

# The Effects of Background Music in Advertising: A Reassessment

JAMES J. KELLARIS  
ANTHONY D. COX\*

Gorn's (1982) pioneering article on the effects of background music in advertising has spurred a significant controversy and inspired vigorous interest in the topic. Following the recommendation of Allen and Madden (1985), we conducted three experiments that attempted to replicate Gorn's results. Contrary to Gorn's findings, there was no evidence that product preferences can be conditioned through a single exposure to appealing or unappealing music.

In 1982, Gerald Gorn published an experiment showing that subjects were more likely to select a specific color of pen if that pen had been paired with pleasant rather than unpleasant music. Since that time, Gorn's study has become quite influential. The study has been cited in at least 34 refereed journal articles (*Social Science Citation Index* 1982–1988). Perhaps more importantly, several consumer behavior textbooks now cite the experiment as proof that consumer preferences can be classically conditioned. For example, Hawkins, Best, and Coney (1983, p. 314) cite Gorn's work in stating that advertising's ability to classically condition product preferences is "well established and widely used."

However, Gorn's findings have also been subject to controversy. Several researchers have been unable to find classical conditioning effects in similar experiments (Allen and Madden 1985; Park and Young 1986). Even those authors who have found evidence of classical conditioning under nearly ideal circumstances (e.g. Bierley, McSweeney, and Vannieuwerk 1985; Stuart, Shimp, and Engle 1987) have noted that Gorn was able to obtain surprisingly strong effects given his relatively weak conditioning treatment. (Gorn's subjects were almost three times as likely to choose the advertised pen after it had been paired once with pleasant music.)

Several factors have led to speculation (see especially Allen and Madden 1985) that Gorn's results

may have been due partially to demand artifacts. First, demand artifacts are a recurring issue in psychological research on classical conditioning in humans. Second, certain aspects of Gorn's procedure might have clued subjects to the purpose of the study. For example, subjects were told that the study's purpose was to evaluate ad music, and pen color choice was measured through a fairly obtrusive procedure—subjects wanting different pen colors had to walk to opposite sides of the room.

Given the importance of Gorn's study, as well as the controversy it has generated, Allen and Madden (1985, p. 304) state, "Certainly Gorn's work merits close inspection, and if his results are maintained in replication studies, he will have opened an important new consumer research stream." Allen and Madden themselves conducted an approximate replication of Gorn's study, but used humor rather than music as the unconditioned stimulus. When no conditioning effect was obtained, the authors speculated that music may have some unique conditioning power that would account for Gorn's results. However, no study has tested Gorn's basic finding that product preferences can be conditioned through a single exposure to appealing or unappealing music. This article presents three experiments that attempt to replicate Gorn's work with particular attention to the potential role that demand artifacts may have played in the research.

## EXPERIMENT 1

### Overview

Our first experiment was essentially a replication of Gorn's: subjects were exposed to slides of a colored pen, accompanied by either pleasant or unpleasant

\*James J. Kellaris is a doctoral candidate in the Department of Marketing, College of Business Administration, Georgia State University, Atlanta, GA 30303. Beginning this fall, he will be Assistant Professor of Marketing, College of Business Administration, University of Cincinnati, Cincinnati, OH 45221. Anthony D. Cox is Assistant Professor of Marketing, School of Business, Indiana University, Indianapolis, IN 46223.

music, and then given the opportunity to select either the pen color in the slide or an alternate color. However, minor modifications were made relative to Gorn's original experiment.

First, in manipulating music appeal, we attempted to hold other music characteristics constant. Two recordings were chosen (a Mozart overture and a selection by Milhaud) that were similar in instrumentation, tempo, and modality, but subjectively judged to differ in appeal. On a seven-point scale (7 = like very much/1 = dislike very much), the pretest among undergraduate subjects showed a significant difference in appeal between these pieces ( $\bar{X} = 5.18$  for Mozart and  $\bar{X} = 2.07$  for Milhaud;  $t = 8.98, p < 0.001$ ).

A second modification involved pen color. The experiment required two pen colors for which subjects did not have strong prior preferences, and which did not differ significantly in appeal. Since color preferences might vary across time and locations, we did not simply use Gorn's colors (beige and blue), but pretested preferences for seven pen colors on a student sample. The two pen colors with means closest to the overall mean were yellow ( $\bar{X} = 3.18$ ) and white ( $\bar{X} = 3.24$ ), and these were used in the experiment.

A third change involved the choice procedure. As mentioned earlier, there was some concern that Gorn's choice procedure (in which subjects had to go to opposite sides of the room to choose different pen colors) may have underscored the importance of color choice in subjects' minds. Therefore, we replaced this with a less obtrusive procedure. A box containing a mixture of yellow and white pens was passed around the room. The researcher paid no attention to which pens subjects chose. However, unknown to the subjects as they chose, the yellow pens wrote in blue ink and the white pens in black ink. Thus, choice behavior could be recorded by examining ink colors on the questionnaire. It is possible that some subjects may have noticed this difference after they began writing, but their choices had already been recorded by this time.

The final change involved varying the cover story. As discussed earlier, Gorn's cover story, in which subjects were told explicitly that the study's purpose was to evaluate music for a pen commercial, could have increased demand artifacts. To test this hypothesis, half of the subjects received Gorn's cover story, and the other half was told simply that the study involved "assessing perceptions of various products." Thus, the study used a  $2 \times 2$  factorial design, varying music appeal and cover story.

## Procedure

Subjects were from five undergraduate marketing classes at a large Southeastern university. One class was taught by the author who administered the exper-

iment, and a second was taught by the other author. A total of 302 subjects participated in the experiment, of which 299 yielded complete data. As per Gorn, each class was split in half randomly (by odd or even Social Security number), and each half heard either appealing or unappealing music. Random assignment of cover stories was achieved by shuffling the questionnaires before they were handed out.

After reading their respective cover stories, subjects were exposed for 60 seconds to slides of a yellow pen, accompanied by either appealing (Mozart) or unappealing (Milhaud) music. After this exposure, subjects were told they could choose a pen as a gift for participating. Then, a box containing a mixture of yellow and white pens was circulated. After everyone had taken a pen, subjects were asked, "Why don't you try out your new pens in filling out the questionnaire?"

The questionnaires included a manipulation check for music appeal, a measure of pen choice behavior, and a debriefing question. Each subject rated the appeal of the music heard on a seven-point "like very much" to "dislike very much" scale. Brand choice behavior was measured by examining the ink color of the pen used to fill out the questionnaire. An open-ended question was used to determine what subjects believed to be the purpose of the exercise.

## Results

Manipulation check data reveal a very strong music appeal manipulation. The mean liking (on a seven-point scale) was 6.01 for Mozart and 1.37 for Milhaud ( $t = 47.48, p < 0.0001$ ). However, music appeal had no effect on choice behavior ( $\chi^2 = 0.20, ns$ ). Fifty-six percent of subjects in the high music appeal group chose the advertised pen, compared to 59 percent in the low music appeal group. Participants and choices were tabulated as follows.

<i>Music appeal</i>	<i>Choice of advertised pen</i>	<i>Choice of nonadvertised pen</i>	<i>Total</i>
<i>High</i>	89	69	158
<i>Low</i>	83	58	141
<i>Total</i>	172	127	299

Following Gorn's procedure, this analysis was repeated, this time excluding subjects who reported liking the Milhaud music or disliking the Mozart music. Results were similar. Fifty-four percent of the remaining subjects in the high music appeal group chose the advertised pen, as opposed to 56 percent in the low music appeal group ( $\chi^2 = 0.03, ns$ ).

A check for the effect of cover story version on hypothesis-guessing behavior was performed by examining the responses to the open-ended question at the end of the procedure. Two judges independently coded sub-

jects' responses, and a comparison of their judgments showed 100 percent agreement. Analysis revealed that subjects exposed to Gorn's cover story were more likely to guess that the study's purpose related to music (33.8 percent versus 19.8 percent for the vague cover story;  $\chi^2 = 6.98, p < 0.01$ ). However, only one out of 299 subjects was judged to have actually guessed the hypothesis (i.e., that music would affect pen choice). Thus, contrary to our expectation, the cover story treatment did not influence true hypothesis guessing. We will return to this point later.

The possible interaction of music and cover story was tested by performing a three-way cross-tabulation of music appeal with pen choice, controlling for cover story. No statistically significant association between music and pen choice was found within either cover story group.

## Discussion

The most noteworthy finding of this experiment is that Gorn's results were not reproduced: music appeal had no effect on pen choice. One explanation for this might be differences in the demand characteristics of our two experiments. Our cover story manipulation did not affect either hypothesis guessing or pen choice. However, even our high demand condition did not replicate one aspect of Gorn's procedure: Gorn's subjects were instructed to go to different sides of the room, depending on which color they selected, and to drop off questionnaires as they chose pens. This procedure may have underscored the importance of color choice. Our subjects, on the other hand, simply selected pens from a box containing a mixture of the two colors as it was circulated around the room. Perhaps our choice measure, as well as our procedure, was significantly less obtrusive than Gorn's.

It is interesting that in Allen and Madden's (1985) replication, the cover story remained in essence the same as Gorn's (i.e., subjects were told that a firm was evaluating humor (music) for ads), but the choice method (i.e., pens mixed in a box) was different from Gorn's. They did not get conditioning effects. Combined with the results of our experiment, this suggests that Gorn's pen selection procedure may have produced a demand artifact, which is a plausible alternative explanation for his results. To further address the possible role of demand artifacts, the following nonexperiment was conducted using Gorn's exact choice procedure.

## EXPERIMENT 2

Sawyer (1975) discusses several methods for evaluating experimental demand artifacts. Two of these—postexperimental inquiry and attempting to manipulate demand characteristics—were used in our first

experiment. Another approach is a nonexperiment, which "involves a reenactment of all experimental procedures except for the actual treatment, which is instead only described" (Sawyer 1975, p. 24). If results of a nonexperiment are similar to the results of an actual experiment, this suggests that subjects may have been reacting to the situation rather than to the treatments. This does not prove that results are entirely attributable to demand artifacts, but it does suggest the plausibility of this explanation.

For our second experiment, we chose undergraduate marketing classes whose members had not participated in the first experiment. One of these classes was taught by an author, who did not administer the experiment. A total of 136 subjects participated in the nonexperiment.

The nonexperiment reenacted Gorn's procedure with only two minor changes: (1) yellow and white pens were used instead of light blue and beige for reasons previously explained, and (2) the treatments were described rather than administered. Subjects were asked to imagine pleasant or unpleasant music, which was described alternatively as "a very pleasant, likable piece of upbeat music taken from a popular film," or "a rather unpleasant piece of Indian classical music that most people would dislike." Subjects were also asked to imagine being exposed to slides of pens as per Gorn's experiment. The slides were described as follows. "A color slide projection showed a (white/yellow) ballpoint pen against a plain background. The pen looked inexpensive and had a (white/yellow) plastic barrel."

The cover story and mechanics of the procedure followed to the letter the description given by Gorn of his experiment. For example, subjects within each class were divided randomly into two groups and were assigned treatments according to Gorn's counterbalancing scheme. One group was sent out of the room while the other group was processed. Subjects were asked to imagine that they had actually participated in the procedure described to them. They were instructed to behave exactly as they thought they would have had the events actually taken place.

They were then exposed both verbally and in writing to the cover story and to a description of the treatments. Asking subjects to read along as the experimenter read out loud was believed to increase the probability of the subjects processing the nonstimuli.

After pausing to simulate a 60-second exposure, subjects were asked to rate the music on a five-point scale (5 = like very much/1 = dislike very much) as though they had actually heard it. Subjects were then told that they would "receive either a yellow or white pen for (their) help, donated by the company that manufactures the pen." The experimenter held up each pen briefly and, following Gorn's procedure, instructed subjects wanting a yellow pen to pick one up

on the left side of the room, dropping off the questionnaire in the box provided. If they wanted a white pen, they were told to go to the right side of the room to pick one up and drop off their questionnaire in the box provided.

At the end of class, a subset of 75 subjects was given a debriefing questionnaire that asked them in an open-ended format to report what they believed to be the purpose of the exercise previously described to them.

## Results

The results of the nonexperiment are shown in the following tabulation. The majority (64 percent) of the subjects in the appealing music group chose the advertised pen. Among subjects in the unappealing music group, the majority (64 percent) chose the nonadvertised pen. A chi-squared test shows the relationship between music appeal and choice behavior to be statistically significant ( $\chi^2 = 10.63, p < 0.05$ ).

<i>Music appeal</i>	<i>Choice of advertised pen</i>	<i>Choice of nonadvertised pen</i>	<i>Total</i>
<i>High</i>	45	25	70
<i>Low</i>	24	42	66
<i>Total</i>	69	67	136

Thus, Gorn's results were replicated in the absence of actual treatments. This suggests that some subjects in Gorn's experiment may have been reacting to the demands of the procedure rather than to the experimental stimuli. To explore this further, the open-ended responses to the debriefing question were examined independently by two judges to determine the extent of correct hypothesis guessing. Twenty-nine percent of the subjects were found to have correctly guessed the main hypothesis (i.e., that the music would affect pen choice). In addition, subjects who were able to state the hypothesis were slightly (about 10 percent) more likely to have behaved in the hypothesized manner, though this difference was not statistically significant ( $\chi^2 = 0.68, ns$ ).

## Discussion

Our first experiment, which had a strong manipulation of music appeal and an unobtrusive procedure, found no evidence of conditioning effects. Our nonexperimental replication, which used no actual music manipulation but a more obtrusive procedure, did produce significant results.

These findings and those of other researchers (especially Allen and Madden 1985) suggest the elusiveness of single-exposure conditioning effects, regardless of whether music or humor is used as the unconditioned stimulus. In addition, these results suggest

that demand artifacts may account at least partially for Gorn's findings.

Of course, there are other potential explanations for the difference between our results and Gorn's. One explanation is that in our first experiment the music did alter subjects' color preferences but that our choice procedure (using the box of mixed pens) encouraged subjects to choose blindly or randomly. In other words, Experiment 1 may have failed to detect a classical conditioning effect because our choice procedure was an insensitive measure of true preference differences.

To test this hypothesis, we selected two pen colors (orange and white) that pretests suggested would differ in preference. Next, a sample of 56 subjects was given the opportunity to select either a white or an orange pen. Roughly half of these subjects were randomly assigned Gorn's choice procedure (in which the two colors were placed at opposite sides of the room) and the other half selected from a single box of mixed colors. The choice method had no significant effect on subjects' color choices ( $\chi^2 = 0.75, ns$ ). Interestingly, it was Gorn's procedure that yielded a choice percentage closer to the 50/50 pattern one would expect from mindless selection (52 percent orange, 48 percent white; the mixed box method yielded 63 percent orange, 37 percent white).

Another potential explanation for the difference between Gorn's result and ours might lie with the choice of music. While our manipulation check indicated a very strong difference in appeal between Mozart and Milhaud, it is possible that Gorn's popular appealing music was somehow a more powerful unconditioned stimulus (at least among a student sample) than our classical music. To test this possibility, we decided to conduct an exact replication of Gorn, in which popular music was used as the positive unconditioned stimulus. In addition, this experiment further examined the role of demand artifacts.

## EXPERIMENT 3

### Overview

While our manipulation of music appeal appeared to be very strong (high appeal music  $\bar{X} = 6.01$ , low appeal  $\bar{X} = 1.37$ ;  $t = 47.48, p < 0.001$ ), we wanted to resolve whether our positive unconditioned stimulus (classical music) was weaker than Gorn's positive unconditioned stimulus (popular music). Therefore, our third experiment used music that, like Gorn's, was appealing and familiar, popular and upbeat. We sensed that Gorn's original musical selection from the album "Grease" might no longer be appropriate, since it is now out-of-date and seldom played. To check this, we pretested a once-popular selection from "Grease" and found it scored much lower than

it had in Gorn's pretests. We then selected several pieces of popular music (both current hits and old standards) and pretested them for appeal using Gorn's five-point liking scale. The highest scoring selection was the Beatles' rock-and-roll classic "Twist and Shout" with a mean of 4.5 (stronger than Gorn's pretest of "Grease," which had a mean of 4.3). The low appeal music was Indian raga music (also following Gorn's example) by Ravi Shankar, with a mean score of 1.6 (which was comparable to Gorn's raga music mean of 1.5). Thus, our music appeal manipulation appears to have been at least as strong as Gorn's.

To further explore the role of demand artifacts, our manipulation of music took place under two procedural conditions. One was an exact replication of Gorn's procedure: subjects were told that an ad agency wanted to test ad music for a pen commercial; after the music/slide presentation, subjects were shown two boxes of different colored pens, on opposite sides of the room, and asked to drop their questionnaires next to the box from which they selected a pen. In the other procedural condition, subjects were given the cover story used in Experiment 1 (which contained no mention of the music's importance). After the music/slide exposure in this condition, they were asked to select a pen from a box of mixed colors; subjects' pen choices were monitored by the ink colors on their questionnaires, apparent only after they had made their choices.

Thus, this was a  $2 \times 2$  experiment, in which music appeal (high or low) and procedure (Gorn's or a more disguised procedure) were manipulated. Half of the study was an exact replication of Gorn, and the other half was an extension, differing only in experimental procedure. These procedural aspects would not be expected to diminish a true classical conditioning effect, but they might affect subjects' hypothesis guessing and perception of experimenter demand.

For this experiment, we selected undergraduate marketing classes whose members had not participated in the previous experiments. One of the classes was taught by an author, who did not administer the experiment. A total of 171 subjects participated in this experiment. As before, subjects were assigned to treatments according to their Social Security numbers.

## Results

Again, there was no significant relationship between music appeal and pen choice, even when popular rock music was used as the positive unconditioned stimulus. Fifty-six percent of those hearing the liked music chose the advertised pen versus 54 percent of those hearing disliked music. This difference does not even approach statistical significance ( $\chi^2 = 0.183$  us-

ing Gorn's procedure;  $\chi^2 = 0.052$  using the disguised procedure). The tabulations for Gorn's procedure were:

<i>Music appeal</i>	<i>Choice of advertised pen</i>	<i>Choice of nonadvertised pen</i>	<i>Total</i>
<i>High</i>	22	19	41
<i>Low</i>	23	24	47
<i>Total</i>	45	43	88

The undisguised procedure yielded the following:

<i>Music appeal</i>	<i>Choice of advertised pen</i>	<i>Choice of nonadvertised pen</i>	<i>Total</i>
<i>High</i>	25	18	43
<i>Low</i>	24	16	40
<i>Total</i>	49	34	83

Among the subjects exposed to Gorn's procedure, those hearing the appealing music were slightly more likely to choose the advertised pen (53.6 percent) than those hearing the unappealing music (48.9 percent); however, this difference was not statistically significant. Among those exposed to the less obtrusive procedure, there was a slight (nonsignificant) negative relationship between music appeal and choice of the advertised color (58.1 percent versus 60.0 percent).

There was a significant relationship between procedure and hypothesis guessing. Twelve of the 15 subjects who correctly guessed the hypothesis (that music appeal would influence pen choice) had received Gorn's procedure. This was a significantly higher proportion than would be expected due to chance ( $t = 2.32, p < 0.05$ ). In addition, subjects who were able to state the hypothesis were slightly (about 25 percent) more likely to have behaved in the hypothesized manner, though this difference was not statistically significant ( $\chi^2 = 3.32, ns$ ).

## CONCLUSIONS

In three experiments, we tried to replicate Gorn's finding that product preference can be classically conditioned through a single pairing with background music. None of the experiments supported this hypothesis. Our findings and those of other researchers (Allen and Madden 1985; Park and Young 1986) suggest that the single-exposure conditioning of product preferences, far from being "well established," is an elusive phenomenon.

In trying to interpret their inability to classically condition product preferences (using a single exposure to humor), Allen and Madden (1985) speculated that Gorn's results may have been due to some unique conditioning power of music. However, the present study does not support this speculation. In Experiment 1, subjects heard music of widely differing appeal, with no apparent effect on their product

preferences. In Experiment 3, the same musical genres used by Gorn (upbeat rock-and-roll and Indian raga music) were employed, but still no conditioning effects were detected.

Our experiments also explored the potential role of demand artifacts in Gorn's results. A nonexperimental replication of Gorn's experiment revealed a high rate of hypothesis guessing and produced an apparent conditioning effect in the absence of actual conditioning stimuli. Experiment 3 also found some support for demand artifacts: subjects receiving Gorn's procedure were significantly more likely to guess the experimental hypothesis than those receiving a less obtrusive procedure, and they were slightly (though not significant statistically) more likely to behave in the hypothesized manner.

Our results cast doubt on the contention that product preferences can be conditioned reliably by a single exposure to appealing or unappealing music. This does not mean that affective conditioning never occurs. For example, there is some evidence that it may occur under more ideal circumstances (e.g., see Bierley et al. 1985; Stuart et al. 1987). Nor does it mean that music cannot affect product preferences in other, more indirect, ways (e.g., Allen and Madden 1985; Gardner 1985). Consumer researchers must continue to examine the classical conditioning phenomenon, so that we may define more clearly the boundaries within which it does and does not take place.

[Received November 1987. Revised January 1989.]

## REFERENCES

- Allen, Chris T. and Thomas J. Madden (1985), "A Closer Look at Classical Conditioning," *Journal of Consumer Research*, 12 (December), 301-315.
- Bierley, Calvin, Frances K. McSweeney, and Renee Van-nieuwkerk (1985), "Classical Conditioning of Preferences for Stimuli," *Journal of Consumer Research*, 12 (December), 316-323.
- Gardner, Meryl Paula (1985), "Mood States and Consumer Behavior: A Critical Review," *Journal of Consumer Research*, 12 (December), 281-300.
- Gorn, Gerald J. (1982), "The Effects of Music in Advertising on Choice Behavior: A Classical Conditioning Approach," *Journal of Marketing*, 46 (Winter), 94-101.
- Hawkins, Del, Roger Best, and Kenneth Coney (1983), *Consumer Behavior*, Plano, TX: Business Publications.
- Park, C. Whan and S. Mark Young (1986), "Consumer Responses to Television Commercials: The Impact of Involvement and Background Music on Brand Attitude Formation," *Journal of Marketing Research*, 23 (February), 11-24.
- Sawyer, Alan G. (1975), "Demand Artifacts in Laboratory Experiments in Consumer Research," *Journal of Consumer Research*, 1 (March), 20-29.
- Social Science Citation Index* (1982-1988), Philadelphia PA: Institute for Scientific Information.
- Stuart, Elnora, Terence Shimp, and Randall Engle (1987), "Classical Conditioning of Consumer Attitudes: Four Experiments in an Advertising Context," *Journal of Consumer Research*, 14 (December), 334-349.