# Measuring Stereotypes in Music: A Commentary on Susino and Schubert (2019)

MANUEL ANGLADA-TORT [1]
*Audio Communication Group, Technische Universität Berlin*

ABSTRACT: In this commentary, I first discuss the strengths of the target paper and provide suggestions for future research. I proceed to point out an important limitation of the target study as well as contribute considerations relevant to measuring stereotypes in music. Finally, I present a novel theoretical account to explain music stereotyping, namely, the representativeness heuristic (Tversky & Kahneman, 1974), which I discuss within the broader framework of the behavioral economics of music.

## STRENGTHS AND SUGGESTIONS

MUSIC stereotyping can play an influential role in a number of music-related phenomena, including musical judgements and preferences, emotion perception, music education, and music and identity (Dunbar, Kubrin, & Scurich, 2016; Negut & Sârbescu, 2014; Lonsdale & North, 2011; Rentfrow & Gosling, 2007; Susino & Schubert, 2017, 2018). The target paper makes an important contribution by investigating negative emotion stereotyping of music genres.

Susino and Schubert (2019) allocated participants (238 undergraduate students) randomly to two groups. Both groups of participants listened to four music excerpts: two test stimuli (either excerpts of heavy metal or hip hop music, depending on the group) and two control stimuli (excerpts of pop music). The lyrical content was identical in both test and control conditions, with only the music changing across conditions. Participants' main task was to indicate which emotion they thought the music was expressing. The results showed a clear effect of music genre on emotion perception, suggesting that heavy metal and hip hop were perceived as expressing more negative emotions than pop music.

In addition to measuring perceived emotions, it could have been interesting to examine the influence of music stereotyping on other evaluative dimensions, such as pleasure or liking, perceived quality, and more behavioral aspects (e.g., the likelihood to recommend the song to a friend or attend a concert by the artist). Moreover, using a similar paradigm, future researchers could compare participants with extremely different music preferences (e.g., hip hop vs. heavy metal vs. pop music fans). Such comparison could help us better understand the relationship between music preferences, stereotypes, and emotion perception in music.

A potentially problematic aspect of the design used by Susino and Schubert (2019) is that lyrics in some music stimuli were more difficult to understand than in others. However, the authors came up with a very practical way of measuring the extent to which lyrics were understood. At the end of the experiment, participants were presented with an unexpected memory task in which they had to indicate which words were featured in the lyrics of the music excerpts. The list of words included target words (present in the lyrics) and foil words (not present). The results revealed that participants discriminated the lyrics above chance level, indicating that the emotional responses were not due to misunderstanding the lyrics.

Another strength of the target paper is the careful control over the emotional content of the song lyrics. By using the *Linguistic Analysis Word Count* software (LIWC; Pennebaker, Francis, & Both, 2007) and *iFeel* system (Araüjo, Gonçales, Cha, & Benevenuto, 2014), the authors determined that the emotionality of the lyrics was overall positive. This computerized approach to analyze the emotional content of lyrics has several advantages: It does not rely on human coders, who are prone to bias, it is not limited by the quantity of lyrics that can be analyzed, and it produces reliable and generalizable results.

Furthermore, researchers interested in measuring perceived emotions in music in a spontaneous manner (instead of restricting participants to a pre-made list of emotions), will find the detailed procedure used to collect and analyze this type of data very useful.

Finally, I would also like to highlight the use of the *Affective Norms for English Words* dataset (ANEW, Bradley and Lang, 1999) to study emotion perception in music. This resource provides normative emotional ratings for more than a thousand words in the English language, including ratings of valence, arousal, and dominance (Bradley & Lang, 1999). In the target paper, the authors used this dataset to quantify the boundaries within the valence-arousal space, established by the affective ratings of the words that were used to measure perceived emotion (e.g., anger, happy, sad, relaxed). The ANEW dataset, however, has many other potential applications to music psychology research. In a recent paper, I used the ANEW dataset to carefully manipulate the emotional content of song titles, creating positive (e.g., "kiss"), negative (e.g., "suicide"), and neutral (e.g., "window") titles (Anglada-Tort, Steffens, & Müllensiefen, 2018). Researchers interested in the control and manipulation of linguistic characteristics of words for music research might also find *NIM* useful, a free search engine designed to provide psycholinguistic research materials, such as word frequency, length, lexical neighbors, and orthographic similarity (Guasch, Boada, Ferré, & Sánchez-Casas, 2013).

## MEASURING STEREOTYPES

I have a major concern regarding the target paper, namely, the effectiveness of the experimental paradigm used to measure the impact of genre-specific stereotypes on music perception and evaluation. The musical input between genre conditions was different and, therefore, it is unclear whether the findings show "negative emotion stereotyping", as claimed by the authors, or merely a general music preference for pop music over heavy metal and hip hop. In fact, the overall fandom scores given by participants seem to support the latter: on average, participants had a lower preference for heavy metal ($M = 1.93$, $SD = 1.08$) and hip hop ($M = 3.32$, $SD = 1.15$) than for pop music ($M = 4.14$, $SD = .94$), which resulted in the highest fandom scores on a scale from 1 (non-fan/never listen to it) to 5 (fan/listen to it all the time). Thus, it is likely that this higher preference for pop music resulted in more positive emotional responses when listening to pop music compared to pieces from less preferred genres.

In addition to generic music preferences, it is plausible that participants simply preferred the specific recordings used in the pop music condition compared to those used in the heavy metal and hip hop conditions. The music stimuli differed in many crucial aspects other than genre, including, *inter alia*, the familiarity of the song and artist, year of release, production and recording quality, instrumentation, and tempo. For example, the first music excerpt in the pop music condition (i.e., *Lay Lady Lay*) was by Bob Dylan, who also composed the song and released it in 1969. The matched song with identical lyrics is a version of Bob Dylan's song performed by the heavy metal band Ministry, released almost 30 years later. Despite having identical lyrics, these two recordings are remarkably different in their popularity, performing artist, and musical content. Arguably, the emotional content of the two songs is also different. Thus, based on the data from the target paper, it is not possible to disentangle whether participants' judgments were based on musical preferences or stereotyping. In this regard, future research should carefully control for the characteristics of the music stimuli, such as familiarity, liking, and emotional content.

To measure music stereotypes successfully using a similar design to Susino and Schubert (2019), I would strongly recommend that the two objects under evaluation (i.e., two music pieces) are identical across conditions. For example, one could present the same musical piece with different explicit information about the genre. The only difference between stimuli should be this piece of information (e.g., a label indicating the genre of the music). In this manner, the effect of non-musical factors (e.g., stereotypes) can be measured and sufficiently isolated because the explicit information presented with the music is the only information that changes while the music remains the same. A similar approach was used by North and Hargreaves (2005), who presented identical music pieces labelled either as "suicide-inducing" or "life-affirming". In both situations, the music pieces were perceived in line with how they were framed.

However, measuring the impact of genre-specific stereotypes while using identical music pieces has an additional difficulty: the music genre of a piece can be easily identified simply by listening to it. One could overcome this challenge by carefully selecting pieces of music that are ambiguous in terms of their genre (e.g., pieces that cannot be identified as belonging to a specific music genre). Margulis, Levine, Simchy-Gross, and Kroger (2017) used a similar approach when investigating the effects of explicit

information on music perception and appreciation. Participants listened to emotionally ambiguous stimuli (i.e., music that could be perceived either as positive or negative) presented either with positive, negative, or neutral information about the composer's intent. The results showed that ambiguous music was evaluated as happier when presented with positive information and as sadder when presented with negative information.

On a final note, it is worth noting that the topic of measuring stereotypes has garnered much attention in social psychology (see Nelson, 2009, for a review). Defining stereotypes is a problematic task itself, with many different definitions. In the context of the target paper and the topic of "problem music" in general (see North and Hargreaves, 2006, for a review), it could also be useful to consider the concept of prejudice, "a negative attitude toward a group or toward members of the group" (Nelson, 2009, p.2). Furthermore, a distinction has to be made between the type of measurement that is used: direct and obtrusive self-report measurements (e.g., Likert scales or trait check-offs) versus indirect and unobtrusive behavioral measurements (e.g., sitting distance or implicit reaction times) (Nelson, 2009). When measuring music stereotyping, researchers have used several self-report measurements, such as agreement scales (Dunbar et al., 2016), the lyrics evaluation scale (Negut & Sârbescu, 2014), and reported emotions (Susino & Schubert, 2018, 2019). However, to the best of my knowledge, there is a lack of research using indirect and unobtrusive measurements.

## THE REPRESENTATIVENESS HEURISTIC

Susino and Schubert (2019) discussed their findings within the *Stereotype Theory of Emotion in Music* (STEM; Susino & Schubert, 2017, 2018), adding a relevant theoretical contribution to the paper. To conclude this commentary, I would like to present an alternative theoretical account for music stereotyping, namely, the *Representativeness Heuristic* (Tversky & Kahneman, 1974). This theory is not exclusive to music, applying to any stereotypical judgement in other domains, and connecting with the broader fields of behavioral economics and the psychology of decision making (see Angner, 2016; Cartwright, 2014; Hastie & Dawes, 2010; Kahneman, 2011; Thaler, 2015, for reviews).

When making judgements and decisions, people are often faced with uncertainty, such as when evaluating who your favorite artist is, or which emotion is expressed by a particular song. In these situations, people rely on mental shortcuts, or heuristics. Although, most of the time, heuristics are economically viable (e.g., speeding up the decision-making process), in some situations they fail in a predictable and systematic manner that can lead to bias, such as stereotypical judgements (e.g., Hastie & Dawes, 2010; Kahneman, 2011).

One of these mental shortcuts is the representativeness heuristic, which refers to the human tendency to estimate the likelihood of an event by comparing it to an existing event of similar characteristics that already exists in people's minds (Tversky & Kahneman, 1974). In other words, with prolonged cultural exposure, people create different categories of stereotypes (e.g., heavy metal expresses anger, fear, and disgust). When they are then faced with a new object (e.g., a given musical piece), people judge the probability that the object in question belongs to the stereotypical category based on the extent to which the object resembles (i.e., is representative of) the category stereotype (Kahneman & Frederick, 2002). This occurs in an automatic, fast, and unconscious manner. Thus, people are often unaware of their own cognitive biases, making judgmental heuristics very difficult to confront and overcome.

My insight here is straightforward: Like any other human judgements, evaluations of music also rely on heuristic principles, such as the representativeness heuristic. I would therefore like to propose the representativeness heuristic as a mechanism underlying music stereotyping. In fact, Lonsdale and North (2011) found empirical evidence supporting the existence of the representativeness heuristic when judging other people's musical taste. In a first experiment, participants evaluated the likely musical taste of 10 fictional individuals, which were described according to stereotypes associated with fans of 10 different music genres (e.g., classical music, heavy metal, rap, chart pop). A significant number of participants were able to identify the particular musical genre as the likely favorite for each of the 10 fictional individuals. In a second experiment, Lonsdale and North (2011) found that participants' predictions of an individual's likely musical taste were significantly correlated with perceived similarity to stereotypical categories of music fans. However, participants' predictions were not correlated with base-rate estimates of general musical taste (i.e., the estimation of the distribution of music genre preferences in the British population). This condition, wherein predictions of likelihood correlate more closely with evaluations of similarity than with base-rate estimates, is crucial to support the existence of the representativeness heuristic (Kahneman

& Tversky, 1973). Nevertheless, the extent to which the representativeness heuristic applies to music evaluation, such as aesthetic and emotional responses to music, remains unclear to this day.

In the context of the target paper, music stereotyping could be explained by the representativeness heuristic in the following way: By being exposed to Western culture, people learn that heavy metal and hip hop music are two representative genres of "problem music", which are normally associated with negative emotions (e.g., anger, fear, or disgust). When listening to a piece of music that is representative of heavy metal or hip hop, listeners search for familiar and stereotypical categories, substituting the attributes of the latter with the former. However, to support the existence of the representativeness heuristic as a mechanism underlying music stereotyping, one should also investigate the correlation between base-rate estimates, judgements of similarity, and probability (Kahneman & Tversky, 1973).

A potential experimental design to test this idea, based on Kahneman and Tversky's (1973) original study and Lonsdale and North (2011), could involve three different groups: (1) the base-rate group, (2) the similarity group, and (3) the probability group. In the base-rate group, participants would be asked to consider the distribution of music genres in the singles sales chart top 100 in the UK today. They would be required to give their best guesses about the percentage of the music that belongs to the following 10 genres: pop, rock, dance, hip hop, R&B, classical, country, jazz, heavy metal, and reggae. Participants in the similarity and probability groups would listen to the same piece of music, previously selected from the actual single sales chart top 100 of a specific music genre (e.g., country). In the similarity group, participants would be asked to rank each of the 10 music genres in order of how similar the music excerpt is to the typical song of that music genre. In the probability group, participants would be asked to rank each of the 10 music genres in order of how likely it is that the music excerpt is charted on the singles sales chart top 100 of that music genre.

To support the existence of the representativeness heuristic in music stereotyping, participants' predictions should be strongly correlated with mean similarity rankings rather than with base-rate estimates. This finding would suggest that listeners do not consider the actual distribution of music genres in a particular culture, which is a complex and time-consuming calculation. Instead, listeners only consider the similarity (or representativeness) to the stereotype category of each music genre. This could also explain the findings from Susino and Schubert (2019) and negative stereotyping of "problem music" in general. But the representativeness heuristic is likely to underlie any decision situation where people try to predict specific probabilities in music. Thus, understanding this heuristic could be central to other music phenomena, such as hit song science or the use of music in advertising. An important part of professionals' job in these areas is to predict the probability of success for songs before they are released to the market or featured in an advertising campaign.

The representativeness heuristic is just one of many decision rules within the heuristic-and-biases framework (Tversky & Kahneman, 1974). This research framework has been hugely influential for understanding human judgements and decision making and is one of the fundamental theories underlying behavioral economics (see Hastie & Dawes, 2010; Kahneman, 2011, for reviews). Somewhat surprisingly, however, the heuristics-and-biases framework has not yet been applied explicitly to music. The scientific potential of this research framework is immense. For example, *The Decision Lab* recognizes more than 80 cognitive biases and heuristics that affect human judgements and decision making (https://thedecisionlab.com/bias/), all supported by empirical findings.

Therefore, *The Behavioural Economics of Music* (Anglada-Tort, 2018; Anglada-Tort & Müllensiefen, 2017; Anglada-Tort et al., 2018; Anglada-Tort, Thueringer, & Omigie, 2019) aims to create a solid understanding of the role that behavioral economics can play in the study of human behaviors related to music. The heuristic-and-biases framework is only one of many areas in behavioral economics that could be useful to music research. Others include time preferences, dual-process theories, nudge theory, or behavioral pricing. In this commentary, I hope to show the potential of the behavioral economics of music and encourage future researchers to apply this research program when investigating issues related to music, such as stereotypical judgements.

## ACKNOWLEDGEMENTS

**NOTES**

[1] Correspondence concerning this commentary should be addressed to Manuel Anglada-Tort, Department of Audio Communication, Technische Universität Berlin, Berlin, Germany. E-mail: m.anglada.tort@campus.tu-berlin.de

**REFERENCES**

Anglada-Tort, M. (2018). Commentary on Canonne (2018): Listening to Improvisation. *Empirical Musicology Review*, *13*(1/2), 16-20. https://doi.org/10.18061/emr.v13i1-2.6387

Anglada-Tort, M., & Müllensiefen, D. (2017). The repeated recording illusion: The effects of extrinsic and individual difference factors on musical judgements. *Music Perception, 35*(1), 92-115. https://doi.org/10.1525/mp.2017.35.1.94

Anglada-Tort, M., Steffens, J., & Müllensiefen, D. (2018). Names and titles matter: The impact of linguistic fluency and the affect heuristic on aesthetic and value judgements of music. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000172

Anglada-Tort, M., Thueringer, H., & Omigie, D. (2019). The busking experiment: A field study measuring behavioral responses to street music performances. *Psychomusicology: Music, Mind, and Brain*, *29*(1), 46. https://doi.org/10.1037/pmu0000236

Angner, E. (2016). *A course in behavioral economics* (2nd edition). New York, NY: Palgrave Macmillan.

Araújo, M., Gonçalves, P., Cha, M., & Benevenuto, F. (2014, April). *iFeel: a system that compares and combines sentiment analysis methods.* Paper presented at the Proceedings of the 23rd International Conference on World Wide Web (pp. 75-78), Seoul, Korea. https://doi.org/10.1145/2567948.2577013

Bradley, M. M., & Lang, P. P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Technical Report C-1). Gainesville, FL: University of Florida, The Center for Research in Psychophysiology.

Cartwright, E. (2014). *Behavioral economics*. New York, NY: Routledge. https://doi.org/10.4324/9780203816868

Dunbar, A., Kubrin, C. E., & Scurich, N. (2016). The threatening nature of "rap" music. *Psychology, Public Policy, and Law, 22*(3), 280-292. https://doi.org/10.1037/law0000093

Guasch, M., Boada, R., Ferre, P., & Sanchez-Casas, R. (2013). NIM: A web-based swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, *45*, 765–771. https://doi.org/10.3758/s13428-012-0296-8

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgement and decision making*. Thousand Oaks, CA: SAGE Publications.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive thought* (pp. 49–81). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237. https://doi.org/10.1037/h0034747

Lonsdale, A. J., & North, A. C. (2012). Musical taste and the representativeness heuristic. *Psychology of Music*, *40*(2), 131-142. https://doi.org/10.1177/0305735611425901

Margulis, E. H., Levine, W. H., Simchy-Gross, R., & Kroger, C. (2017). Expressive intent, ambiguity, and aesthetic experiences of music and poetry. *PLOS ONE*, *12*(7), e0179145. https://doi.org/10.1371/journal.pone.0179145

Neguţ, A., & Sârbescu, P. (2014). Problem music or problem stereotypes? The dynamics of stereotype activation in rock and hip-hop music. *Musicae Scientiae, 18*(1), 3-16. https://doi.org/10.1177/1029864913499180

Nelson, T. D. (2009). *Handbook of prejudice, stereotyping, and discrimination*. New York, NY: Psychology Press. https://doi.org/10.4324/9781841697772

North, A. C., & Hargreaves, D. J. (2005). Brief report: Labelling effects on the perceived deleterious consequences of pop music listening. *Journal of Adolescence*, *28*(3), 433-440. https://doi.org/10.1016/j.adolescence.2004.09.003

North, A., & Hargreaves, D. (2008). *The social and applied psychology of music*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198567424.001.0001

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2007). Linguistic inquiry and word count: LIWC [computer software]. Austin, TX. Retrieved from http://LIWC.net

Rentfrow, P. J., & Gosling, S. D. (2007). The content and validity of music-genre stereotypes among college students. *Psychology of Music*, *35*(2), 306-326. https://doi.org/10.1177/0305735607070382

Susino, M., & Schubert, E. (2017). Cross-cultural anger communication in music: A framework towards a stereotype theory of emotion in music. *Musicae Scientiae, 21*(1), 60-74. https://doi.org/10.1177/1029864916637641

Susino, M., & Schubert, E. (2018). Cultural stereotyping of emotional responses to music genre. *Psychology of Music, 47*(3), 342-357. https://doi.org/10.1177/0305735618755886

Susino, M., & Schubert, E. (2019). Negative emotion responses to heavy-metal and hip-hop music with positive lyrics. *Empirical Musicology Review, 14*(1-2), 2-15. https://doi.org/10.18061/emr.v14i1-2.6376

Thaler, R. H., & Ganser, L. J. (2015). *Misbehaving: The making of behavioral economics*. New York, NY: WW Norton.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124