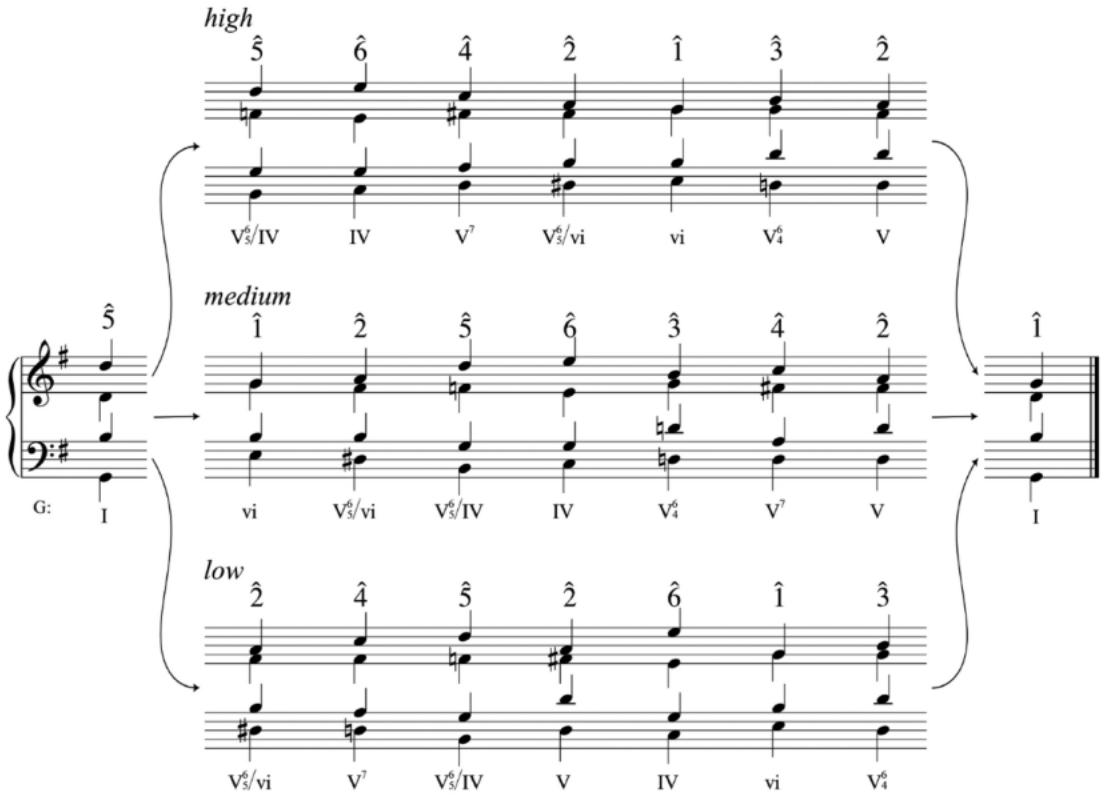
- Discrimination (*d'*; all trials)
- Bias (*c*; all trials)
- Reaction Time (RT; correct in-tune trials)

Independent Variables

- Temporal Coherence (*high*, *medium*, *low*)
- Musical Expertise (*musician*, *general*)

Figure 2

An Example Chord Sequence in Each Coherence Condition



Note. Chord sequences from the medium and low conditions were scrambled using IDyOM. The first and last chords in all stimuli are not scrambled, and the target chord is acoustically identical across all conditions.

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

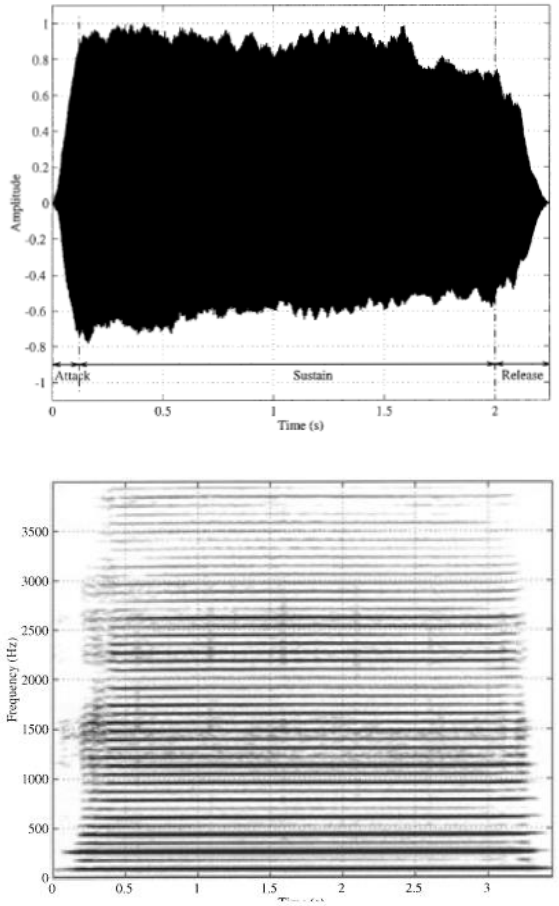
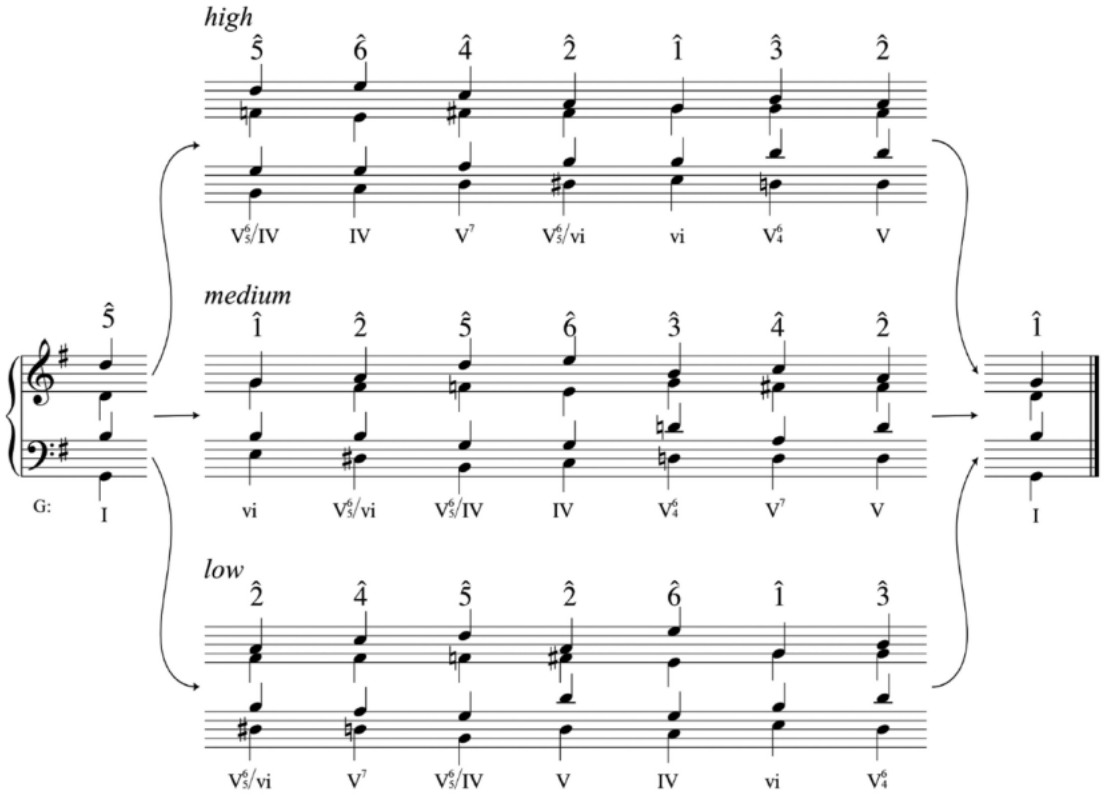


Figure 2

An Example Chord Sequence in Each Coherence Condition



Note. Chord sequences from the medium and low conditions were scrambled using IDyOM. The first and last chords in all stimuli are not scrambled, and the target chord is acoustically identical across all conditions.

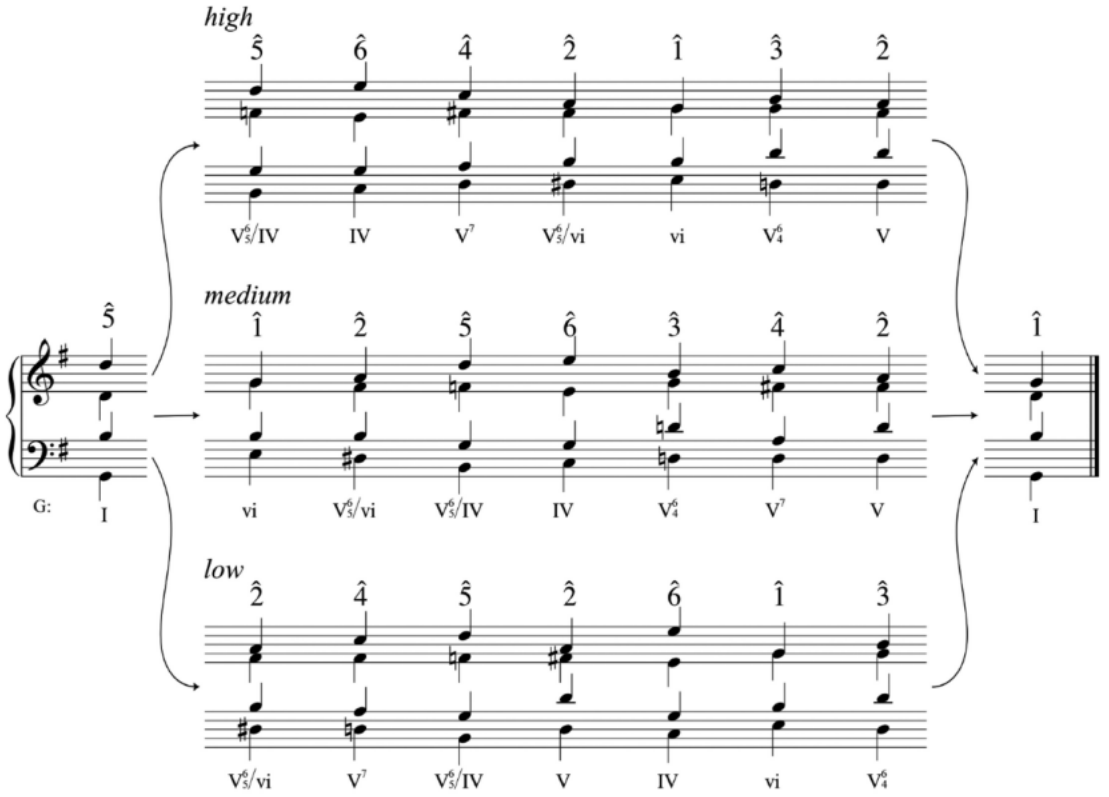
Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Visualizations should serve a purpose!

Figure 2

An Example Chord Sequence in Each Coherence Condition



Note. Chord sequences from the medium and low conditions were scrambled using IDyOM. The first and last chords in all stimuli are not scrambled, and the target chord is acoustically identical across all conditions.

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**

- Sampling & Recruitment – study design (e.g., between-subjects), site selection
- Sample Size – power ($1 - \beta$ in NHST, where β refers to the Type II error)
- Participant Characteristics – interindividual differences (descriptive/inferential)
- Inventories/Instruments – interindividual differences (descriptive/inferential)
- Ethics – ethics approval (e.g., IRB2023-85)

★ **Materials & Design**

- Experimental Conditions – conditions/manipulations

★ **Procedure**

- Data Collection – session walkthrough, task(s), trials and study design

Apparatus

- Equipment – data collection site(s), equipment/tools

Analysis

- Analysis Strategy – measures, inferential statistics, experiment-wise error correction

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Procedure. The experiment was divided into two blocks: block A with the 50 discrete emotion excerpts and block B with the 60 dimensional model excerpts. The order of the blocks was counterbalanced across the participants (67 participants did block A first followed by block B), and the order of the examples within blocks was individually randomized. The participants were instructed to rate the emotions represented by the music excerpts (perceived emotions), and the difference between perceived and induced emotions was explained to them. While one group of participants were rating excerpts in block A on a scale of 1–9 for each discrete emotion, the other group was rating excerpts on bipolar scales of 1–9 for each of the three axes of the dimensional model. Then the blocks were switched for each group. This meant that the experiment used both models to rate the emotions in all 110 excerpts without taking an unreasonably long time. The total duration of the experiment was between 50 and 60 minutes, depending on the participant's rating speed.

– Eerola & Vuoskoski (2011)

Session Walkthrough Task(s) Study Design

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Procedure. The experiment was divided into two blocks: block A with the 50 discrete emotion excerpts and block B with the 60 dimensional model excerpts. The order of the blocks was counterbalanced across the participants (67 participants did block A first followed by block B), and the order of the examples within blocks was individually randomized. The participants were instructed to rate the emotions represented by the music excerpts (perceived emotions), and the difference between perceived and induced emotions was explained to them. While one group of participants were rating excerpts in block A on a scale of 1–9 for each discrete emotion, the other group was rating excerpts on bipolar scales of 1–9 for each of the three axes of the dimensional model. Then the blocks were switched for each group. This meant that the experiment used both models to rate the emotions in all 110 excerpts without taking an unreasonably long time. The total duration of the experiment was between 50 and 60 minutes, depending on the participant's rating speed.

– Eerola & Vuoskoski (2011)

Session Walkthrough Task(s) Study Design

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Participants were also asked to mark how much they liked each example (with a preference rating) and how beautiful they considered each example to be (with a beauty rating). In both cases this was on a scale from 1 to 9. These additional measures were added to clarify the role of valence because valence and preference have been shown to be separate constructs (Schubert, 2007). For example, one can be fond of harsh and rough sounding music despite the fact that most people associate those qualities with negative valence. Ratings of preference and beauty would also briefly allow the exploration of the relation between sadness and valence in this context.

Before the actual experiment, a short practice session was carried out by each participant to become familiar with the interface, likert scales, and type of music used. The participants also had the possibility to ask questions about the task before the start of the experiment.

– Eerola & Vuoskoski (2011)

Session Walkthrough Task(s) Study Design

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**

- Sampling & Recruitment – study design (e.g., between-subjects), site selection
- Sample Size – power (1 – β in NHST, where β refers to the Type II error)
- Participant Characteristics – interindividual differences (descriptive/inferential)
- Inventories/Instruments – interindividual differences (descriptive/inferential)
- Ethics – ethics approval (e.g., IRB2023-85)

★ **Materials & Design**

- Experimental Conditions – conditions/manipulations

★ **Procedure**

- Data Collection – session walkthrough, task(s), trials and study design

Apparatus

- Equipment – data collection site(s), equipment/tools

Analysis

- Analysis Strategy – measures, inferential statistics, experiment-wise error correction

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Apparatus. The listening experiments were conducted in a soundproof room. To gather the emotion ratings, a special patch was designed in *Pure Data* graphical programming environment (Puckette, 1996), running on Mac OS X. The patch enabled the participants to move from one excerpt to the next at their own pace and to repeat an excerpt if needed. Participants listened to the excerpts through studio quality headphones (AKG K141 Studio), and were able to adjust the sound volume according to their own preferences.

– Eerola & Vuoskoski (2011)

Data Collection Site(s) Equipment/Tools

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Apparatus. The listening experiments were conducted in a soundproof room. To gather the emotion ratings, a special patch was designed in *Pure Data* graphical programming environment (Puckette, 1996), running on Mac OS X. The patch enabled the participants to move from one excerpt to the next at their own pace and to repeat an excerpt if needed. Participants listened to the excerpts through studio quality headphones (AKG K141 Studio), and were able to adjust the sound volume according to their own preferences.

– Eerola & Vuoskoski (2011)

Data Collection Site(s) Equipment/Tools

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**

- Sampling & Recruitment – study design (e.g., between-subjects), site selection
- Sample Size – power ($1 - \beta$ in NHST, where β refers to the Type II error)
- Participant Characteristics – interindividual differences (descriptive/inferential)
- Inventories/Instruments – interindividual differences (descriptive/inferential)
- Ethics – ethics approval (e.g., IRB2023-85)

★ **Materials & Design**

- Experimental Conditions – conditions/manipulations

★ **Procedure**

- Data Collection – session walkthrough, task(s), trials and study design

Apparatus

- Equipment – data collection site(s), equipment/tools

Analysis

- Analysis Strategy – measures, inferential statistics, experiment-wise error correction

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Analysis

Discrimination performance on the intonation task (estimated using sensitivity (d') and bias (c) measures from signal detection theory; Kingdom & Prins, 2016) and correct RTs were analyzed with linear mixed-effects models (LMMs; West et al., 2007). All mixed-effects analyses were conducted with the software R (4.0.3) using the package lme4 (Bates et al., 2015). The LMM for correct RTs included crossed random effects for participants and items (musical stimuli; Baayen et al., 2008), along with the maximal random effects structure that would converge (Barr et al., 2013), which was determined using the buildmer package (Voeten, 2020).[2] The maximal model includes random intercepts for each participant and by-participant slopes for the within-participant fixed factors of temporal coherence (*high*, *medium*, and *low*), and intonation (*in*, *out*), and with intercepts for each musical stimulus and by-stimulus slopes for the between-participants factor of musical training (*musicians*, *general*). LMMs for the performance measures did not include a random effect for musical stimuli, however, because these measures aggregate the data across stimuli for each level of the fixed factors.

Fixed factors were sum-coded so that levels of each factor would represent deviations from the grand mean, as is the approach in traditional ANOVA pedagogy (Barr et al., 2013). Omnibus tests and planned comparisons were calculated using the emmeans package (Lenth, 2022). To examine the potential differences between the levels of temporal coherence, we also included planned comparisons between the unscrambled and scrambled conditions (high vs. medium, high vs. low) for both training groups using Bonferroni adjustment. Finally, figures present the means and *SE*s estimated by the LMM, but since correct RTs were log-transformed to address issues of normality in the model residuals, estimated means and *SE*s for that variable were back-transformed to a milliseconds scale for the sake of transparency.

Primary (Omnibus) Tests

Secondary (Comparison) Tests

study design

– Sears, Verbeten, & Percival (2023)

Measures Statistics Error Correction

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

Analysis

Discrimination performance on the intonation task (estimated using sensitivity (d') and bias (c) measures from signal detection theory; Kingdom & Prins, 2016) and correct RTs were analyzed with linear mixed-effects models (LMMs; West et al., 2007). All mixed-effects analyses were conducted with the software R (4.0.3) using the package lme4 (Bates et al., 2015). The LMM for correct RTs included crossed random effects for participants and items (musical stimuli; Baayen et al., 2008), along with the maximal random effects structure that would converge (Barr et al., 2013), which was determined using the buildmer package (Voeten, 2020).[2] The maximal model includes random intercepts for each participant and by-participant slopes for the within-participant fixed factors of temporal coherence (*high*, *medium*, and *low*), and intonation (*in*, *out*), and with intercepts for each musical stimulus and by-stimulus slopes for the between-participants factor of musical training (*musicians*, *general*). LMMs for the performance measures did not include a random effect for musical stimuli, however, because these measures aggregate the data across stimuli for each level of the fixed factors.

Fixed factors were sum-coded so that levels of each factor would represent deviations from the grand mean, as is the approach in traditional ANOVA pedagogy (Barr et al., 2013). Omnibus tests and planned comparisons were calculated using the emmeans package (Lenth, 2022). To examine the potential differences between the levels of temporal coherence, we also included planned comparisons between the unscrambled and scrambled conditions (high vs. medium, high vs. low) for both training groups using Bonferroni adjustment. Finally, figures present the means and *SE*s estimated by the LMM, but since correct RTs were log-transformed to address issues of normality in the model residuals, estimated means and *SE*s for that variable were back-transformed to a milliseconds scale for the sake of transparency.

– Sears, Verbeten, & Percival (2023)

Primary (Omnibus) Tests

Secondary (Comparison) Tests

study design

Measures

Statistics

Error Correction

Stage 2: Journal Article Reporting Standards (JARS)

Results Section

Data Analysis

- Preliminary Analysis – frequencies/percentages, rationale, exclusion methods, reliability
- Model Description – model details, estimation problems (if no Method: Analysis)
- Inferential Statistics (NHST) – test statistics, effect-size estimates, primary/secondary hypotheses
- Exploratory Analyses – continuous IVs for materials (features) or participants (inventories)

Visualizations

- Tables/Visualizations – study design, test statistics

Stage 2: Journal Article Reporting Standards (JARS)

Results Section

Data Analysis

- Preliminary Analysis – frequencies/percentages, rationale, exclusion methods, reliability
- Model Description – model details, estimation problems (if no Method: Analysis)
- Inferential Statistics (NHST) – test statistics, effect-size estimates, primary/secondary hypotheses
- Exploratory Analyses – continuous IVs for materials (features) or participants (inventories)

Visualizations

- Tables/Visualizations – study design, test statistics

Stage 2: Journal Article Reporting Standards (JARS)

Results Section

Results

To rule out extreme mood states that might affect the participants' emotion ratings, POMS-A ratings were aggregated and the distance from the mean rating (1.87, SD = 0.47 on a scale of 1–5) was calculated for each participant. One participant whose score was more than three SDs above the mean was removed from the analysis as that person's current mood appeared considerably pessimistic, tired and negative. The intersubject correlations were used to identify possible outliers and anyone who scored more than three SDs from the mean intersubject correlation was removed from the dataset. This resulted in the removal of five participants, leaving the total of 110 that was eventually used. Cronbach's alpha was used to assess the inter-rater reliability across both experimental blocks. All emotion ratings received alphas above .99 and only the preference ratings were slightly lower (.94), probably since personal opinions are prone to vary across individuals. Subsequently the data was pooled together for further analyses.

We reviewed the contribution of individual variables (personality, musical preferences, film genre preferences, and musical training) to the emotion ratings by correlating the individual ratings with the background variables. Only a few statistically significant correlations emerged: the personality trait 'openness to experience' (John & Srivastava, 1999) appeared related to increased ratings of anger ($r = 0.25$, $p < .05$) and valence ($r = 0.33$, $p < .05$), and 'extroversion' seemed related to decreased ratings of tension ($r = -0.38$, $p < .01$). These traits may therefore

– Eerola & Vuoskoski (2011)

Preliminary Analysis Model Description Inferential Statistics Exploratory Analyses

Stage 2: Journal Article Reporting Standards (JARS)

Results Section

Discrimination of emotion categories and levels. An examination of the discrete emotion rating was carried out next using a mixed design repeated measures ANOVA for groups of excerpts representing each target emotion. The five emotion concepts (anger, fear, sadness, happiness, and tenderness) as the within subjects variable. The two levels (high and moderate) at which excerpts conveyed an emotion (i.e., high and moderate examples) provided the within group variable. Taken together this gave significant main effects for all target emotion concepts, and a significant main effect for levels in one target emotion. There was also significant interaction effects between these two factors in four of the target emotions (see Table 2 for effect sizes and *p* values). In other words, the effect sizes were robust for concept (between 0.79 and 0.83; see Table 2), but negligible for emotion levels (0.000–0.006). And most of the interactions between concept and level were within each target emotion were significant. These analyses were later followed by post hoc tests, in which *p* values were adjusted using the Holm-Sidak procedure to avoid the effects of multiple comparison tests (Ludbrook, 1998). These analyses revealed that in the high examples the target emotion was never confused with other emotion concepts, but the moderate examples exhibited confusions between one or two other concepts of emotion (see Table 2). Figure 2 clearly shows this pattern. For example, moderate examples of anger are indistinguishable from fear and sadness, and moderate examples of tenderness could easily be

– Eerola & Vuoskoski (2011)

Missing/Excluded Data Model Description Inferential Statistics Exploratory Analyses

Stage 2: Journal Article Reporting Standards (JARS)

Results Section

Data Analysis

- Missing/Excluded Data – frequencies/percentages, rationale, exclusion methods
- Model Description – model details, estimation problems
- Inferential Statistics (NHST) – test statistics, effect-size estimates, primary/secondary hypotheses
- Exploratory Analyses – continuous IVs for materials (features) or participants (inventories)

Visualizations

- Tables/Visualizations – study design, test statistics

Stage 2: Journal Article Reporting Standards (JARS)

Results Section

Table 2. Mixed-design repeated measures ANOVA results for groups of excerpts representing each target emotion (η^2 for effect sizes). Post-hoc tests display target emotions that do not reliably differ in means (using Holm-Sidak adjusted values for $p < 0.05$)

Type of excerpt	Category	Level	Interaction	Post-hoc (moderate)
Happy	0.70^{***}	0.003^{**}	0.21^{***}	S, T
Sad	0.83^{***}	0.003	0.05^{*}	T
Tender	0.78^{***}	0.000	0.05	H, S
Fearful	0.79^{***}	0.001	0.08^{**}	A
Angry	0.79^{***}	0.006	0.06^{**}	F, S
Pos. valence	0.52^{***}	0.01	0.08^{*}	En, Te
Neg. valence	0.70^{***}	0.000	0.08^{*}	En, Te
Pos. energy	0.44^{***}	0.11^{*}	0.06	Va, Te
Neg. energy	0.54^{***}	0.01	0.11^{*}	Te
Pos. tension	0.81^{***}	0.02^{**}	0.07^{*}	En
Neg. tension	0.58^{***}	0.09^{*}	0.03	En

Notes: $p < .05$; $^{**}p < .01$; $^{***}p < .001$; $df = 4,49$ for basic emotion concepts, $df = 2,29$ for dimensional concepts.
S = Sadness, T = Tenderness, H = Happiness, A = Anger, F = Fear, Va = Valence, En = Energy, Te = Tension.

– Eerola & Vuoskoski (2011)

Missing/Excluded Data Model Description Inferential Statistics Exploratory Analyses

Stage 2: Journal Article Reporting Standards (JARS)

Results Section

For NHST, report effect sizes!

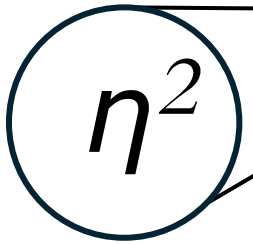


Table 2. Mixed-design repeated measures ANOVA results for groups of excerpts representing each target emotion (η^2 for effect sizes). Post-hoc tests display target emotions that do not reliably differ in means (using Holm-Sidak adjusted values for $p < 0.05$)

Type of excerpt	Category	Level	Interaction	Post-hoc (moderate)
Happy	0.70^{***}	0.003^{**}	0.21^{***}	S, T
Sad	0.83^{***}	0.003	0.05^{*}	T
Tender	0.78^{***}	0.000	0.05	H, S
Fearful	0.79^{***}	0.001	0.08^{**}	A
Angry	0.79^{***}	0.006	0.06^{**}	F, S
Pos. valence	0.52^{***}	0.01	0.08^{*}	En, Te
Neg. valence	0.70^{***}	0.000	0.08^{*}	En, Te
Pos. energy	0.44^{***}	0.11^{*}	0.06	Va, Te
Neg. energy	0.54^{***}	0.01	0.11^{*}	Te
Pos. tension	0.81^{***}	0.02^{**}	0.07^{*}	En
Neg. tension	0.58^{***}	0.09^{*}	0.03	En

Notes: $p < .05$; $^{**}p < .01$; $^{***}p < .001$; $df = 4,49$ for basic emotion concepts, $df = 2,29$ for dimensional concepts.
S = Sadness, T = Tenderness, H = Happiness, A = Anger, F = Fear, Va = Valence, En = Energy, Te = Tension.

– Eerola & Vuoskoski (2011)

Missing/Excluded Data Model Description Inferential Statistics Exploratory Analyses

Outline

- **Stage 2: APA Journal Article Reporting Standards (JARS)**
  - Resources (Quant, REC)
  - APA Manuscript (Layout)
  - Method Section (Participants, Materials, Procedure, etc.)
  - Results Section (Statistics, Visualizations, etc.)

- **Stage 1: Research Ethics and Approval**
  - Drafting Proposals
  - Cross-Cultural Research Guidelines

- **Stages 1-2: New Publication Standards**
  - Supplemental Material (Open Resources)
  - Pre-registrations
  - Registered Reports

Stage 1: Research Ethics & Approval

Drafting Proposals

- Human Research Protection Programs protect the rights and welfare of individuals participating in research.
  - Institutional Review Board (IRB) – USA
  - Research Ethics Board (REB) – Canada
  - Research Ethics Committee (REC) – United Kingdom

- They review proposals for conducting research involving human participants.

- Projects that fall under their purview:
  - Experiments – behavioral, psychophysiological, neural
  - Questionnaires/Surveys/Inventories
  - Interviews – structured and semi-structured

Stage 1: Research Ethics & Approval

Drafting Proposals

- At many institutions, a faculty member must serve as PI and submit the proposal on a student's behalf.
- Data collection at multiple sites typically requires ethics approval for each site.
- **Approval can take up to 6 weeks at some institutions!**

Common Review Types

- *Exempt*
  - Involves minimal risk, with data de-identified (e.g., surveys, interviews, participant observation).
- *Expedited*
  - Involves collection of minimally invasive data (e.g., blood, hearing or vision screenings)
- *Full Board*
  - Involves greater than minimal risk (e.g., single-cell recording).

Stage 1: Research Ethics & Approval

Drafting Proposals

Submission Types

- Initial
- Modification
- Renewal
- Incident
- Withdrawal/Closure

Status

- In-Draft – Research team is writing.
- Awaiting Certification – Awaiting PI review/certification
- Pre-Review – Pre-screening
- Under Review – Sent to a reviewer
- Approved – Reviewed/approved by the review board.

Stage 1: Research Ethics & Approval

Drafting Proposals

Recruitment

- Review Boards often provide templates for common procedures:
  - Recruitment announcements
  - Email
    - Flyer

Common Forms

- Consent (template provided)
- Instructions
- Research materials (e.g., interview questions, survey, etc.)
- Receipt
- Debriefing

Stage 1: Research Ethics & Approval

Drafting Proposals

Common Research-Related Questions

- What is the purpose of the research? (~100 words)
- Enter the research question(s) the study will address.
- Describe the target population to be recruited for this study.
- Describe the experimental session in detail.
- Enter or attach screenshots of the experimental interface.
- Enter or attach the instruments/surveys used for this study.
- Enter or attach all experiment forms (informed consent, instructions, etc.).

Stage 1: Research Ethics & Approval

Cross-Cultural Research Guidelines

Navigating cross-cultural research: methodological and ethical considerations

Tanya Broesch[1,†], Alyssa N. Crittenden[2,†], Bret A. Beheim[3], Aaron D. Blackwell[4], John A. Bunce[3], Heidi Colleran[3,5], Kristin Hagel[3],

Pitfalls with Cross-Cultural Research

- *positivism* – the hypothetico-deductive method produces "authoritative" knowledge.
- *epistemicide* – the erasure of (nonwestern) knowledge systems.
- *extractivism* – taking the knowledge from its place of origin to an intellectual center.

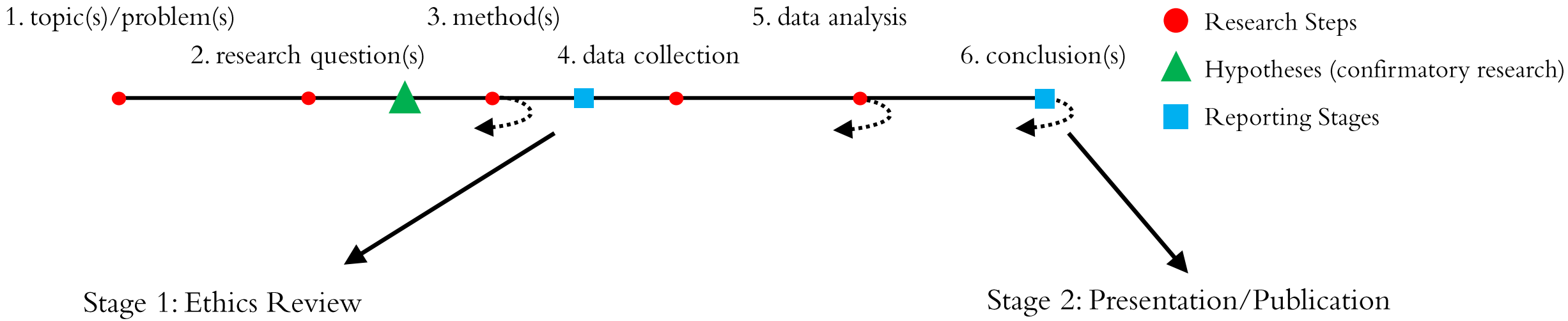
Solutions

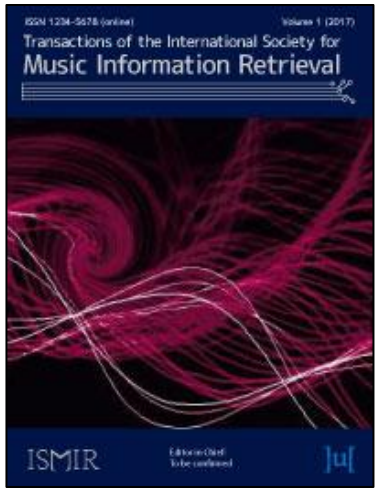
- *study-site selection* – select non-WEIRD populations according to the study's research Qs.
- *community research* – involve cultural experts and adhere to OCAP.
- *experimental design* – prefer full-factorial experiments in cross-cultural research.
- *research methods* – employ culturally appropriate and/or converging methods.
- *positionality* – disclose the researchers' academic backgrounds and personal motivations.

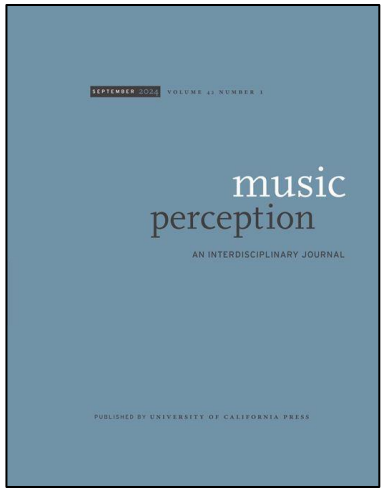
- Review the provided bibliography! (repo)

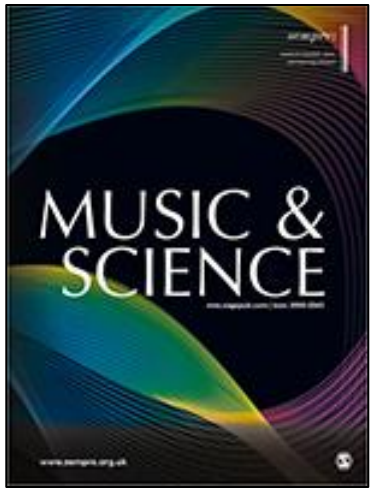
– Broesch et al. (2020); Sauvé et al. (2022)

Research Life Cycle (with Human Subjects)

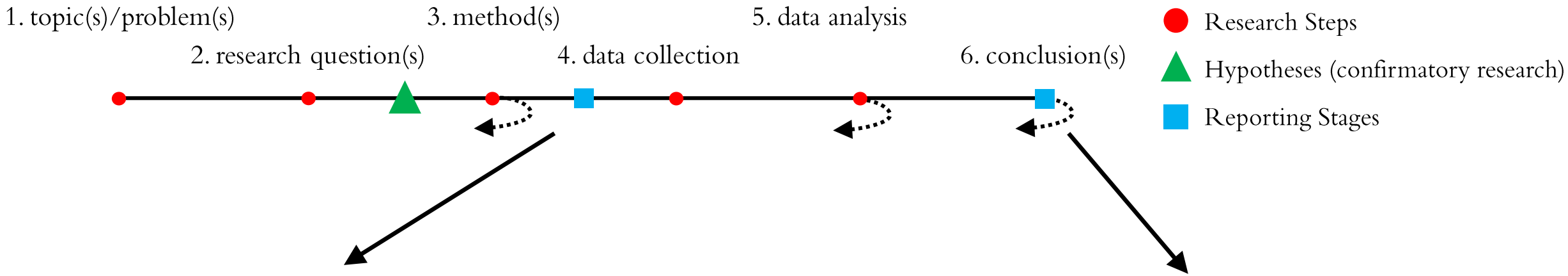




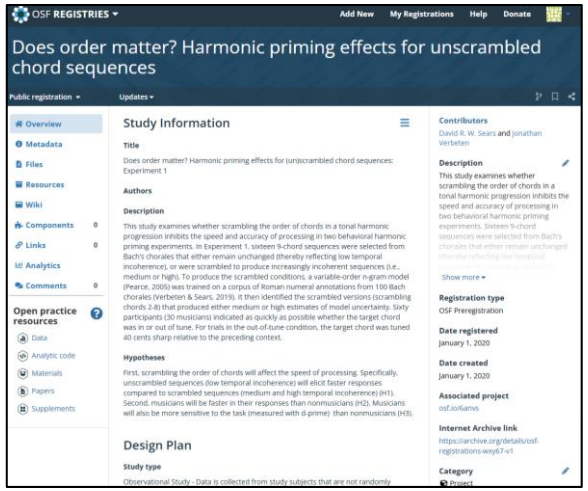




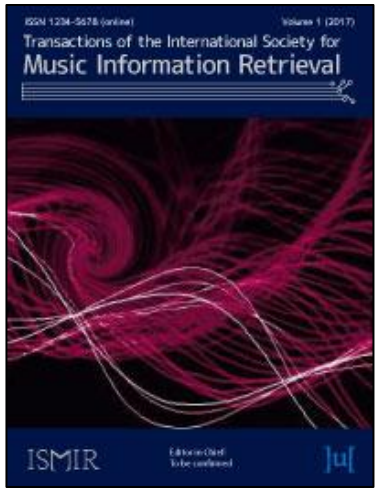
Research Life Cycle (with Human Subjects)

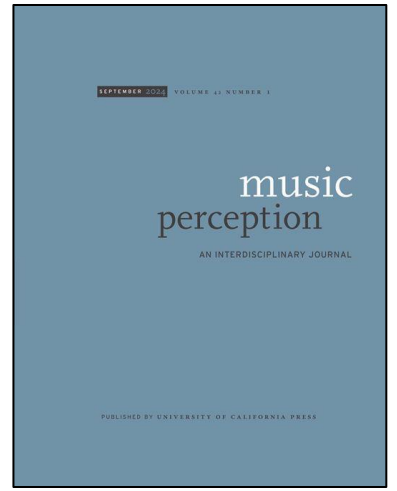


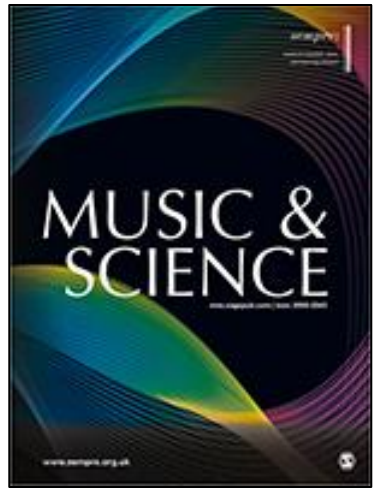
Stage 1: Ethics Review & Pre-Registration



Stage 2: Presentation/Publication







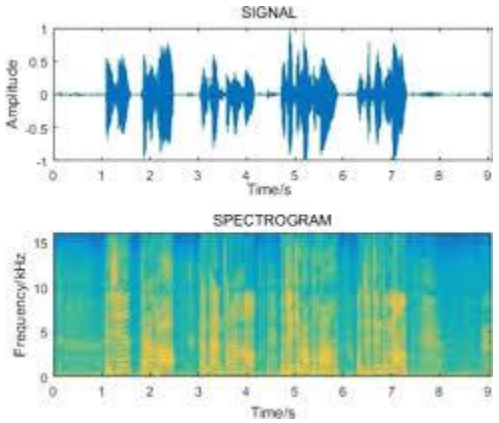
Stages 1-2: New Publication Standards

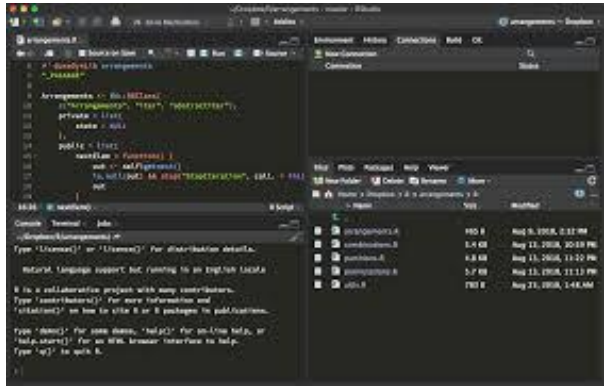
Open Resources

- Funding agencies and academic journals expect/require experimental studies to include an accompanying digital repository of experimental materials.
- Avoids p-hacking and positive-test strategies.
- Improves the quality and transparency of experimental research.

https://osf.io/6anvs/

  - ***Anonymized datasets*** – contingent upon ethics approval
  - ***Stimuli*** – contingent upon adherence to copyright laws (e.g., "fair use")
  - ***Analysis scripts*** – should reproduce the published results exactly





118.1MB Make Private Public 0

Does order matter? Harmonic priming effects for unscrambled chord sequences

30–50% more time/labor!

Contributors: **David R. W. Sears,** Jonathan Verbeten

Date created: 2019-12-29 04:18 PM | Last Updated: 2023-05-15 10:16 AM
Create DOI
Category: Project

Description:

This study examines whether scrambling the order of chords in a tonal harmonic progression inhibits the speed and accuracy of processing in two behavioral harmonic priming experiments. Sixteen 9-chord sequences were selected from Bach's chorales that either remain unchanged (thereby reflecting low temporal incoherence), or were scrambled to produce increasingly incoherent sequences (i.e., medium or high). To produce the scrambled conditions, a variable-order n-gram model (Pearce, 2005) was trained on a corpus of Roman numeral annotations from 100 Bach chorales (Verbeten & Sears, 2019). It then identified the scrambled versions (scrambling chords 2-8) that produced either medium or high estimates of model uncertainty. Sixty participants (30 musicians) indicated as quickly as possible whether the target chord was in or out of tune. In Experiment 1, the target chord was tuned 40 cents sharp relative to the preceding context. In Experiment 2, the optimal level of 'out-of-tuneness' for each participant was selected using an adaptive staircasing procedure in the training phase to ensure task difficulty remained constant across participants.

Wiki

Add important information, links, or images here to describe your project.

Files

Click on a storage provider or drag and drop to upload

Filter

Name	Modified
Does order matter? Harmonic priming effects for unscrambled chor...	
− OSF Storage (United States)	
+ analysis_scripts	
+ datasets	
+ stimuli	

Citation

Components

Add Component Link Projects

Add components to organize your project.

Tags

Add a tag to enhance discoverability

Recent Activity

Stages 1-2: New Publication Standards

Pre-Registrations – Nosek et al., 2018

https://osf.io/wxy67

- a research plan specified in advance of the experiment. Intended for confirmatory research.
- common submission portal: osf.io
- Some journals expect pre-registration; others require it.

Study Information

- Description (abstract) and hypotheses.

Design Plan

- Study design and randomization procedure

Sampling Plan

- Procedures for data collection, sample size, and sample size rationale (i.e., power analysis)

Variables

- Dependent (e.g., 2-AFC), independent (e.g., violin vs. piano), and indices (e.g., d-prime)

Analysis Plan

- Description of statistical models, inference criteria, and exclusion criteria

Does order matter? Harmonic priming effects for unscrambled chord sequences

Public registration Updates

Open practice resources

Data
Analytic code
Materials
Papers
Supplements

Study Information

Title

Does order matter? Harmonic priming effects for (un)scrambled chord sequences: Experiment 1

Authors

Description

This study examines whether scrambling the order of chords in a tonal harmonic progression inhibits the speed and accuracy of processing in two behavioral harmonic priming experiments. In Experiment 1, sixteen 9-chord sequences were selected from Bach's chorales that either remain unchanged (thereby reflecting low temporal incoherence), or were scrambled to produce increasingly incoherent sequences (i.e., medium or high). To produce the scrambled conditions, a variable-order n-gram model (Pearce, 2005) was trained on a corpus of Roman numeral annotations from 100 Bach chorales (Verbeten & Sears, 2019). It then identified the scrambled versions (scrambling chords 2-8) that produced either medium or high estimates of model uncertainty. Sixty participants (30 musicians) indicated as quickly as possible whether the target chord was in or out of tune. For trials in the out-of-tune condition, the target chord was tuned 40 cents sharp relative to the preceding context.

Hypotheses

First, scrambling the order of chords will affect the speed of processing. Specifically, unscrambled sequences (low temporal incoherence) will elicit faster responses compared to scrambled sequences (medium and high temporal incoherence) (H1). Second, musicians will be faster in their responses than nonmusicians (H2). Musicians will also be more sensitive to the task (measured with d-prime) than nonmusicians (H3).

Design Plan

Study type

Observational Study - Data is collected from study subjects that are not randomly

Contributors

David R. W. Sears and Jonathan Verbeten

Description

This study examines whether scrambling the order of chords in a tonal harmonic progression inhibits the speed and accuracy of processing in two behavioral harmonic priming experiments. Sixteen 9-chord sequences were selected from Bach's chorales that either remain unchanged (thereby reflecting low temporal incoherence), or were scrambled to

Show more

Registration type

OSF Preregistration

Date registered

January 1, 2020

Date created

January 1, 2020

Associated project

osf.io/6anvs

Internet Archive link

https://archive.org/details/osf-registrations-wxy67-v1

Category

Project

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**
- Sampling & Recruitment
- Sample Size
- Participant Characteristics
- Inventories/Instruments
- Ethics

★ **Materials & Design**
- Experimental Conditions

★ **Procedure**
- Data Collection

Apparatus
- Equipment

Analysis
- Analysis Strategy

Stage 2: Journal Article Reporting Standards (JARS)

Method Section

★ **Participants**
- Sampling & Recruitment
- Sample Size
- Participant Characteristics
- Inventories/Instruments
- Ethics

★ **Materials & Design**
- Experimental Conditions

★ **Procedure**
- Data Collection

Apparatus
- Equipment

Analysis
- Analysis Strategy

Transparency and Openness

issues of normality in the model residuals, estimated means and SDs for that variable were back-transformed to a milliseconds scale for the sake of transparency.

Transparency and Openness

All data for Experiment 1 were collected in 2019–2020. The experimental design, hypotheses, and analysis plan were preregistered on the Open Science Framework (see https://osf.io/wxy67), and the stimuli, behavioral data, and analysis scripts are available for download at https://osf.io/6anvs/.[3]

Results

Responses preceding the onset of the target chord were deemed

– Sears, Verbeten, & Percival (2023)

Stage 1: New Publication Standards

Registered Reports

- The researchers submit a pre-registered plan for peer review in advance of data collection.
- Research designs that pass peer review receive "in principle acceptance" (IPA).
- Registered Reports are published in two stages!

Stage 1

- consists of the Introduction, Method, Proposed Analyses, and Pilot Data (if applicable).
- Stage 1 manuscripts receiving IPA are published if the authors adhere to the protocol, and their stated conclusions are supported by the collected evidence.

Stage 2

- Consists of the Stage 1 manuscript, but now includes the Analysis, Results, and Discussion.

***Music Perception* publishes registered reports!**

- https://online.ucpress.edu/DocumentLibrary/MP/RR_Guidelines_for_Authors_MP_June_2019.pdf (repo)
- https://online.ucpress.edu/DocumentLibrary/MP/RR_Reviewers_Authors_June_2019.pdf (repo)

Conclusions

- **Journal Article Reporting Standards (JARS)**
  - Provide visualizations for the experimental materials and/or procedure.
  - Statistical analyses should include estimates of effect size!
  - Visualizations and tables should reflect the study's experimental design and analyses.
- **Research Ethics and Approval**
  - Report the site(s) that approved the study and the ID (e.g., IRB2024-64).
  - Review current guidelines for conducting cross-cultural research.
- **New Publication Standards**
  - Include anonymized data, materials, and code in an accompanying repository.
  - Pre-register experiments and/or publish registered reports.
- **Publication Planning**
  - Review the journal's submission guidelines carefully.
  - Include a cover letter with your manuscript submission, and identify reviewers.