MPATC-GE 2042:
Psychology of Music

Development
Empirical Research
Topic choice (2% of project grade)

• Due today, September 26, before midnight


• Students in each group must come up with an ordered list of three preferred topics. The votes should be tallied by adding up the order number for each topic chosen. The topic with the lowest sum will be chosen for the group project.

• Groups should meet to (1) discuss the topic ideas before voting and (2) elect a group leader. The group leader will submit all group assignments on behalf of the rest of the group.
• For the topic choice assignment, groups must add their chosen topic description in the appropriate box on the Google doc: https://docs.google.com/spreadsheets/d/1wsEwq_yYisZC2_VfGbT0wOSM1wB_JVB3djzwM9cdKk4/edit#gid=0

• If there is a tie, a note should be made and both topics should be listed. I will cast the tie-breaking vote and Google doc accordingly. The name of the elected group leader should also be added to the doc.

• Multiple groups may work in the same topic area (though separately). If there is a tie vote for the top choice, I will cast the tie-breaking vote.
Empirical research: Types of studies

• Many of the things music scholars do is empirical: deciphering manuscripts, studying a score, and listening to a performance. That is, much music scholarship is empirical.

• A distinction can be made between *formal* and *informal* observation.

• What makes something Empirical Research is that the observations are planned in advance.

• At least seven types of empirical studies can be distinguished: (1) exploratory study, (2) descriptive study, (3) correlational study, (4) pilot study, (5) experimental study, (6) meta-study, and (7) modeling study.
Types of studies:
1. Exploratory study

- An exploratory study is intended to help the researcher gain familiarity with a new field or phenomenon. When Charles Darwin sailed on the Beagle in 1831-6, the purpose was exploratory. The goal was to see what plants and animals existed in different parts of the world.
- An ethnomusicologist might go to the island of Yap, simply to expose him/herself to the culture and music. A historical musicologist might poke around in a dusty archive, simply to see what is there. A music psychologist might attach a heart-rate monitor to a listener, simply to see if the music has any measurable effect.
- Exploratory studies are not hypothesis-driven. They are common when a researcher enters a new field or encounters a new phenomenon.
- The principal purpose of the exploratory study is to alert the researcher to new possibilities.
- It is usually difficult to get exploratory studies published although this sometimes occurs in the field of ethnomusicology.
Types of studies:
2. Descriptive study

- Like the exploratory study, the descriptive study is not hypothesis-driven.
- It differs from the exploratory study in that usually some measurements are made.
- When an anthropologist discovers the skull of a long dead human ancestor, the first order of business is to describe the skull. This typically involves reporting a series of detailed measurements.
- Publishing a detailed description is useful, even if the anthropologist has no theory or hypothesis to test.
Types of studies:
3. Correlational study

• A correlational study is hypothesis-driven.
• It involves collecting at least two different sets of measurements, and determining whether there is any relationship between the two sets.
• Note that it cannot identify causation.
• A common type of correlational study is the survey.
Types of studies:  
4. Pilot study

• A pilot study is a form of experimental study (next category).

• It is carried out simply as a way of testing the experimental procedure.

• The goal is to determine whether the full-fledged experiment is feasible.

• The difference between a pilot study and an experimental study is typically the number of participants.
Types of studies:
5. Experimental study

- Measure. Hypothesis-driven.
- Can say something about causation. Dependent and independent variables.
- Manipulate independent variables, observe dependent variables.
- Of all the different kinds of empirical studies, the experimental study is the most highly regarded by researchers. There is a reason for this: the experimental study is the only type of study that allows the research to say something about causation.
- “No causation without manipulation.”
Types of studies: 6. Meta study

- A meta-study is a “study of studies.”
- It is typically done when a large number of studies have been carried out related to some problem.
- For example, many studies have been carried out related to whether television violence promotes violent behavior in viewers.
- Some of the studies seem to show a link, whereas other studies seem to show no link.
- In a meta-analysis, the researchers identify all of the pertinent studies. They then evaluate the quality of each study, including the quality of the samples used, the number of participants, the quality of the stimuli, the extensiveness of the controls, and other factors.
- Poor studies are simply discarded if they fail to achieve the minimum quality criteria established by the researchers.
- Then the researchers combine together all of the good studies, and do a statistical analysis on the aggregate data. The aim is to see if all of the studies ultimately tell a coherent story.
Types of studies: 7. Modeling study

• Theories can often be implemented as models.
• An example of a physical model is a large model of San Francisco Bay built by the U.S. Army corps of Engineers. The actual bay is 100 km long.
• The model is 1 km in length so the scale is 1 meter = 100 meters.
• Models are useful for testing hypotheses that are impossible (or unethical) to test in reality.
• For example, how long will it take an oil spill in Oakland to reach the mouth of the Sacramento River?
• More commonly, models are rendered as computer programs.
Empirical research: Six big ideas

1. We never prove anything

- Only logicians and mathematicians can talk about “proof.” Any set of observations can be interpreted in more than one way.

- Conclusions should be expressed as follows:
  - The results *are consistent* with the view that …
  - The observations *indicate* that the theory …

- Avoid words like “establish,” “confirm,” or “prove.” Use words like “suggest” or “imply.”
  - Our study *suggests* that …
  - The results from our experiment *imply* that …

- Just because we can’t prove anything doesn’t mean we aren’t interested in truth. In fact, the pursuit of truth is one of the main motivations for people who engage in research.

- Our slogan reminds us of the motivation, and simultaneously tells us that the truth is not accessible to us:

  **Slogan:** Motivated by truth, with no hope of proof.
Empirical research: Six big ideas
2. Research invites failure

• Suppose you were trying to convince someone that your idea is right. Which of the following strategies provides a more compelling argument?
  – Identify lots of existing evidence that fits with your theory.
  – Identify lots of the ways in which your theory might be wrong—and show that your theory survives all of these challenges.

• Any set of observations is consistent with innumerable theories. So showing that the evidence is consistent with your theory doesn’t make the most compelling argument.

• In formal empiricism, researchers follow the second rhetorical strategy: Instead of trying to “prove” your theory, try to make your theory fail. Your audience will be more impressed.

• It’s not research if you don’t allow failure. Good research seeks opportunities to refute your own ideas.

• Good research does not involve collecting evidence that supports an idea. Instead, good research involves collecting evidence that tests an idea.

• Research is often motivated by our intuitions, hopes, and (sometimes) secret beliefs. Without these motivations, we wouldn’t have the energy to do all the work involved in research.

• Even if you have no ulterior motive, other researchers may think that you have an ulterior motive. You will convince your most skeptical critic by persistently testing your own ideas.

• **Slogan**: It’s not research if you don’t invite failure.
Empirical research: Six big ideas

3. Make a prediction

• What is a test? If a theory is good, then you should be able to make a prediction about the future.

• An empirical test is a prediction. You predict what as-yet unobserved data should look like.
  – Biblical Adage: You can recognize a false prophet by his/her false prophecies.
  – Empirical Test: You can recognize a false theory by its false predictions.

• Slogan: We invite failure by making predictions.
  (We invite failure by testing hypotheses)
Empirical research: Six big ideas

4. Refutation is easier than confirmation

- The statement “All swans are white” can never be confirmed because you could never be sure that you have observed all swans (David Hume, 18th century).
- However, the statement “All swans are white” can fail by observing a single nonwhite swan. Refutation is easier than confirmation (Karl Popper, 1934).
- The essence of modern science: Tempt failure by trying to show that a set of observations is not consistent with your hypothesis. If this test fails, then you can say that “the observations are consistent with the hypothesis.”
- Aim not to be right, but to be not wrong. Or, expressed as our slogan:

  **Slogan**: *Aim not to be right, but to be not not right.*
Empirical research: Six big ideas

5. Abstractions can be tested only by making them concrete

• If an abstract idea correctly describes the world, then we should see evidence of the idea in the concrete organization of the world.

• While theories are abstract ideas, theories can be tested only by making predictions that have observable consequences.

• Predicting things that can’t be observed won’t allow you to test your theory. In order to test an idea, we need to predict things that can be observed.

• Transforming abstract ideas into concrete observations is called operationalizing.

• Slogan: Test hypotheses by operationalizing terms.
Empirical research: Six big ideas

6. Compare actual observations with control observations

- Most people who catch a cold feel better in three or four days. Suppose you catch a cold and take a drug. You feel better in three or four days. Did the drug contribute to your recovery?
- If you always took this drug whenever you caught a cold, how would you ever know whether it was useful or useless?
- When making observations, ask yourself, “What would one normally see without this change or intervention?”
- In order to determine whether a drug helps you recover from a cold, you must compare the effect of taking the drug with what would happen if you didn’t take the drug.
- Each time you get a cold, count the number of days before you recover. Take the drug only every second illness. If the drug is effective for you, then it should, on average, shorten the duration of the colds when you took the drug compared with the colds when you didn’t take the drug. In this research, you will compare the results for the treatment condition against the control condition.
- We learn things only by comparison. All empirical research involves comparing two or more situations, measurements, or conditions.
- **Slogan:** Compare, compare, compare.
Empirical research continued:
Questions, theories, conjectures, and hypotheses

• Questions
  – A question is an expression of inquiry that invites or calls for a reply. Research is typically motivated by a question.
  – “Why do Beethoven’s metronome markings seem so fast?”

• Conjectures
  – A conjecture is a supposition, prediction, or speculative claim made about the world.
  – Conjectures are not directly testable.
  – “The first musical instruments were probably drums.”
Questions, theories, conjectures, and hypotheses (continued)

- **Hypotheses**
  - A *hypothesis* is a testable conjecture. It is a claim or prediction that can, in principle, be compared to existing or future observations.
  - “The music people experience as adolescents will remain the most emotionally important music throughout their lives.”

- **Theories**
  - A *theory* is an explanatory framework for understanding a set of observations. Theories typically propose some *cause*—as in “X causes or influences Y.” Theories often involve words like “because,” “due to,” “affects,” “influences,” or “causes.” Some theories may be untestable. Theories are often used to generate conjectures and hypotheses.
  - “The reason why music from adolescence is so memorable is because of the high levels of the hormone oxytocin experienced during adolescence.”
Group task: Is it a question, theory, conjecture or hypothesis? (or none)

• Why do performers tend to slow down at the ends of phrases?
  – This is a question

• Beethoven’s metronome markings are too fast because his metronome was broken.
  – This is a (small) theory: it proposes a causal explanation. Note the use of the word “because.”

• Much of the popularity of World Music is due to commercial exploitation.
  – This is a theory: it proposes a causal explanation—although the mechanism of influence (“commercial exploitation”) may be a bit vague. Note the use of the phrase “due to.”
Group task: Is it a question, theory, conjecture or hypothesis? (or none)

• How is it that listening to music can sometimes cause shivers to run up and down your spine?
  – This is a question

• The language we use shapes the way we think.
  – This is also a theory. Note that the word “shapes” could be replaced by “affects” or “influences.”

• Brahms uses a lot of hemiolas in his music.
  – This is a hypothesis. It is easily tested.

• Brahms likes hemiolas.
  – Unless Brahms wrote a letter or otherwise communicated that he “likes” (or dislikes) hemiolas, it would be difficult or impossible to determine what he “likes.” Consequently, it is better to call this a conjecture rather than a hypothesis.
Group task: Is it a question, theory, conjecture or hypothesis? (or none)

• Africans have a better sense of rhythm than Europeans.
  – If you think “sense of rhythm” is not possible to measure, then this is a conjecture. Otherwise it is a hypothesis.

• The music of Carl Nielsen echoes the spirit of the Danish people.
  – As written, this statement allows for several different interpretations. It is often helpful to try rewriting a statement in order to gain some clarity. For example, we might rewrite this statement as follows: “The Danish spirit is echoed in the music of Carl Nielson.” This suggests that being Danish influenced Nielson’s music. Formulated this way, the statement would be regarded as a theory.

• Music can make people happier.
  – This is a conjecture or hypothesis.
Group task: Is it a question, theory, conjecture or hypothesis? (or none)

- Raag Shree sounds sad to experienced Indian listeners.
  - This is a hypothesis. In principle, this claim could be easily tested.
- The music of the Lakota has been influential primarily because of frequent portrayals of Plains Indians in Hollywood films.
  - This is a theory: it proposes a causal explanation for the widespread influence of Lakota music. Note the use of the word “because.”
- The purpose of our research is to study the relationship between music and ritual.
  - This is not a question. It is not a theory since no cause is proposed.
  - It could be construed as a conjecture or hypothesis if there were some doubt about the purpose of the research (say, if two collaborators were arguing about what they are doing). e.g.,
    - Researcher 1: The purpose of our research is ... music and ritual.
    - Researcher 2: No, no, the purpose of our research is ... music and dance.
  - But this is a stretch. The best answer is “none.”
Empirical research: Six big ideas

#5: Abstractions can be tested only by making them concrete

- If an abstract idea correctly describes the world, then we should see evidence of the idea in the concrete organization of the world.
- While theories are abstract ideas, theories can be tested only by making predictions that have observable consequences.
- Predicting things that can’t be observed won’t allow you to test your theory. In order to test an idea, we need to predict things that can be observed.
- Transforming abstract ideas into concrete observations is called operationalizing.
Operationalizing

• Operationalize this: Listening to Mozart makes you smarter.
  – Who are the listeners? NYU grad students?
  – What exactly does “Mozart” entail? His piano sonatas?
  – What does “listening” mean? How much? How long? In what situations?
  – What does “smarter” mean? Higher IQ? Better math scores?
Group task: Operationalizing

• Would an interval starting from one scale degree be more memorable if it were starting from a different scale degree?
  – What sort of scale? Atonal, diatonic??
  – What interval?
  – What does “memorable” mean? How is that tested? How long do you have to remember it?
  – Example: Is a tritone more memorable if it starts on a major scale degree other than ^4 or ^7?
  – Is a minor third easier to differentiate from its major counterpart if it begins from ...
  – Task: recognizing that interval from others (e.g., P5, P4...).
  – e.g., putting tritone in different contexts; also tonicization.
  – Is it easier to identify a tritone in an expected context vs. unexpected (where there is a tonal violation/not in key)?
Group task: Operationalizing

• How do melodic contours in speech differ based on emotional sentiment?
  – What kind of sentiment?
  – Which languages?
  – Details about the melodic intervals
  – Felt vs. perceived emotion
Group task: Operationalizing

• How much uncertainty in melodic phrases do listeners find most pleasing?
  – Quantifying levels of uncertainty
  – What kind of melodic phrases? How long are they? What style?
Group task: Operationalizing

• How does music affect shoppers in a retail environment?
  – What kind of music?
  – What time of year?
  – What type of retail environment?
  – What demographic?
  – What does “affect” mean?
Group task: Operationalizing

- How does harmonic structure in R&B music contribute to listener preference?
  - What does harmonic structure mean? Level of harmonic variety?
  - Too much change vs. less change?
  - How many times does it modulate?
  - What kind of listeners? People who typically listen to R&B?
  - What is R&B?
  - Preference: depends on mood of subject? So more than one rating over time? Testing for consistency?
  - “Preference” in what context?
  - What does the stimuli sound like?
Group task: Operationalizing

• How is syntax in music different than syntax in language?
  – What is the definition of syntax in music/language?
  – This is very broad question – needs to be more specific.
  – Make question more meaningful: how are they similar rather than different.
Group task: Operationalizing

- What kind of information can be best conveyed through music, and what musical and acoustic parameters are best for conveying that information?
Group task: Operationalizing

• How does perception visual scene differ in musical vs. non-musical environments?
Group task: Operationalizing

• How does noise in urban environments affect us?
Thompson, Chapter 5: Music Acquisition

• The effects of music on a human fetus
• Active and passive forms of learning and their role in musical development
• Studies for evaluating the sensitivity of preverbal infants to various attributes of music
• Musical attributes for which infants show early sensitivity vs. musical attributes for which sensitivity emerges later in development
• Two ways in which early sensitivity to consonance and dissonance is manifested
• Development of sensitivity to harmony and key
Article: Hannon & Trehub, 2005

- Discussion leaders: Tyler Nabinger and Willie Payne
The authors increased MIDI note amplitude on the metrically strong beats to indicate the pulse. Does this bias the experiment in favor of people who use that indicator to perceive the meter? If the metrically strong beats were not accented, would the results have been similar? (Sri)
The author says that the infant testing method includes introducing the monitors with a red light and then playing long clips. Is this a decent testing method for the attention span of an infant? How can the test rule out the possibility that the parents had some influence subconsciously on the infants’ test results? (Shannon)

In experiment 3, there was a heavy reliance on ending the test as soon as an infant looked away for more than 2 seconds. While it could be a good way to judge an infant’s attention to visual stimuli, is it really an effective way of testing an infant’s perception of auditory stimuli? (Mihir)
• On their article, Hannon and Trehub describe a 3 part experiment that involved a group of North American college students who have lived exclusively in North America, a group of college students that were Bulgarian or Macedonian descendants and a group of 6 months old infants. The results on the first two groups were obtained by contrasting an comparing a sample audio with two altered versions, while on the third group, the music stimuli was accompanied by a visual element. Since the music sample used in the experiment pertained to Bulgarian folklore, the results clearly reflected the familiarity with the style of music used, not only in a metrical aspect, but instrumental, and melodic as well. What would have happened if the music sample used as a reference belonged to a style more familiar to the North American students or if the sample consisted on just a rhythm played on a piano or percussion? If the Bulgarian students had only lived in Bulgaria, would North American students show better, same or worse results in differentiating the samples? How accurate do you consider this type of perceptual responses to be, since, with such a wide range of possible answers (1 to 6), doesn’t it become a subjective experience? (Julian T.)
Reading questions: Too many experiments?

• Did this experiment tackle too many topics at once, and did it go in depth enough into each individual study? Is it possible that the whole experiment could have been divided to fit several separate research papers? (Shannon)
Reading questions: Relearning?

- The study results suggest that biases in temporal pattern processing are a result of musical enculturation, and that infants are born without a preference to simple or complex metric measures. If the brain reorganizes how it processes these details with added exposure to unique musical characteristics, is it possible to reorganize or "retrain" those brain processes so that there is a decreased bias to any specific meter? (Eugenio)
Reading questions: Preference for complex meters possible?

- Part of the study focused on Romanians and Bulgarians that have been exposed to both simple and complex metrical structures. However, would there be a bias towards complex metrical structures from people who have been minimally exposed or not exposed at all to Western music? (Eugenio)

- If one were to perform an identical experiment with adults who have only really had experience with complex meters, would they have as much trouble noticing changes in simple meters as North American adults have in noticing changes in complex meters? Essentially, are simple meters actually "simple" or are they as confusing to those who don't have experience with them as complex meters are to North Americans? (Max)
Reading questions: Stimuli

• By using the same familiarization melodies for both adult groups (from traditional Serbian & Bulgarian melodies), is it possible the results are confounded by a deep familiarity with the melodies prior to the experiment for one group and not the other?

Thoughts:

The authors say that the same procedure and familiarization/test stimuli were used for the North American Adults and the adults of Bulgarian and Macedonian origin. If the familiarization stimuli (from which the test stimuli were derived) were 'traditional folk-dance melodies of Serbia and Bulgaria,' I think it is likely that the second test group (Bulgarian/Macedonian) would be much more likely to have these melodies already engrained in their music culture. It is also probable that some of the participants in this group may experience playing and/or dancing to these songs throughout their life, which would make them much more familiar with the melody, structure, and meter of the stimuli than their North American counterparts. Even though the North Americans did have a familiarization period in the study, this would not be equal to potentially growing up with a song as is likely for the Bulgarian/Macedonian participants. Thus, I think it is quite possible that the second group could spot metrical violations due to a much greater familiarity with the specific stimuli.

• A better way to design this experiment would be to create completely novel melodies that both adult groups have never heard before in both metrical conditions. This would eliminate the possibility of familiarity with the specific melodies used affecting the answers. (Henry)
Reading questions: Rating type

• Is the use of a scale measurement (1, *very well*, to 6, *very poorly*) the best way to test whether subjects are sensing the meter violations? Wouldn't a binary measure work better/be more consistent with the stimuli used in test?

• Thoughts: In general, I think gradual scales with many options (1-6 in this case) are good measures when there is a gradual change in stimuli. If the test stimuli for meter violations in this experiment had varying in how much they violated the meter, this may be effective. However, it seems that the researchers only created 1 difficulty level for the violations condition (adding 1 extra note without changing the previous or following note durations). To me, it seems like using a gradual scale would be confusing for participants. What would be the difference between a 4 and a 6 on this scale when there is only 1 type of violation? They were only really testing rhythmic deviations that fit the meter perfectly (which would be 1 on the scale) and rhythmic deviations that do not fit the meter (which would be 6?). I think adding 6 options for answers would only make the participants feel as if they needed to vary their answers and could possibly just add more variation to the results that does not actually reflect differences in the processing of the meter. In addition, since each participant has different criteria for how to use the scale (and may adjust their criteria while the experiment is progressing) it is hard to compare one person's 4 to another person's 6. It is even hard to compare 1 person's 4 to the same person's 6 a few trials later.

(Henry)
According to Hannon, E. E., & Trehub, S. E. (2005), North American infants perceived alterations in metric structure more readily than North American Adults. Similarly, Trainor, L. J., & Trehub, S. E. (1992) demonstrated how infants perceived changes in both non-diatonic and diatonic melodies equally well, while their adult counterparts were able to detect changes in the non-diatonic tests but had difficulty with diatonic changes. It follows that the process of enculturation of adults in our North American society limits our perceptions of musical properties outside of the dominant mainstream tonal and rhythmical context. Are there any aspects of music perceptions in which adults fare better than infants or are we doomed to forever be bested by 7-month olds? (Johnny)
Reading questions: Unhappy babies

* In experiment 3, is it possible that a factor like hunger could have an effect on the results? (Julian C.)
Reading questions
Reading questions
Reading questions
Reading questions
Reading questions
Reading questions
Reading questions