

Commentary on “The Effects of Musical Fit On Choice Between Competing Pairs of Cultural Products” by Joanne P.S. Yeoh & Adrian C. North

WARREN BRODSKY

Department of the Arts, Ben-Gurion University of the Negev, Beer-Sheva, Israel

ABSTRACT: Yeoh and North (2009) believe their findings are a clear indication that musical fit can influence product choice among a non-Western sample; pieces of Indian and Malay music primed the selection of one corresponding type of product over another, while when there was no music played the choices indicated no reliable preference for either Indian or Malay items. However, a closer look at their methods of sampling, stimuli, procedures, and results, makes it clear that these conclusions are unfounded.

Submitted: 2009 October 15; accepted 2009 October 20.

KEYWORDS: *musical fit, non-Western, selection priming, consumers, ecological validity*

THE basis of Yeoh and North’s (2009) study is a claim that music can prime the selection of products and influence consumers, because it activates related knowledge of the world, and subsequently raises the salience of certain products and behaviors associated with that music. The authors rally the premise that music, which corresponds to attributes of certain products, can *prime* the selection of those products. Thus far, there is nothing new here. The first claim was already demonstrated by North and colleagues’ groundbreaking (now landmark) study (North, Hargreaves & McKendrick, 1997; 1999), while the second claim was eloquently illustrated in a later study (North, MacKenzie, Law, & Hargreaves, 2004). The current investigation is not only dwarfed by these previous efforts, but the empirical work and theoretical contributions of the present investigation seem minimal.

Yeoh and North may feel that their research corroborates previous contributions but in a Non-Western environment. After all, the researchers recruited a sample of ethnically-Chinese Malaysian residents and presented them with Malay and Indian music. However, the methods used in the study raise serious questions concerning ecological validity and empirical reliability, and I am dubious that much can be learned from this investigation. There is perhaps another, and somewhat politically fueled, agenda at bay: it is often noted that psychological investigations of music are tarnished by their exclusive focus on Western culture, Western Art Classical music genres, and elitist performances by trained musicians. However, I will not address this issue in the current commentary.

There is a lack of ecological validity and reliability concerning the methods employed by Yeoh and North. These limitations are outlined below in four major areas as they relate to sampling, stimuli, procedural methods, and main findings.

Sampling

Certainly, the *raison d’être* of the study was to recruit a non-Western sample. More effort and detail should have gone into this task. First of all, the authors claim that the residents of Malay are made up of three major ethnic groups: Malays (54.1%), Chinese (21.3%) and Indians (7.5%). Yet they offer no documentation to substantiate such frequencies. But if these percentages are correct, then perhaps they could have provided some justification as to why they chose to single out ethnically-Chinese Malaysians as the targeted sample. Especially since the overriding strategy behind the study was to explore the effect of music on preferences between products both associated with two resident cultures in Malaysia among participants drawn from a third cultural group, then looking at the frequencies, one might have thought that recruiting Malays would be warranted because they represent the majority or typical average cultural milieu among the population.

The authors make several questionable assumptions. For example, they claim that people prefer products from their own ethnic group. Further, they assume that individuals from a particular ethnicity are not usually

familiar with cultural aspects, behaviors, and products associated with other ethnicities. These assumptions lead to the questions driving the investigation. The authors argue that “lack of familiarity with products may leave participants even more susceptible to the influence of external cues such as musical fit.” (Yeoh & North, 2009, p. 133). This argument leads to the prediction that ethnically-Chinese residents of Malay should prefer Malay goods when presented with Malay music, but Indian goods when presented with Indian music. But in the local setting of Malaysia, where three ethnicities live side-by-side, cultural isolation and ethnic uniqueness may not exist.

Clearly, we live in a 21st century world, whereby products such as *Nike* and *Nokia* cross all geopolitical and cultural boundaries. Today, ethnologists question whether there *are* remnants of cultural innocence and isolation. As a resident of Israel, I have personally witnessed ethnic contagion resulting in cultural blurring on a massive scale. Historically, over the past 62 years since its establishment, The State of Israel has welcomed a post-Holocaust homecoming of roughly 65% of its Jewish inhabitants through immigration from European-American descent (35%)^[1], African origins (15%)^[2], and Asian derivations (12%)^[3] (CBS, 2009). As a result, even our best ethnomusicologists find it exasperating to isolate the earliest musical traditions among the 1st generation immigrants, while there is no doubt that the 2nd generation has experienced a potpourri of local verse, rhythms, and tonalities that smear cultural distinctiveness. Moreover, today’s 3rd generation descendents, as well as the children of those arriving in this millennium, encounter a stalwart *emptyoducation*^[4] in opposition their ethnic individuality.

Like in so many other regions of the world, we see that folk musics, as well as ethnic religious-based music cultures, have been eclipsed by widespread music marketing efforts heralding personalities such as Michael Jackson and Madonna as the prophets to an alternative ethnicity. Therefore, in the context of the current study, Yeoh and North should have demonstrated greater control over sampling. Readers may ask: How were Chinese participants recruited from University Putra in Malaysia? How long have they lived in Malaysia? If born in Malaysia as descendants, how far removed are they from the original émigré generation? Were they raised in close-quarter ghetto-like villages retaining greater ethnic heritage? Do they currently live in student dormitories and/or apartment buildings allowing for a higher degree of exposure to global marketing and social/national enculturation? No details are provided, while readers are to focus their attention on the use of ‘a non-Western sample.’ What we can read between the lines is that participants were well versed in Western-cultured commercial, intellectual, cultural, and digital worlds.

Stimuli

The text details the fact that classical Malay or Indian music was presented to each participant (at least those in the music-exposure groups). Each piece was taken from a commercially available CD, performed on authentic instruments by a well-known master. The authors offer details that impress upon the readers that the music stimuli used in the study are highly reliable. For example, they state that a pre-study pilot was designed “to ensure that the music chosen was recognizable as either classical Malay or classical Indian music among the population from which the sample for the main study was drawn.” Yet, other essential information required to assess ecological validity is missing, and as such, the consequences are suspicions to the contrary. That is, while the authors concluded from the pilot that seven seconds was the most appropriate and reasonable time frame to complete their forced-choice decision task, and hence employed this duration for both music exemplars, there is no attempt to substantiate if such an exposure timeframe for music effects within a market-research context is of any advantage.

It is crucial that experimenters control the stimuli they work with, especially as these are as potent an influence on *effect-size* as is the sample-size on *statistical power*. Readers may also ask: Is 7 seconds enough? Moreover, one wonders whether these few bars of music were taken from the beginning of the piece or some other segment. The authors offer no details, but such details are especially essential for non-Western music. In the classical Indian raga, 7 seconds of the introductory *alap* section would involve long free-flowing melodic improvisations, whereas 7 seconds of the *chor-alap* section would be more animated with distinctive rhythmic pulses, and 7 seconds of the *jhala* section in which the percussionist joins-in would have sounded most rhythmically complex and even reaching climatic excitement. I am hopeful that the crux is not: *it simply doesn't matter as long as there is a sitar in it.*

The study also used a procedure whereby 24 pairs of products in similar categories, positioned side by side, were presented. The example given is: “A Canon digital camera versus a Kodak digital camera.” However, in actuality, the researchers were only interested in 6 pairs of pictures which featured Indian and Malay items: Malay rendang curry versus Indian tikka masala; Malay keropoks versus Indian papadams; Malay popiahs versus Indian samosas; a Malay gambus versus an Indian sitar; a Malay gendang versus an Indian tabla; and a picture of Malay versus Indian dancers. I find it unacceptable that the authors offer no clear information as to their motives for choosing 3-out-of-6 items from a food category and two items depicting musical instruments. One wonders why

more effort wasn't put into developing a varied set of items that represent a wider range of ethnic behaviors, ornaments, and figures associated to either Malay or Indian cultures? As their reported production method was "downloading pictures from the internet," there really is no excuse here.

Procedure

In the procedures described by Yeoh and North, the participants sat facing a laptop computer and were shown 24 pictures of paired products while hearing music via headphones at a constant comfortable volume, and then after seven seconds indicated their preference between the products by writing 'A' or 'B' on a piece of paper. First, given that a laptop computer was employed, one questions why *E-Prime* or other experiment presentation software wasn't used. These research tools secure more reliable data, offer options to log various types of responses (such as reaction time), and enable the randomization of presented stimuli. For example, what assurances (i.e., empirical controls) can the researchers offer that participants didn't jot down 'A' in place of 'B' incorrectly? Second, no information about what happened in-between each 7 second picture presentation is mentioned. Totally swept aside is the fact that there is a pen and paper response interim, and some type of visual image is either still on the computer monitor or a different one has replaced the original picture on the screen. Third, there are no details concerning the 'no-music' condition: Did these participants wear headphones to control for masking effects and other perceptual idiosyncrasies present in fixed-field listening experiments?

I feel that it is essential for studies that compare music and no-music in fixed-field exposure conditions to take all precautions to control for experimental confounds or 'noise.' This means, for example, that all subjects should wear headphones whether or not they are assigned to hear music. Fourth, the text is clear that the order of picture presentation was not randomized: "The [six] pairs appeared as the fourth, eighth, twelfth, sixteenth, twentieth and twenty-fourth slides in the *Powerpoint* presentation." This leads the reader to assume that the spatial layout of the two pictures was held constant. That is, Malay products were viewed screen right, while Indian products were viewed screen left (or visa versa), or perhaps they were in a standard alternating sequence. With this in mind, and considering the fact that the participants were exposed to either Malay, Indian music or no music (i.e., single condition), then, another disturbing dilemma surfaces.

I summarize thus far: three groups each heard one music or no-music as background to a computerized slide presentation, in which each slide was viewed in identical orders and duplicate layouts; participants were then required to make a forced-choice A/B decision. What about training effects? After how many items did the university undergrad participants, who are somewhat savvy at taking exams, realize that there was a constant correct answer? Had response times been collected, or had more items been entered into the data set (see comment below), this query would not remain a mystery. I feel the researchers should have taken on a more cogent method. They should have sampled all 90 subjects together as one group, in which all participants would have viewed all 24 pairs of items in random orders and spatial layouts – perhaps even twice or three times – whereby both music types and no-music as three background conditions would be presented in random order sequences. In my most modest estimate, given that each item is to be displayed for 7000 msec, with another 5000 msec for response time, a full run of each set would be approximately 5 minutes; therefore three full runs, even with required briefing, practice, and debriefing stages, would take approximate 25-30 minutes. Moreover, had the study employed this suggested paradigm, then, the authors could have reported findings from a two-way [2(product) x 3(background)] within-subjects repeated measures analyses of variance design – a more appropriate method ensuring a higher level of rigor and reliability.

I now take this opportunity to question a methodology used in the current study, which North has employed in past investigations. The issue I raise here is his excessive use of decoys (i.e., lure items). Although a University ethics committee presumably approved the research, it is always troubling when researchers use trickery and misrepresentation in research projects. Collectively, we teach our students to adhere to *APA Ethical Standards*, which call for humane treatment of participants including their *informed* consent and right to withdraw. Moreover, Helsinki ethics approval forbids researchers from collecting samples or data that are not directly required or relevant to the study, including: demographic information, body fluids, audio/video recordings, and empirical responses. In a previous study on time-perception in a gymnasium (North, Hargreaves, & Heath, 1998), not only was the purpose of the investigation misrepresented to participating consumers of a university sports center (by telling them that the study was implemented by the management seeking to improve their services), but in an effort to substantiate such a 'cover-story' each participant was required to complete an additional 12 out of 17 (70%) questions about the service provider – even though the researchers were clearly not interested in such matters. This strategy can be found in several other publications of North and colleagues.

In the current study, each participant was required to respond to an additional 18 out of 24 (75%) items in

order to disguise the true research goals. I worry about this strategy of occupying the time of participants by asking for irrelevant information. Ironically, had the researchers had the foresight to recognize the importance of the data they discarded, they might have explored differences of response characteristics to congruent (pictures of ethnic products) versus non-congruent (pictures of non-ethnic products) items. Findings from this kind of effort might have offered a much greater understanding of the effects that music can have within a multi-cultural advertising context.

Results

Readers may find it difficult to understand the results. While the researchers calculated mean frequencies of correct responses out of 6 trials per group, they offer no visual aid to understand the essence of findings. I have taken the liberty to collate the data, presenting their described results as Hits (i.e., mean frequencies). I have added the same information as percent correct (PC) to allow for ease of interpretation.

	<u>Malay Music</u>		<u>Indian Music</u>		<u>No Music</u>	
	<u>N Hits</u>	<u>PC %</u>	<u>N Hits</u>	<u>PC %</u>	<u>N Hits</u>	<u>PC %</u>
Malay Products	3.53	59	2.07	34	2.80	47
Indian Products	2.47	41	3.94	66	3.20	53
Total	6	100	6	100	6	100

Table 1. Collated data of the group mean frequencies presented by Yeoh and North.

The researchers implemented a between-groups one-way analysis of variance on the group mean frequencies. They conclude that there is a clear indication that musical fit can influence product choice: “the pieces of Indian and Malay music...primed the selection of one corresponding type of product over another. When there was no music played, participants were not primed to choose one product over the other, and in that case the choices indicated no reliable preference for Indian or Malay items.” (Yeoh & North, 2009, p. 133). Here the research has taken on a maneuver that in my opinion has tainted the findings: on the one hand, Yeoh and North focus on between-groups comparisons (attesting to interactions^[5] highlighting differences of the 60%-40% split based on music condition), while on the other hand they employ a within-group comparison (attesting to no preference, or chance levels, as seen in the no-music group). I believe that this inconsistency has led the authors down the wrong road, and in fact, the findings they present, and their interpretations, are no more than ‘wishful thinking.’

I would suggest that before comparing groups, the main issue that needs to be resolved, no doubt, is the reliability of preferences among each and every group. Otherwise, calculating differences between the groups might have no ‘clinical significance’ even though statistical evidence would show otherwise. In some circles, we refer to such situations as a Type I Error. If we re-examine the group means using a sign test, whereby if both outcomes (A/B) have equal probabilities (i.e., probability of success is 0.5), then even four successes out of six trials would not achieve statistical significance ($p = 0.69$, 2-tailed test, or $p = 0.34$, 1-tailed test). That is, the 60%-40% preferences seen between Malay versus Indian products, taking into account only six trials, may reflect statistical noise – reflecting no more than chance levels. Obviously, with only six trials per groups, it is no wonder that statistical power places the data against all odds. So, while Yeoh and North attempt to present the figures in such a way as to illustrate differences between groups regarding their choice of products as based on the effects of music background, preferences within groups are most likely undifferentiated, and hence, meaningless. At best, the current study should be viewed as a pilot study with interesting trends. Without a more serious follow-up, no further conclusions should be drawn.

The subject matter at hand *is* very important. Musical-fit effects do have implications for manufacturers, marketing agencies, and advertisers. Further, retail outlets should be sensitive to the diversity of a multi-cultural society and employ music in an appropriate manner. However, if studies such as this wish to address broader issues, especially when considering non-Western music cultures and ethnicities, then a higher level of empirical rigor should be incorporated.

NOTES

[1] Jewish immigrants from: USSR; Poland; Romania; Bulgaria & Greece; Germany & Austria; Czech Republic, Slovakia & Hungary; France; UK; USA, Canada & Oceania; Argentina & Latin America.

[2] Jewish immigrants from: Morocco; Algeria & Tunisia; Libya; Egypt; Ethiopia; Other.

[3] Jewish immigrants from: Turkey; Iraq; Yemen; Iran; India & Pakistan; Syria & Lebanon; Other

[4] Emptypoducation is an *onomatopoeia* descriptor term consisting of the word 'education' interrupted by two phenomena: 'MTV' and 'iPod'. When said quickly, MTV+iPod become 'emptypod'.

[5] Clearly, the supplied statistic (i.e., $F_{(2, 90)} = 10.16, p < .001$) is incorrectly cited for a sample of 90.

REFERENCES

Central Bureau of Statistics (CBS) (2009) Statistical Abstract Of Israel 2009. <http://www.cbs.gov.il>.

North, A. C., Hargreaves, D. J., & Heath, S. J. (1998). Musical tempo and time perception in a gymnasium. *Psychology of Music, 26*, 78-88.

North, A. C., Hargreaves, D. J., & McKendrick, J. (1997). In-store music affects product choice. *Nature, 390*, 132.

North, A. C., Hargreaves, D. J., & McKendrick, J. (1999). The influence of in-store music on wine selections. *Journal of Applied Psychology, 84*, 271-276.

North, A. C., MacKenzie, L., Law, R. M., & Hargreaves, D. J., (2004). The effects of musical and voice 'fit' on responses to advertisements. *Journal of Applied Social Psychology, 34*, 1675-1708.

Yeoh, P.S. & North, A.C. (2009). The effects of musical fit on choice between competing pairs of cultural products. *Empirical Musicology Review, 4*(4), 130-133.