VERIFICATION AND EXTENSION OF THE MOZART EFFECT

JASON M. HABERMAN

Abstract:

Attempted to verify and extend the findings of F. H. Rauscher, G. L. Shaw, and K. N. Ky, who found a significant improvement on a spatial-temporal reasoning task after exposure to a selection by W. A. Mozart. The current study investigated whether the findings of Rauscher et al. could be replicated using a different dependent measure, reflected in both accuracy scores and overall reaction times. A modification to the original paradigm, the present study includes compositions by Beethoven and Schoenberg, testing whether there exists an effect for these radically different composers as well. Participants (10 men, 2 women) each listened to four selections, took a spatial relations task after each listening, and answered a questionnaire examining their musical experience with the piece. No significant effects were found for either accuracy or reaction time as a function of listening condition. No significant correlations were found between likeability of a given piece and accuracy. The present data fail to confirm the existence of the Mozart effect, but given the controversial nature of this area, further tests are required before any definitive conclusion can be reached. Possible explanations are also offered to explain this data, and that of other studies.

Introduction

The field of music cognition is a rapidly expanding and highly controversial field in the psychological realm. One of the most contested areas of music cognition involves exploring the existence of the putative Mozart effect, first discovered by Rauscher, Shaw, and Ky in 1993. The finding that listening to a certain piece by W. A. Mozart could temporarily enhance spatial-temporal reasoning abilities inspired a veritable media frenzy (Shaw, 2000, p. 163). Popular newspapers and magazines published numerous articles heralding classical music as an intelligence booster, claiming that an individual's I.Q. could actually change through brief musical exposure. A cornucopia of research followed attempting to verify the Mozart effect, but the results have been anything but clear-cut. Many studies, using comparable methodology outlined by Rauscher et al., have replicated the significant improvement on certain spatial tasks, although the effect they found was often not as dramatic as that obtained in the original research (i.e. Rideout, Dougherty, and Wernert, 1998; Rideout and Laubach, 1996; Rideout and Taylor, 1997) while many others, often varying in their methodological approaches, have failed to replicate this effect (i.e. Carstens, Huskins, and Hounshell, 1995; Newman et al., 1995; Stough, Kerkin, Bates, and Mangan, 1994).

Rauscher and Shaw attribute these replication failures to a number of sources, the most significant being the wide variation of dependent measures used in measuring spatial reasoning abilities (1998). They also claim that other factors, including presentation order of stimuli, musical selection, and the inclusion of a distractor task may have occluded results. In their original study, the researches used three indices from the Stanford-Binet Scales of Intelligence (paper folding and cutting task, pattern analysis, and matrices). The authors assert that only one of these tasks, namely the paper folding and cutting task (PF&C), tests spatial-temporal reasoning ability, specifically. Other research, which made use of test items from Raven's Progressive Matrices (Newman et al., 1995) or other tests, do not accurately test this specific ability, the authors claim. Indeed, their own original study did not show an effect on pattern analysis and matrices tasks, which they explicitly state are non-spatial-temporal tasks. It did, however, show a highly significant, large effect for the first task.

In 1995, Rauscher et al. attempted to replicate and extend their study. They reused the Stanford-Binet PF&C task, but this time included a minimalist music condition, in which a piece by Philip Glass was used, as well as some British "trance" music. Both of these pieces, used as an additional control condition, are characterized by highly repetitive, easily recognizable patterns. Adding support to their claims, the authors found a large significant effect as participants showed vast improvement on the PF&C/Mozart task compared to the control conditions. In other words, the music of Mozart continued to improve, at least temporarily, spatial-temporal tasks, while the minimalist music of Philip Glass showed no differentiation. They administered varying formats of this 16-item task to participants over a five-day period, until a ceiling effect was found for those in the Mozart condition. Many critics point out that improvements on test scores may be the result of practice effects, but the authors maintain that on every test day, the Mozart condition participants performed consistently better than controls, although the differences became attenuated as more test days expired. They attribute the immediate improvement on test scores, though, exclusively to the listening condition.

Finally, research done by Sarnthein et al. (1997) seems to provide substantial support to the existence of the Mozart effect in the form of neurophysiological evidence. Researchers administered the same PF&C task given to participants in Rauscher et al.'s experiment, with the added element of measuring EEG correlations during music both listening and testing. Significant findings indicate that regions activated during listening in the Mozart condition, namely right frontal and left temporo-parietal cortices, are also activated during PF&C test taking. As the authors suggest, the music acts as a priming agent, and places certain regions of the brain in an activation synchrony that is valuable during spatial-temporal tasks. Additionally, the researches found that the effects of this cortical activation remains for just over 12 minutes, suggesting that there is significant carryover from the listening session to the test taking session (for an additional EEG correlational study, see Rideout and Laubach, 1996).

It is a common thought that the utility of an experimental finding can only be determined by its verification across several comparable dependent measures and/or varying general approaches. Certainly, it may be correct to blame dependent measures as the reason for other researchers' inability to replicate the Mozart effect, but if these alternate tests are quantifying the same capacities as the one used by Rauscher et al., how could that be considered a compelling defense? If indeed there does exist a Mozart effect, but it can only be seen on a specific index, how much external validity does it hold? Although neurophysiological studies such as the one outlined above can provide powerful evidence in support of a theory, in addition to the multitude of behavioral studies that seem to verify the existence of the Mozart effect, it must be remembered that the dependent measure used to quantify spatial-temporal reasoning abilities remained the same throughout. Most attempts to use alternative measures have failed to replicate the Mozart effect.

The current study, therefore, attempted to verify the Mozart effect by using a spatial reasoning task never before used by music cognition experts, namely one developed by The Psychological Corporation (1974). Given the nature of this task, in which participants had to determine which specific shape is the correct representation of the figure "unfolded", it was thought that it would effectively measure the same abilities as those used to complete the PF&C task. It was logical, then, to maintain the conditions of the original task to see if such an enhancement in spatial-temporal abilities could be seen using an alternative dependent measure. Additionally, if there does exist a Mozart effect reflected in this particular measure, I hypothesize that such an effect would extend to other forms of music written by "great" composers. I presented participants, then, with the music originally used in Rauscher et al.'s seminal study, and also presented participants with a work by Beethoven, as well as a work by the 20th century serialist composer, Arnold Schoenberg[1]. Given the radical differences in composing styles of the first two composers versus that of the latter (and, as some musicologists would contend, the radical differences in composing styles of the first two composers themselves), such explorations may delineate some of the limitations of a Mozart effect, and highlight the point at which sound no longer provides a facilitating, beneficial effect on spatial-temporal reasoning abilities. Due to its discordant and disharmonious nature, then, I expect to see a negative effect on spatial tasks for Schoenberg, reflected as lower accuracy and higher reaction times in response to questions relative to controls[2], while a positive effect should be seen

[1] Schoenberg was the founder of the Second Viennese School of Music (Stolba, 1998, p. 595) and created a highly intellectualized style of composition, in which the pitches of a piece are based upon a pre-established pattern. It is highly atonal, breaking nearly every rule of western harmony listeners have been socialized to

[2] This is not to place a qualitative judgment on the works of Schoenberg, but serves only to suggest that, when it comes to spatial-temporal tasks, his music may serve as a distraction rather than a facilitator.

associated with the aesthetic and organized works of Mozart and Beethoven, shown by higher accuracy and lower reaction times relative to controls.

Methods

Participants

Participants were 12 undergraduate students (10 men, 2 women), the majority of whom volunteered as in order to fulfill course requirement. Consistent with previous experiments, all participants were musically naïve, as determined by the experimenter prior to testing. Consent was obtained from each participant before testing began.

Apparatus

A Macintosh G3 was used to present the various test stimuli, measuring reaction time in milliseconds as well as overall test accuracy. The musical selections were all between 9 and 11 minutes long, and were taken from their original CD's (or tape for the control condition) and copied onto computer by the I-tunes computer program. I-tunes was used to present the musical selections, played through Koss headphones. Participants had the opportunity to manually adjust the listening level for each piece.

Stimuli

The selections presented were: 1) Progressive relaxation tape (control), 2) W. A. Mozart's Sonata for Two Pianos in D Major (K. 448), Allegro, 3) Ludwig van Beethoven's Hammerklavier Sonata, Allegro, and 4) Arnold Schoenberg's Five Piano Pieces (op. 23). Additionally, a 7-question questionnaire pertaining to the listeners' musical experience was administered to participants after each test. The dependent measure was the Space Relations test, taken from Form T of the Differential Aptitude Tests (The Psychological Corporation, 1972). The test was comprised of 60 images of unfolded shapes, and participants had to select the correctly folded shape from among four choices by pushing 1, 2, 3, or 4. Since there were four test conditions, the test was partitioned into four separate tests, 15 questions each, counterbalanced for difficulty.

Procedure

Each participant was tested individually. Before testing began, a brief screening was administered to verify participants' musical naiveté. The experimenter presented each participant with a musical selection (the first page from J. S. Bach's E-Major Partita for Solo Violin), and asked if the participant could identify the first note of the selection (an "E") and if the participant could name the key the piece was written in (E-Major). Nearly all participants were unable to answer these questions, and therefore met the requisite for being musically naïve.

Testing began immediately after by asking participants to place the headphones on and adjust the volume to a comfortable level. The experimenter then presented the first musical stimuli. Presentation of the four stimuli was randomized, as well as the order in which the tests were administered. Therefore no test corresponded with any musical condition, and in fact, no two participants received the same order of musical stimuli and test presentation. During listening, the computer monitor was turned off and the experimenter left the room. Participants were instructed to state when the selection was complete, at which time the experimenter turned the monitor back on and set up the first spatial task. Individuals input their name, age, and gender, and were then presented with an instructions screen detailing the nature of the test. Speed was emphasized, but not at the expense of accuracy. After completion of the spatial task, participants were given the 7-question questionnaire, which asked questions regarding likeability and perceived complexity of each selection. Since there were four conditions, each participant underwent this cycle four times. Total testing time varied from participant to participant, depending upon proficiency with the spatial task, ranging from one to two hours each. Participants were given the opportunity after the experiment to ask questions, which the experimenter answered fully.

Results

A one-way repeated measures analysis of variance was used to determine the relationship between listening condition (control, Mozart, Beethoven, or Schoenberg) and overall accuracy on the spatial reasoning task. For an alpha level of .05, the obtained F ratio was found to be statistically non-significant, $F(3, 27) = .59$, $p > .05$, suggesting that the listening condition had no impact on test accuracy. Additionally, another one-way repeated measures analysis of variance was used to determine the relationship between music condition and reaction time for correct answers. The obtained F ratio was also found to be statistically non-significant, $F(3, 27) = .50$, $p > .05$, suggesting again that there was no relationship between listening condition and reaction time. The strength of the relationship, as indexed by eta^2, is not reported due to the non-significant natures of the statistical tests.

Individual correlated groups t tests were also run to determine differences between each musical condition (Mozart, Beethoven, and Schoenberg) and the control group. All tests were found to be statistically non-significant, for both accuracy and reaction times, although a near-significant level was obtained when the Mozart condition was compared to the control group, $t(9) = 1.94$, $p = .08$ for accuracy and $t(9) = 1.43$, $p = .18$ for reaction times. However, the near-significant level for accuracy was contrary to the original prediction, indicating that the Mozart listening ($M = 10.6$, $SD = 2.22$) may actually have had a negative impact on the performance on this spatial task relative to controls ($M = 11.7$, $SD = 1.49$). The prediction for reaction time, however, was in the correct direction, implying that listening to Mozart may decrease reaction times ($M = 20.87$, $SD = 12.27$), although at the expense of accuracy, relative to controls ($M = 23.95$, $SD = 15.13$). However, these tests reflect only trends and caution should be used in interpreting such data. They are non-significant, and quite feasibly could be merely a result of chance. Due to the likelihood that these statistics are derived from chance, further digression into the possible interpretations of such data will be avoided

Since no significant relationships between accuracy and spatial-temporal reasoning or between reaction times and spatial-temporal abilities were found, I decided to investigate the possibility that performance on these tests was a function of emotional arousal, reflected by the likeability of a given piece of music, since some previous studies attributed the Mozart effect artifacts of preference (e.g. Nantais and Schellenberg, 1999). Correlated group t tests were run to compare favorite piece accuracy with least favorite piece accuracy and also for reaction times (questions regarding perceived complexity were not analyzed), but were also found to be non-significant: $t(9) = 2.84E-07$, $p = 1$ for accuracy on favorite piece ($M = 11.3$) compared with accuracy on least favorite piece ($M = 11.3$) and $t(9) = 1.08$, $p > .05$ for reaction time on favorite piece ($M = 20.382$) compared with reaction time on least favorite piece ($M = 24.59$). Finally, a Pearson correlation was also used to address the relationship between likeability of a given piece ($M = 4.15$, $SD = 1.64$) with overall accuracy ($M = 11.2$, $SD = 2.45$). This test was statistically non-significant, $r(39) = -.13$, $p > .05$, reinforcing the idea that likeability had no bearing on the participants' performance on these spatial tasks.

Discussion

The inability to find any statistical significance makes one immediately wonder if there was not some methodological problem associated with the experimental design. Several potential problems come to mind including, low participant number, carryover effects between listening conditions, and difficulty differentiation among spatial tests. For the first problem, statistical analyses revealed no general trends besides the one mentioned above, and given the within subjects design of the experiment, the number of participants should have provided enough power to evoke certain patterns, patterns that were not seen. Therefore it is unlikely that the results reflect a low participant number.

Since in previous studies the Mozart effect was determined to last only a little over 12 minutes, it seems improbable that any effect exerted by the listenings would carry over to the test following the next listening condition. The participants all took the spatial test directly after the listening was completed, and most participants took around 10 minutes to finish. Additionally, participants had to finish a questionnaire after each test, extending the amount of time between each listening stimulus.

Finally, the four 15-question tests were taken from the original 60 question assessment, which gradually increases in difficulty as the test progresses. In designing the various test stimuli the items were taken with this gradation in mind. Although the four 15-question tests were not professionally rated to determine their overall difficulties in relation to each other, participant's performance on these tests did not differ overall as a function of test discrepancies.

Given the fact that none of these explanations reasonably account for the data observed , I must consider the possibility that the current study is a valid assessment of the Mozart effect, and that it may be simply a failure to replicate the original findings. This may be due to any number of factors, and I imagine Rauscher would claim that my dependent measure was not an accurate measure of spatial-temporal reasoning ability. However, as alluded to previously, how valid is the Mozart effect, then, if it can only be seen on one kind of dependent measure? Faith in an experimental discovery is always strengthened by the ability not only to replicate its original findings, but also to replicate those findings using alternate methodological paths. The scientific literature, including the present study, have yet to achieve that goal.

Because of these discrepancies, many researchers have shied away from searching for alternative paths and turned their attention to searching for alternative explanations. The power of placebo must therefore not be overlooked. Music is highly mood evocative, and moods can often influence performance on a given task, even a spatial-temporal one. For this reason, several researchers maintain that the overall affinity for a piece may have influenced in no small way certain findings, and which is why the current investigation attempted, unsuccessfully, to correlate likeability with accuracy and reaction time. For example, a meta-analysis performed by Chabris (1999) revealed an average cognitive enhancement of only $\underline{d} = 0.09$, which accounts for an average "IQ boost" (even the scientific literature has adopted the notion that reasoning abilities are synonymous with IQ) of only 1.4 points, as opposed to the eight or nine points reported by Rauscher et al. This, Chabris contends, is a result of enhanced hemispheric arousal, what he calls "enjoyment arousal", and could positively influence overall test scores. Additionally, poorer performance in controls (after listening to a relaxation tape) remains consistent with this hypothesis, since relaxation tapes are generally designed to reduce arousal. Such emotional reduction may negatively impact performance on these special tasks. This "enjoyment arousal" idea was confirmed, the author cites, in children who performed better on certain tasks after exposure to popular music (supposedly more enjoyable for them) than after exposure to Mozart.

Unfortunately, the current study did not find any such emotional correlations, and as Rauscher might claim (1999), this is not surprising. In his response to Chabris, he cites several findings that contradict the "enjoyment arousal" hypothesis put forth, stating that despite participants' ratings of a Mendelssohn piece as being the most arousing stimuli of the set, the spatial-temporal effect was most prominently seen for the Mozart condition. He also claims that students scored higher after listening to the Mozart than after listening to other conditions, regardless of their reported musical preferences. This is consistent with my finding that piece likeability is not correlated with performance on the spatial relations task. Additionally, Rauscher dismisses Chabris' meta-analysis as an inaccurate representation of the studies involving the Mozart effect, since he included assessments beyond the PF&C task used, almost exclusively, by Rauscher. Once those other tasks were removed from the analysis, Rauscher claims, the Mozart effect became very prominent.

Despite the inability of this study, and several others, to replicate a Mozart effect, it does not serve to nullify the significant findings of Rauscher and his colleagues, who continuously find an effect. This reliability, at least on the PF&C task, suggests that this type of music exerts some kind of facilitating effect. As Rauscher remarks, "Because some people cannot get bread to rise does not negate the existence of a 'yeast effect'" (1999).

The long-term beneficial effects of music are not being debated here, just the question of whether a short-term, immediate benefit exists. Had I used the original PF&C, I predict there would have been a Mozart and Beethoven effect, independent of likeability ratings, and there would have also been a negative Schoenberg effect. Such revelations could have helped to begin the long process of determining what it is about music that influences the brain in such a seemingly positive manner. Also, as an aside, it could have shed light on the question of whether this effect is restricted to those elite composers who possess an unprecedented talent for composing, synthesizing works that create an intrinsic synchrony with our minds, or if this effect could be evoked by a mediocre composer simply because of the nature of music However, since such insights could not be provided presently, and given the controversial nature of this research, further investigations, especially neurophysiological ones involving varying spatial-temporal tasks and fMRI scans (which, incidentally, are currently being done) are warranted before the practical applications of a Mozart effect can be fully understood.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1972). Differential Aptitude Tests: Space Relations. New York, NY: The Psychological Corporation.

Carstens, C. B., Huskins, E., & Hounshell, G. W. (1995). Listening to Mozart may not enhance performance on the revised Minnesota Paper Form Board Test. Psychological Reports, 77(1), 111-114.

Chabris, C. F., Steele, K. M., Bella, S. D., Peretz, I., Tracey, D., Dawe, L. A., Humphrey, G. K., Shannon, R. A., Kirby, J. L. Jr., Olmstead, C. G., & Rauscher, F. H. (1999). Prelude or requiem for the “Mozart effect”? Nature, 400(6747), 826-828.

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: an artifact of preference. Psychological Science, 10(4), 370-372.

Newman, J., Rosenbach, J. H., Burns, K. L., Latimer, B. C., et al. (1995). An experimental test of the “Mozart effect”: Does listening to his music improve spatial ability? Perceptual and Motor Skills, 81(3, Pt 2), 1379-1387.

Rauscher, F. H. & Shaw, G. L. (1998). Key components of the Mozart effect. Perceptual and Motor Skills, 86, 835-841.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. Nature, 365, 609-611.

Rideout, B. E., Dougherty, S., & Wernert, L. (1998). effect of music on spatial performance: A test of generality. Perceptual and Motor Skills, 86, 512-514

Rideout, B. E., & Laubach, C. M. (1996). EEG correlates of enhanced spatial performance following exposure to music. Perceptual and Motor Skills, 82, 427-432.

Sarnthein, J., vonStein, A., Rappelsberger, P., Petsche, H., Rauscher, F. H., & Shaw, G. L. (1997). Persistent patterns of brain activity: An EEG coherence study of the positive effect ofmusic on spatial-temporal reasoning. Neurological Research, 19, 107-116.

Shaw, G. L. (2000). Keeping Mozart in Mind. San Diego: Academic Press.

Stolba, K. M. (1998). The Development of Western Music: A History (3rd Ed.). Boston, MA: McGraw-Hill Companies, Inc.

Stough, C., Kerkin, B., Bates, T., & Mangan, G. (1994). Music and spatial IQ. Personality and Individual Differences, 17(5), 695.