Effects of Brand-Fit Music on Consumer Behavior: A Field Experiment

Daunfeldt, Sven-Olov*
Rudholm, Niklas*
Sporre, Hampus†

March 27, 2017

Abstract

Businesses might use music to align consumers with brand values, thereby influencing consumers' choices and perceptions. However, previous studies have focused on the effects of various characteristics of the music choice (e.g., tempo and style) and not on the effect of the congruence between music and brand values. Our cooperation with Soundtrack Your Brand, the exclusive provider of Spotify Business, makes it possible for us to test the effect of congruence between music and brand values on consumers in a field experiment using 16 chain restaurants within the Stockholm metropolitan area. Our results show that a playlist that only includes brand-fit songs increases revenues by 9.1 percent in comparison to selecting music that does not fit the brand. We also find that brand-fit music has a positive impact on consumers' emotions and that music seems to have an unconscious effect on consumers.

Keywords: Consumer choice, store atmospherics, food consumption, brand values, emotions, self-control.

JEL-codes: D03, D12, M31, L83.

*HUI Research, SE-103 29 Stockholm, Sweden; and Department of Economics, Dalarna University, SE-781 88 Borlänge, Sweden.

†HUI Research, SE-103 29 Stockholm, Sweden; and Department of Economics, Lund University, SE-220 07 Lund, Sweden. E-mail: hampus.sporre@nek.lu.se.

1 Introduction

It has long been recognized that music says something about an individual's personality and can affect individuals' social relationships (Cattell and Saunders, 1954). Individuals often use music to give others an image of themselves and to create perceptions of other individuals (Rentfrow and Gosling, 2006). However, can businesses use music in the same manner to convey brand values and appeal to consumers' self-perceptions? Some argue that the answer is yes and that in-store music can align consumers with the brand or the marketplace and thus influence consumer behavior and perceptions (Bruner, 1990; Schmitt and Simonson, 1997).

The effect of in-store music can, to a large extent, be unconscious and influence consumer behavior through customers' emotions (Doyle, 2002). If in-store music affects consumer behavior and perceptions through emotions, music influences behavior through, or as, visceral factors (Loewenstein, 2000). Visceral factors include mood and emotions, and they influence consumer behavior, unconsciously, even if they do not change preferences (Loewenstein, 1996). Thus, if the congruence between in-store music and brand values influences the consumer's mood and emotions, then a change in these visceral factors can cause a change in consumer behavior.

We know that some companies use in-store music as a brand-building tool. Abercrombie & Fitch, for example, plays loud music to attract teenagers to their stores, claiming "Music first! Merchandise second!". They also attempt to play songs that are beginning to become popular among college students but are less well known among people in general (Morisson and Beverland, 2003). Starbucks is another example of a company that uses in-store music to strengthen its brand image. Starbucks often also plays songs from less known artists, with the result that these artists often become more well known among the general public (Yorkston, 2010).[1]

The effect of congruence between in-store music and brand image on consumer behavior might thus depend on the popularity of the songs. However, surprisingly few studies investigate the effects of congruence between in-store music and brand values (Beverland et al., 2006), and to the best of our knowledge, no previous study tests whether also using less well-known songs in the brand-fit playlists can help businesses to attract consumers and increase sales. Previous studies instead focus on investigating the effects of the mere presence of music (Garlin and Owen, 2006), tempo (Milliman, 1982, 1986; Oakes, 2003), different music styles (Areni and Kim, 1993; North et al., 1999; Wilson, 2003) and interaction effects with other sensory cues (Mattila och Wirtz, 2001).[2]

We conduct a field experiment with support from Soundtrack Your Brand, the exclusive provider of Spotify Business. Soundtrack Your Brand uses data from Spotify on listening frequencies and can therefore create playlists that

[1]The way in which Starbucks uses in-store music to affect its customers led the New York Times to call Timothy Jones, the manager who developed Starbucks' music strategy, "one of the quiet shapers of American culture "(New York Times, 2010).

[2]For overviews, see Hargreaves and North, 1997; Kellaris, 2008; and Yorkston, 2010.

differ in terms of song popularity. It also has developed a model for including songs that are in congruence with brand values, which means that we can test the effects of creating playlists that are supposed to reflect a company's brand values. It also means that we can test for the additional effect, if any, of including less well-known songs in the brand-fit playlist.

Furthermore, previous evidence on the effects of in-store music tends to be based on field experiments covering only one store (Areni and Kim, 1993; North et al., 1999), restaurant (Wilson, 2003), coffee shop (North and Hargreaves, 1996), or shopping mall (Yalcht and Spangenberg, 1990). The results might therefore be context-dependent (Radocy and Boyle, 1997) and driven by store-specific circumstances. Another limitation is that the experiment period is often very short. The combination of using only one store and a short time period makes causal inference less convincing.

To identify the causal effect of brand-fit in-store music on sales and customer perceptions, we conduct a field experiment in 16 restaurants of a market-leading restaurant chain in the city of Stockholm, Sweden. The experiment runs over 12 weeks, and we also gather sales data 4 weeks before and after the experiment period. This means that we have a total of 2240 daily observations. We divide the restaurants into a treatment and a control group, making it possible to use a difference-in-difference regression model (Card and Kruger, 1994). Difference-in-difference models are used extensively in empirical economic research because, under the assumption of parallel trends in the treatment variable in the absence of treatment, they enable causal inference from a comparison between treated and control restaurants.

Our experiment is based on a Latin Square Design with four different music treatments. The first treatment (our baseline) is a brand-fit (A) playlist that includes songs that reflect the brand values of the company, both well-known songs from Spotify's Top 1000 Sweden playlist and songs that are less well known.[3] The second treatment, brand-fit B, includes only those brand-fit songs that are on Spotify's Top 1000 Sweden playlist. Because both these playlists are included, we can investigate the effect of including less well-known songs in the brand-fit playlists. The third music treatment is based on Spotify's Top 1000 Sweden playlist but without any selection of songs that represent the brand of the restaurant chain. The no-brand-fit playlist makes it possible to test the effect of congruence between music and the brand. Finally, we include a situation in which there is no music played in the restaurant. This makes it possible to distinguish between the effect of the mere presence of music and the effect of music selection.

We directly gather data on revenues and quantity sold from the restaurant chain to investigate whether the presence and the choice of music influence the real purchasing decisions of customers. In addition, we gather data from a customer survey to investigate the effect of the music treatments on consumer perceptions. We also conduct a separate analysis of eight restaurants that have a drive-through option for their customers. Drive-through customers are not

[3]Note that songs with explicit lyrics are excluded from all playlists.

affected by the music treatments in the restaurants when they place their orders, making it possible to use drive-through customers as a control group when investigating the effect of music treatments on sales inside the restaurants.

We find that the congruence between in-store music and brand values is important for sales. Revenues are 9.1 percent higher when brand-fit music is played instead of no-brand-fit music, and the customer survey shows that the music affects customers' emotions. The survey responses also show that the customers often are unaware of the in-store music, suggesting that the effect of in-store music on consumer behavior is mostly on an unconscious level, which corresponds to the economic literature on mood and emotions (Loewenstein, 1996; Capra, 2004). We are unable to confirm whether including less well-known songs or the mere presence of music affects sales.

Section 2 describes the theoretical framework and presents the hypotheses that we want to test. Section 3 presents the experiment and the empirical model, while Section 4 presents the results on the effect of the music treatments on revenues and quantity sold. Section 5 presents the findings from the survey, mainly showing how the music influences the consumers' emotions. Section 6 summarizes our findings and discusses our conclusions.

2 Theoretical framework and hypotheses

Many purchasing decisions are made in the store, implying that consumer behavior might be influenced by perceptions and store atmospherics, e.g., visual cues, scents, and auditory stimuli (Bellenger et al., 1978; Babin and Dardin, 1995).

Most previous studies use the Mehrabian-Russell environmental psychology model (Mehrabian and Russell, 1974) to explain why store atmospherics affect consumers' experiences and behavior. It assumes that the environment sends out signals that affect the individual's emotional state, which lead consumers to either avoid or approach the environment. Thus, it is consumers' emotional responses to various atmospheric cues that influence their behavior. Three different emotional states can be influenced by store atmospherics according to the model, namely: (i) pleasure; (ii) arousal; and (iii) dominance. Pleasure refers to the degree to which consumers feel well and happy, while arousal relates to the extent to which consumers feel stimulated, excited, alert and active. The last emotional state is domination and captures the degree to which consumers feel influential and important.

The first study that used this model to explain the behavior of consumers in a marketplace was Donovan and Rossiter (1982). In contrast to Mehrabian and Russell (1974), they exclude dominance from the model because previous studies indicate that it lacks explanatory power (Russell and Pratt, 1980). Store atmospherics are thus assumed to influence consumers' level of enjoyment and excitement, i.e., their emotions, which are then assumed to influence their behavior. The emotional reactions are assumed to result in one of two contrasting behaviors, namely that consumers either approach or avoid the store environment. Consumers' emotional reactions might also be moderate, resulting in an increasing or decreasing desire to explore the marketplace. This thus suggests that music can affect consumers' emotions when the consumers are in the store or, as in this study, the restaurant. Exactly how the emotions then influence economic behavior is discussed more extensively in the field of behavioral economics. One of the main ideas is that emotions distort self-control (Loewenstein, 1996).

The concept of a dual self, with a hot and cold mode, dates back to Adam Smith in his *Theory of Moral Sentiment* (1759). Schelling (1960, 1978) and Buchanan (1975) highlight the importance of the dual self model for understanding self-control, which is incorporated into economic models (e.g., Thaler and Shefrin, 1981; Gul and Pesendorfer, 2004). The basic idea is that emotions (or other visceral factors[4]) can change behavior, even if they do not change preferences, by switching the dual self from cold to hot mode. Specifically, emotions can lead to more impulsive actions (Loewenstein, 1996). Especially interesting for this study, Loewenstein (1996) also notes that visceral factors can be induced by external influences in a predictive manner. Experiments often induce mood and emotions through recollection and imaging emotional events, as well as through audio (Capra, 2004).

A change in self-control can be part of the explanation for *how* emotional states affect consumer behavior, according to Mehrabian and Russell (1974). Supposing that consumers' purchasing decisions are dependent on their self-control, the marketplace environment can increase sales by influencing emotions. The predictability of the effect of external factors on self-control depends on the factor, and it might not be obvious how music would lower self-control. For example, Capra (2004) finds that a good mood induces more rational decisions. A good mood can still, however, influence consumer behavior by magnifying pre-decisional distortions (Meloy, 2000). An individual who enters a certain restaurant likely has a positive inclination toward the menu, and Meloy (2000) suggests that this positive approach is magnified by the good mood. This may well be how emotions make consumers approach the marketplace.

As Capra (2004) emphasizes, the effect of mood may vary substantially depending on the context. The context of this study, chain restaurants, makes self-control an important factor. Discussions of self-control often use restaurants as examples because many customers struggle to choose the more healthful alternative or to not have any dessert. Because music affects consumer emotions and mood (e.g., Doyle, 2002), the theory of the dual self and self-control provides an appealing framework for us to use when we study *how* music affects consumer behavior and sales in a restaurant.

By inducing the hot mode of the dual self, the resulting increased intensity in emotions leads to more impulsivity in decision making. We can consider this through the distinction between actual and desired utility (Loewenstein, 1996), which is in line with the distinction between predicted and decision utility made by Kahneman, Wakker and Sarin (1997). Increasing intensity in emotions

[4]Emotions, mood and drive states (e.g., hunger, pain and anger).

increases the desired value (decision utility), which leads the individual to act impulsively. Therefore, if the congruence between music and the brand increases emotional intensity, as Mehrabian and Russell (1974) suggest, the studies on self-control can help explain how this translates into increased consumption.

Another reason that brand-fit in-store music can increase sales is that it reminds the consumer of the good and in this way leads to an increase in consumption. According to Bernheim and Rangel (2004), such environmental cues can have substantial impacts on behavior. In-store music can thus directly influence the consumer, just as a visceral factor does. In this line of reasoning, the congruence between the music and the brand increases consumption in a more direct, and more predictable, way.

We now use these findings to hypothesize about how music can increase sales in a restaurant. It is essential that the music reflects the brand, and it affects consumption in two ways. One way is through increased intensity in emotions and that this causes the consumer to approach the environment and thus consume more. The second way is more direct and simply reminds the consumer of the specific good. To achieve an optimal effect, we therefore need the music to reflect the brand but nothing else. If the music also reminds the consumers of something else, it does not work in a direct way. We therefore expect that the best music for sales is music that, while congruent with the brand, does not remind consumers of other experiences. Our first hypothesis is therefore as follows:

H1. *Among brand-fit music, less well-known songs increase sales.*

This is simply because consumers do not relate the less well-known music to anything else, while they might already relate well-known songs to different experiences. Familiar songs also reduce the perceived complexity of the music, and consumers might react differently when the music is more predictable (Bruner, 1990).

Our second hypothesis does not distinguish between the direct and indirect effects but simply focuses on establishing the congruence between music and the brand:

H2. *Brand-fit music increases sales and influences the customers' emotions.*

This hypothesis, however, also states that congruence affects the customers' emotions. Because we do not at present distinguish between the direct and indirect (through emotions) effects, it is important to test whether emotions actually depend on the music to determine whether an indirect effect exists.

To test the effect of congruence between music and the brand in comparison to a music-neutral situation, our third and final hypothesis is as follows:

H3. *The presence of music increases sales and customers' positive emotions.*

This neutral situation makes it possible to compare the effect of the mere presence of music with the effect of aligning the music with the rest of the restaurant environment.

3 The Experiment

We conduct a field experiment in one of the market leading restaurant chains in Sweden to study the effects of in-store music choice on sales and customer perceptions. Our experiment is conducted with support from Soundtrack Your Brand (SYB), the exclusive provider of Spotify Business.

3.1 Music Treatments

SYB constructs the playlists used in the experiment, using information on music preferences from Spotify and the company's knowledge and experience concerning what type of music reflects the brand of the restaurant chain. To decide which songs reflect the brand, the restaurant chain provides SYB with value words that it wishes to be associated with the company. Then, the music experts at SYB select which songs fit the value words. We can thus test a real business solution that aims to create brand-fit playlists. The playlists are controlled from the headquarters of SYB, and the local staff in the restaurants therefore cannot influence the music played during the experiment and control periods without contacting the headquarters of SYB. This reduces non-compliance, which is a common problem in field experiments (Krueger, 1999; Ortmann, 2005; Duflo et al., 2007).

To estimate the effect of music in the restaurants, we use the following four music schemes:

1. *Brand-fit A* music (the baseline), a playlist that includes 100 well-known songs from Spotify's Top 1000 Sweden playlist and 260 less well-known songs. All songs are chosen because they have a sound that fits the brand of the restaurant chain.

2. *Brand-fit B* music, a playlist of 360 well-known songs from Spotify's Top 1000 Sweden that have a sound that fits the brand of the restaurant chain.

3. *No brand-fit* music, a playlist that includes 360 well-known songs from Spotify's Top 1000 Sweden playlist.

4. *No music*, in which the music is simply off.

The *brand-fit A* treatment is the baseline in all estimations because it is the playlist that is usually played in all restaurants that belong to the restaurant chain. The *brand-fit A* playlist only includes songs that reflect the brand image. Here, this means that the playlist creator at SYB excludes songs that sound too excluding, low-key, traditional, technological or serious. The final playlist includes 100 of the most popular songs in Sweden based on information from

Spotify's Top 1000 Sweden for both genders and all ages, as well as 260 additional less well-known songs that sound similar. Popular and less well-known songs are thus distinguished using data from Spotify on listening frequencies.

The second music treatment is the *brand-fit B* playlist. Only songs that have a sound that fits the brand of the company are included in this playlist. The difference from the *brand-fit A* playlist is that only well-known songs are included. This makes it possible for us to test the effect of including less well-known songs that still reflect the brand values of the company (H1).

The third music treatment is the *no brand-fit* playlist, where there is no selection of songs in terms of whether they fit the brand. This tests whether selecting music that fits the brand increases sales and influences customers' emotions (H2).

The last music treatment is the *no music* situation, where the music is turned off. We include this situation in the experiment because we want to distinguish between the effect of playing music at all (H3) and the effect of music choice (H1 and H2). We are thus able to investigate whether it is the presence of music or the choice of music that influences sales and consumer emotions.

3.2 Experimental Design

To measure the effect of these different music treatments, we conduct a field experiment in 16 restaurants belonging to one of the market-leading restaurant chains, of which 8 restaurants are exposed to different music treatments and 8 belong to the control group. All restaurants are located in the Stockholm metropolitan area to reduce geographical heterogeneity. The restaurants are selected into treatment and control groups by the management of the restaurant chain. We control for the potential effects from this non-random selection by using restaurant-specific fixed effects (see Section 3.4).

We apply a Latin Square Design (Hinkelmann and Kempthorne, 2008) to study how the different music treatments affect sales in the restaurants. Specifically, the eight experiment restaurants are randomly assigned into four experiment groups (EG1-EG4) based on the different music treatments, with two restaurants in each group. The experiment period is then divided into four subperiods (three weeks each) such that each of the four experiment groups are exposed to each of the four music treatments at different times. By assigning different treatments to different experiment groups in each time period, this setup generates variation across groups within each time period and variation over time within each group. We also observe both control and experiment restaurants for four weeks before (pre-period) and four weeks after the experiment period (post-period), when they are all exposed to the baseline treatment, i.e., the *brand-fit A* playlist. Table 1 visualizes the experimental design.

Table 1: Latin Square Design

	Pre	Experiment Period				Post
		Weeks				
Group	1-4	5-7	8-10	11-13	14-16	17-20
1	T1	T4	T3	T2	T1	T1
2	T1	T3	T2	T1	T4	T1
3	T1	T2	T1	T4	T3	T1
4	T1	T1	T4	T3	T2	T1
Control	T1	T1	T1	T1	T1	T1

The different treatments are denoted T1 = *brand-fit A* (baseline), T2 = *brand-fit B*, T3 = *no brand-fit*, and T4 = *no music*.

The Latin Square Design minimizes the risk that some other, unobserved, factor drives the results. If some odd event were to affect the estimates of the music treatments, the timing of this event must coincide with the different specific experiment periods for the experiment groups of interest. Such an event is highly unlikely. Thus, using a Latin Square Design experiment in combination with a difference-in-difference regression analysis, after confirming the parallel trend assumption, makes it likely that any observed differences between the treatments are in fact causal effects of the treatments.

3.3 Implementation and partial compliance

When conducting a field experiment, one often encounters instances in which treated units do not comply with the setup of the experiment (Krueger, 1999; Ortmann, 2005; Duflo et al., 2007). In this experiment, two restaurants simply refused the *no music* treatment and instructed SYB to play *brand-fit A* instead, creating a potential endogeneity problem. With partial compliance, the appropriate estimation method is to use the initial random assignment as an instrumental variable (IV) for observed treatment (Duflo et al., 2007; Angrist and Imbens, 1994). Under the assumptions that (i) potential outcomes are independent of the instrumental variable (independence) and (ii) that initial assignment makes the restaurants weakly more or weakly less likely to receive the treatment (monotonicity), the IV estimates the *Local Average Treatment Effect* (LATE) (Angrist and Imbens, 1994; 1995). Because initial assignment was random, it is convincing that potential outcomes are independent of the assignment. It is also very unlikely that the assignment would make any restaurant *less* likely to receive the assigned treatment. Therefore, the IV mitigates the selection bias from partial compliance and estimates LATE, which is the effect of treatment on the restaurants that comply with the initial assignment.

In addition to selection bias, non-compliance creates variation in treatments within the "squares" of the experimental setup. When two restaurants within a group and experiment period do not follow the same music scheme, treating them as a group serves no purpose. To account for the variation within squares, we use restaurant-specific fixed effects, i.e., within-estimation, which is common practice when comparing groups that are different in some time-invariant aspects. In our experimental setup, it means dissolving the experiment groups and experiment periods in the estimations and controlling for differences across restaurants and weeks instead of differences between groups and periods. The problem of variation within squares is thereby solved, and the fixed effects strategy actually controls for more heterogeneity than using the experimental design in Table 1 without the inclusion of restaurant- and week-specific fixed effects.

Finally, note that because SYB manages the music centrally, we are certain of what music the restaurants actually played during the experiment.

3.4 Empirical model

To test whether the presence and choice of music can influence business performance, we estimate the following equation:

$$\begin{aligned} lnSALES_{it} &= \alpha + \beta_2 T2_{it} + \beta_3 T3_{it} + \beta_4 T4_{it} + \\ &\gamma_1(WEEK) + \gamma_2(WEEKDAY) + \gamma_3 R_i + \epsilon_{it} \end{aligned} \tag{1}$$

where the dependent variable in our main estimations ($lnSALES$) is the natural logarithm of revenues. In some additional estimations, we also estimate equation (1) using the natural logarithm of the quantity sold[5] as the dependent variable. The main advantage of taking logarithms before estimation is that it enables one to interpret the estimated effects in terms of percentage changes.

The β coefficients estimate the effect of the music treatments relative to the *brand-fit A* treatment (the baseline). The variables $T2$, $T3$ and $T4$ are indicator variables, taking value 1 if the observation is from the denoted treatment and zero otherwise. Each of these indicator variables is instrumented with the corresponding variable from the initial random assignment to estimate LATE.

($WEEK$) is a vector of week indicator variables, and ($WEEKDAY$) is a vector of day-of-the-week indicator variables. We include the latter set of covariates to control for both week- and weekday-specific heterogeneity in sales.

One potential problem with our experimental design is that the restaurants were not randomly assigned to treatment and control groups, which implies that the control group might not be the perfect counterfactual for the experiment restaurants. The empirical model thus also contains a restaurant-specific fixed effect R_i, which is included to account for the non-randomness of the selection into treatment and control stores. If the differences across the restaurants are time-invariant and do not correlate with the different treatments, then this suffices to address the non-randomness of the treatment and control group selection. Finally, ϵ is an error term with expected value zero and constant variance.

In addition to the restaurant-specific fixed effects model, we also apply an alternative strategy to address the potential problem of the non-random

[5]For example, a value meal that contains one hamburger, one box of French fries and one soda is reported as one unit of each item, totaling three units.

selection into the treatment and control groups. This strategy exploits a specific feature of eight of the restaurants (of which five belong to the treatment group), namely that they also provide a drive-through option for the customers, and sales data are collected separately for the drive-through. The drive-through customers are obviously not affected by the music played in the restaurant, which means that we can compare sales at the drive-through window (not affected by the music) with sales at the front counter (affected by the music treatments) to estimate the effect of the music.

In this case, there also might be time-invariant differences across the restaurants, for example in the level of average sales, and we again control for such differences using fixed effects estimation. The fixed effect difference-in-difference regression model used in the estimations thus compares changes in sales (revenues and quantity sold) inside the restaurants when the different music treatments are implemented to changes in sales at the drive-through window in the same restaurants during the same periods of time. Again, descriptive statistics verify that the sales data show parallel trends in sales at the front counter and at the drive-through window in the pre-experiment period. Thus, the change in sales when the different music treatments are implemented are interpreted as causal effects of the different music treatments.

3.5 Data

The restaurants report daily sales data through two different measures: revenues in SEK[6] and quantity sold. Although revenues and quantity sold are highly correlated, they are not identical. For example, one possible change in consumer would be that customers select more expensive meals, increasing revenues but holding quantity sold constant. It is also possible that customers buy additional smaller items, such as desserts or drinks, which increases both the quantity sold and revenues. Analyzing both of these outcome variables thus helps us to draw conclusions regarding what is happening in the restaurants when the music changes.

The experiment period starts on February 8, 2016, and lasts until May 1, 2016. We also use data from a pre-experiment period of four weeks, starting January 11, and a post-experiment period of four weeks, ending May 29. Because the experiment period lasts for 12 weeks, the total number of weeks is 20 and the number of days is 140. Observing 16 restaurants over 140 days yields a total of 2240 daily observations.

Table 2 provides descriptive statistics for revenues and quantity, disregarding drive-through sales.[7] Because the restaurant chain wants to keep its sales information confidential, we are not allowed to present average revenues or quantity. Therefore, in the descriptive statistics, the means have been converted into indexes with the mean of the experiment group in the pre-experiment period as

[6] 1 SEK = 0.12 USD, February 8, 2016.

[7] In our main specification in which we compare sales in the experiment restaurants to sales in the control restaurants during the experiment period, we exclude the sales from the drive-through as these cannot be affected by the music treatments.

Table 2: Descriptive statistics

Group	# of obs.	Mean Daily Revenues (in SEK) Full Period	Pre	Experiment	Post
Control	1 113	105.92	95.74	106.56	112.80
		(55.24)	(52.53)	(52.35)	(63.35)
Experiment	1 119	107.48	100.00	106.83	116.84
		(37.19)	(37.24)	(36.81)	(35.61)
All	2 232	106.70	97.87	106.70	114.85
		(47.06)	(45.53)	(45.24)	(51.19)

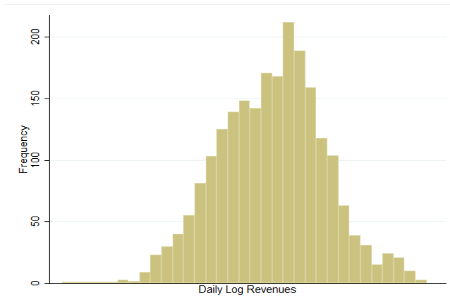
Group	# of obs.	Mean Daily Quantity Sold Full Period	Pre	Experiment	Post
Control	1 113	102.04	90.23	102.96	109.68
		(56.48)	(51.85)	(53.89)	(65.05)
Experiment	1 119	106.94	100.00	106.81	114.03
		(44.35)	(45.42)	(44.49)	(40.56)
All	2 232	104.49	95.09	104.91	111.85
		(50.79)	(48.94)	(49.44)	(54.03)

a base equaling 100. All means are thus presented in comparison to this base value. Henceforth, we do not present any information that would make it possible to infer average sales levels in the tables or graphs. For the same reason, the constants are not presented in the tables that report our regression results.

Because some restaurants were temporarily closed (e.g., due to maintenance work), there are 2232 observations available for analysis. The missing days are all in the post-experiment period.

Figure 1 plots the distribution of daily revenues (in logs). As shown, the variable is almost normally distributed, with some outliers at the top of the distribution. We also ensure that outliers do not drive our results by excluding observations outside two standard deviations from the mean in some additional regressions estimated as a robustness check.

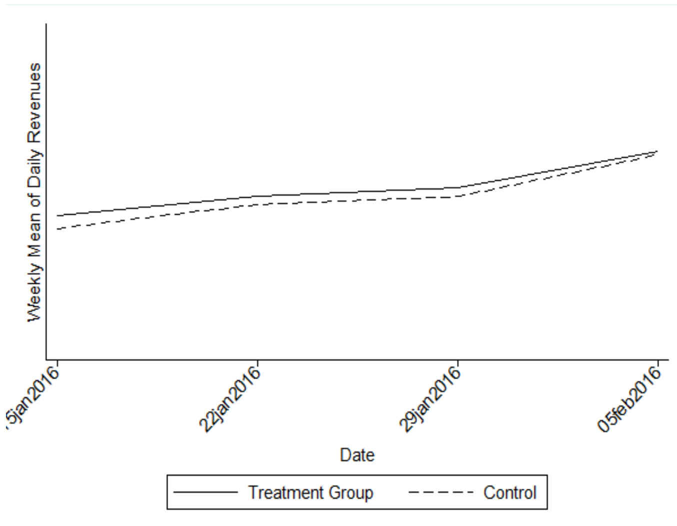
Figure 1: Distribution of Daily Log Revenues



For a more specific comparison between the intervention and control groups, Figure 6 (see appendix) plots daily revenues over all periods. The sales data are plotted on a weekly basis, displaying peaks on weekends, and an increase over time is visible.[8] Because we control for general trends within the week in the estimations using weekday-specific indicator variables, we are interested in how the weekly averages of daily sales evolve over time, which is difficult to discern from Figure 6. Figure 2 thus instead plots weekly averages of daily revenues over time in the pre-experiment period. The important feature of the graph is that the trends are parallel in the pre-experiment period, indicating that the treated restaurants would likely have experienced the same sales pattern as the control group restaurants in absence of treatment. The differences observed in the sales during treatment can therefore be referred to as causal effects of the treatment.

[8] We test for a unit root with the ADF test on both weekly and daily mean revenues, and we reject a unit root in both series.

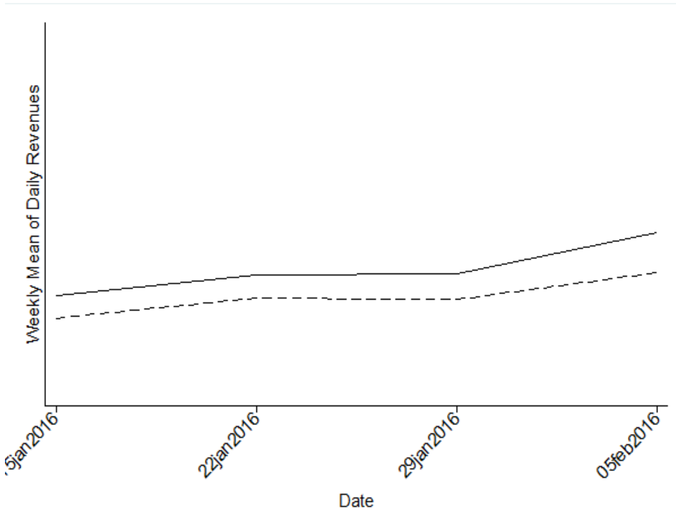
Figure 2: Revenues - Parallel Trend



Regarding the alternative strategy, i.e., using drive-through customers as our control group, Table 9 (see appendix) provides descriptive statistics corresponding to Table 2, and the descriptive statistics show that sales are generally lower at the drive-through than inside the restaurants. To ensure that the general difference in the level of sales between what is now the control group (drive-through) and the experiment group (front counter) do not drive the results, we control for this difference in the regression specification using a drive-through-specific fixed effect.

Figure 3 plots the average revenues for each week during the pre-experiment period in the drive-through setting. The front counter revenues are plotted as a line and the drive-through revenues as a dashed line. The interpretation is the same as that of Figure 2. If the trends are parallel in the pre-experiment period, the drive-through sales are expected to be valid counterfactuals for the change in front counter sales. Thus, this graphical analysis supports the causal inference interpretation of the fixed effects difference-in-difference estimations.

Figure 3: Drive-Through - Parallel Trends



The strength of the drive-through setup is the convincing counterfactual assumption. Each restaurant has its own control group, i.e., drive-through sales. The trends of the front counter sales and the drive-through sales are very similar, as Figure 3 shows.

Both revenues and quantity are reported per product item and part of the day. Our main analysis focuses on total daily sales, but we also test whether the results are different for different product items and parts of the day. This is useful to gain specific insights into which type of sales responds to music choice. The data cover over 500 different product items, and we use this information to also investigate whether the effects differ for hamburgers, fries, soda, side-dishes, hot drinks, smoothies and milkshakes, and desserts. The different parts of the day used in the analysis are breakfast, lunch, snack, dinner, evening and late night.[9] These divisions help us understand whether the music affects some customers more than others and if the effect is greater for certain items.

4 Main results

4.1 Effects on sales

Table 3 presents the main results from the experiment, where the coefficients are interpreted as the percentage change in sales when switching from brand-

[9]Part of the day – hours: Breakfast – 06-10, lunch – 10-14, snack – 14-17, dinner – 17-19, evening – 19-23, late night – 23-06.

fit A to the respective music treatment.[10] Columns (1) and (2) present the estimates based on Equation 1 for revenues and quantity, respectively. Estimations presented in columns (3) and (4) exclude outliers, i.e., observations of net sales and units sold that are more than two standard deviations from the mean, to investigate whether the results are driven by a few influential observations. Columns (5) and (6) present the drive-through estimates.

Table 3: Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Main Analysis		Excluding Outliers		Drive-Through	
Treatment	Revenues	Quantity	Revenues	Quantity	Revenues	Quantity
Brand-Fit B	$-.036$	$-.030$	$-.038$	$-.021$	$-.094^{*}$	$-.079^{*}$
	(.032)	(.032)	(.035)	(.026)	(.050)	(.042)
No Brand-Fit	$-.091^{**}$	$-.090^{**}$	$-.094^{***}$	$-.080^{**}$	$-.114^{***}$	$-.107^{***}$
	(.038)	(.040)	(.036)	(.034)	(.035)	(.033)
No Music	$-.048$	$-.048$	$-.061$	$-.041$	$-.093^{*}$	$-.088$
	(.038)	(.035)	(.042)	(.027)	(.052)	(.063)
Observations	2,232	2,232	2,130	2,143	2,236	2,236
R^2 within	.430	.451	.381	.451	.454	.432
# restaurants	16	16	16	16	8	8
Excluded obs.	-	-	2 Std	2 Std	-	-

Note: Bootstrapped standard errors in parentheses (number of replications: 1000. Grouped on restaurant). Revenues and quantity sold are in logarithmic values. All estimations include fixed effects for restaurant, week and day of the week. P-values (inferred from degrees-of-freedom adjusted t-values): *** < 0.01, ** < 0.05, * < 0.1.

The estimates in columns (1) and (2) show that the music treatment that has a statistically significant negative effect on revenues when the *no brand-fit* condition is compared to the *brand-fit A* playlist. The coefficient for *no brand-fit* indicates that revenues (quantity) are 9.1 (9.0) percent lower with *no brand-fit* than with *brand-fit A*. This result confirms the hypothesis that music that fits the brand increases sales (H2). Excluding outliers does not seem to substantially impact the estimated effect, indicating that our findings are robust. The coefficients for *brand-fit B* and *no music* are not significant in the main specification, leading us to conclude that they have no significant effect on sales compared to *brand-fit A*. However, they are significant in the drive-through specification, which will be discussed soon. If we were to only consider the findings in the main specification, we would conclude that switching from *brand-fit A* to *brand-fit B* or *no music* will not significantly impact sales. *Brand-fit A* and *B* are very similar, and thus we expect it to be difficult to confirm

[10] Specifically, for a percentage change interpretation, the coefficients should be recalculated as $100 \times [e^{\beta} - 1]$. This is however rarely done when coefficients are small because $e^{\beta} \approx 1 + \beta$ when $\beta \approx 0$.

the hypothesis that less well-known songs are better for sales (H1). However, the lack of a significant difference between *brand-fit A* and *no music* is not as expected. This means that we cannot confirm the hypothesis that the mere presence of music increases sales. In some sense, this means that if the only music available is *no brand-fit*, it might be better to not play any music at all. However, none of the coefficients are statistically significantly different from one another [11], and thus, the only difference on which we can make inference is that between *brand-fit A* and *no brand-fit*.

The additional analysis using drive-through sales as a control group strengthens our results in the main analysis and suggests even stronger effects. The estimated coefficients from the drive-through analysis, presented in columns (5) and (6), translates into a 11.4 (10.7) percent difference in revenues (quantity) between *brand-fit A* and *no brand-fit*. This setting also suggests that *brand-fit B* and *no music* both significantly decrease sales compared to *brand-fit A*. This indicates that all three hypotheses are supported. The differences in revenues and quantity between *brand-fit A* and *brand-fit B* (*no music*) are 9.4 and 7.9 (9.3 and 8.8) percent, respectively. Here, the only coefficient that is not statistically significant is the *no music* coefficient on quantity.

The reason that the estimated effects are much stronger in the drive-through setting is not clear. First, it could simply be the case that customers at drive-through restaurants, i.e., road restaurants, are more receptive to music. If so, these results are very important for restaurant chains that locate near roads. However, the main lesson is the same, namely that the best music choice is *brand-fit A*. Second, the effect might be the same, but the drive-through sales are simply a better comparison group than the front counter sales in the restaurants assigned to the control group in the main analysis. If this is true, then we should take the drive-through results more seriously than the main results. However, we cannot know what the reason is. We therefore retain the main setup in the further analysis of the effect on sales. For a more detailed insight into how music affects consumer behavior, we estimate separate regressions for different product categories. Our results are presented in Table 4, where columns (1)-(3) present estimates for the product categories that are default in most value meals (hamburgers, fries, soda); column (4) presents estimates for all side dishes, including French fries; column (5) presents estimates on hot drinks (mainly coffee); column (6) presents estimates on smoothies, iced coffee and milkshakes (henceforth defined as smoothies); and column (7) presents estimates on dessert items.

These results tell us two things. First, playing music that fits the brand increases all item sales. The *no brand-fit* treatment has a negative significant effect across the value meal items and hot drinks, suggesting that this treatment reduces standard item sales. However, the *no brand-fit* also decreases all non-value meal items. Second, and especially interesting, the effect is much greater for sides, smoothies and desserts than for value meal items. This implies that playing music that fits the brand image makes people more likely to buy ad-

[11] We use the Wald test to test the differences between coefficients.

Table 4: Item by item (Log Revenues)

Treatment	(1) Hamburgers	(2) Fries	(3) Soda	(4) Sides	(5) Hot Drinks	(6) Shakes & smoothies	(7) Desserts
Brand-Fit B	−.034	−.040	−.058	−.019	−.063*	−.096	−.057
	(.036)	(.035)	(.038)	(.033)	(.037)	(.072)	(.058)
No Brand-Fit	−.086**	−.082	−.076**	−.111**	−.067*	−.150***	−.156**
	(.039)	(.051)	(.059)	(.019)	(.033)	(.033)	(.035)
No Music	−.052	−.051	−.059	−.019	−.033	−.033	.003
	(.044)	(.036)	(.040)	(.042)	(.048)	(.115)	(.041)
Observations	2,232	2,232	2,232	2,232	2,232	2,232	2,232
R^2 within	.379	.362	.437	.478	.182	.557	.402

Note: Bootstrapped standard errors in parentheses (number of replications: 1000. Grouped on restaurant). Revenues and quantity sold are in logarithmic values. All estimations include fixed effects for restaurant, week and day of the week. P-values (inferred from degrees-of-freedom adjusted t-values): *** < 0.01, ** $<$ 0.05, * < 0.1.

ditional items than if the restaurant is playing random popular music, i.e., *no brand-fit*.

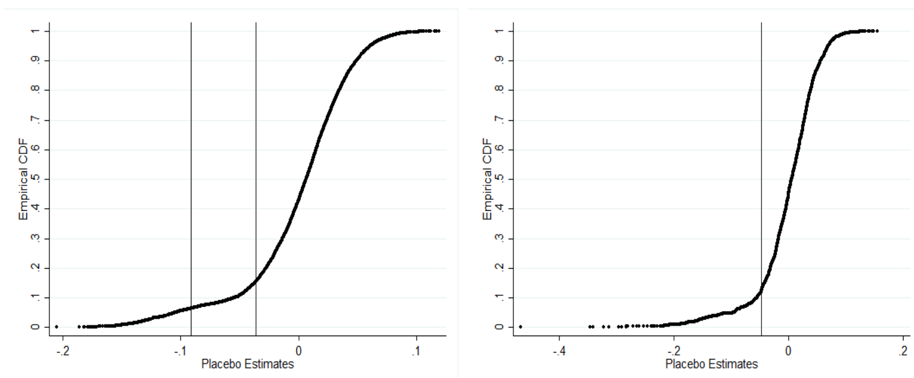
We also estimate the effects of music conditions on revenues separately for each part of the day (Table 10, appendix). The negative effect of the no brand-fit treatment on revenues is only significant during daytime hours (lunch, snack and dinner). The estimated effects are insignificant during breakfast and late night. The reason for this heterogeneity is likely that the meals and the customers are very different during these time intervals. During breakfast, the menu is very different and desserts are unusual, and customers are probably more influenced by other factors during late night. It could be that the significance during only daytime hours is because of increased statistical power due to more frequent sales. However, it is not the standard errors that drive the significance, it is the size of the coefficients. The standard errors during daytime hours are actually larger than the standard errors during breakfast and dinner. Because the difference is in the magnitude of the coefficient, we consider our conclusions to be well founded.

4.2 Statistical inference

As noted in the table notes, we use bootstrapping to derive reliable standard errors. The data we have are problematic because the observations are not independent within each restaurant. We would like to cluster the standard errors to allow for the intra-restaurant correlation structure, but this is not a reliable approach when the number of clusters is only 16 (Cameron and Miller, 2008). We turn to a bootstrap technique that treats the restaurants as clusters, and we then calculate p-values based on a t-distribution because the number of clusters is small.

We also perform non-parametric permutation tests, an approach that is very similar to bootstrapping. We randomize placebo treatments such that eight restaurants are randomly chosen for "treatment". Then, for each of these restaurants, we randomize which experiment period to denote as the period for "treatment". We then estimate Equation (1) but with this randomized placebo treatment. We repeat this procedure 10,000 times and then plot the empirical CDF of these estimates. If our estimates are not random, they should lie in the tail of the distribution. We can use the empirical distribution of estimates to obtain reliable p-values for our true treatments. Because this procedure does not make any parametric assumptions and treatment is randomized at the restaurant level, we allow for the intra-restaurant correlation without biased inference.

Figure 4: Distribution of Placebo Estimates



(a) No Brand-Fit (L) and Brand-Fit B (R)

(b) No Music

Figure 4 plots the empirical CDFs, one for *brand-fit B* and *no brand-fit* and one for *no music*. The procedure is a somewhat different for *no music* to account for the partial compliance. The vertical lines denote the treatment coefficients in Table 3, column (1). The p-values derived by the empirical CDFs are 0.156 (*brand-fit B*), 0.064 (*no brand-fit*) and 0.119 (*no music*). At least, it is evident that the estimate of *no brand-fit* is unlikely to be random.

5 Effect on emotions

5.1 The survey

Customers were asked when they left the restaurant if they would be willing to participate in an anonymous survey that required 2-3 minutes. They were thus outside the restaurant and could no longer hear the music. The customers were not told what the purpose of the survey was and had no reason to suspect that it concerned the music.[12] Respondents were at least 16 years old and answered the questionnaire on iPads. The survey was collected at all experiment restaurants at six points in time during the experiment. The survey staff wore black clothing.

The survey consists of 10 main questions with sub-questions. The full survey can be obtained upon request. The literature suggests that emotions play an important role in the effect of music on consumer behavior. To test this, we analyze how the different music treatments affect the customers' emotions, evaluated in question 2, *emotion*.[13]

In the 10 sub-questions in question 2, *emotion*, the respondent grades the intensity with which she feels each emotion in that moment. The scale is 1 to 7, where 1 is not true at all while 7 is completely true. The emotions are content (nöjd), happy (glad), hopeful (optimistisk), relaxed (avslappnad), satisfied (tillfredställd), frenetic (full av energi), aroused (lycklig), stimulated (inspirerad), alert (pigg) and excited (motiverad).

5.2 Empirical set-up

We regress the emotion variables on the music treatment variables and a set of control variables that are constructed using the responses to other questions from the survey. A simple comparison of means across the different treatments is not an appropriate strategy, mainly because respondents are not randomly sampled. Because participation in the survey was optional, there are good reasons to believe that sample selection would inflict bias on a simple difference in means analysis. However, we can use many of the questions to control for differences between subjects that correlate with their responses, as well as factors determining participation.

The equation we estimate is as follows:

$$\begin{aligned}(ANSWER_{cit}) = \alpha + \beta_2 T2_{it} + \beta_3 T3_{it} + \beta_4 T4_{it} \\ +\theta_1(CONTROLS_{cit}) + \epsilon_{cit}\end{aligned} \tag{2}$$

The outcome variables are the different emotions, reported by customer c in restaurant i at time t, and the treatment variables are the same as in the sales data analysis. The vector of control variables consists of other responses that might also determine the outcome responses. We are not interested in the estimated coefficients for the control variables. The interesting coefficients are, again, those for the different music treatments.

Because the responses are given in integers (1 to 7), and hence deliver a discontinuous and bounded outcome variable, linear estimation is inappropriate (Cox, 1958). We therefore use ordinal logit regression to estimate the effect

[12] They might have realized that the survey concerned music when they reached question 4. However, at that point, the questions about emotions had already been answered.

[13] Many of the questions are standard questions in the restaurant chain's regular evaluations. The results for questions other than those referring to mood are available upon request.

of music treatments on customers' emotions. We re-calculate the coefficients and present them as "odds ratios". The odds ratio indicates how much greater the odds are for the variable to take a higher value on the scale because of the treatment. Because it is a ratio, 1 means that the treatment makes no difference, in comparison to the baseline treatment. If the ratio is above 1, the effect is positive. Analogously, if the ratio is below 1, the treatment effect is negative.

5.3 Variables and data

The set of controls we use consists of the questions 1.1 (c) "I received what I ordered", 1.1 (d) "My food was served hot and fresh", 1.1 (f) "A clean restaurant", 5 "How often do you visit a [brand name] restaurant?", 6.1 "Which meal did you order today?", 7 "I visited [brand name] today accompanied by?", 8 "How far away from this [brand name] restaurant do you live?", 9 "Age" and 10 "Gender". These factors might affect the likelihood of participation and the answers we study. Controlling for them likely gives more precision to our estimates. The reason for using only some of the questions in question set 1, "today's visit", is that these are more objective. Questions that are less objective are more likely influenced by the music. Controlling for variables that are also affected by the treatment variables introduces bias into the treatment estimates. Therefore, we do not control for these.

Table 5 presents mean values for the control questions. A brief scan of the control variables across treatments suggests balance across the treatment groups, which is important for causal inference. This means that customers are similar in the different treatments, and thus, differences in outcomes depend on the treatment. However, we still use the control strategy for precision.

Table 6 presents descriptive statistics for the emotion questions, 2(a)-(j). We observe a pattern across treatments. Mean values are higher during *brand-fit A* and *brand-fit B* than during *no music* and *no brand-fit* for almost all emotions. The only exception is question 2(a). This very simple comparison suggests that brand-fitted music makes the customers feel better. We also specify an aggregated variable, which simply sums the values for these different emotions into a single measure. This creates more variation across observations and tells a more general story about how music affects mood and emotions overall. The mean of the variable follows the same pattern across treatments.

Table 5: Treatment balance

QUESTION	All	Brand-Fit A	Brand-Fit B	No Brand-Fit	No Music
1(c) *I got what*	6.608	6.663	6.550	6.518	6.705
I ordered	(1.076)	(.991)	(1.189)	(1.210)	(.856)
1(d) *My food was*	6.376	6.352	6.392	6.461	6.299
served hot and fresh	(1.076)	(1.041)	(1.053)	(1.028)	(1.179)
1(f) *The restaurant*	5.586	5.478	5.608	5.644	5.619
was clean	(1.428)	(1.395)	(1.451)	(1.475)	(1.386)
5 *How often do you*	3.977	3.978	4.050	3.856	4.016
visit this chain?	(1.435)	(1.438)	(1.452)	(1.413)	(1.432)
Distance (in min)	35.943	39.324	48.799	28.919	25.142
	(102.521)	(88.152)	(152.125)	(76.449)	(61.568)
Age	39.575	41.659	35.871	39.266	41.792
	(16.632)	(17.409)	(17.455)	(15.540)	(15.146)
Male	.596	.577	.554	.646	.613
# of obs.	2101	534	558	508	501

Note: Because questions 6.1 and 7 are unordered indicator variables, mean statistics are not presented.

Table 6: Descriptives - Emotions

QUESTION	All	Brand-Fit A	Brand-Fit B	No Brand-Fit	No Music
2(a) *Satisfied*	6.090	6.099	6.181	5.970	6.102
	(1.108)	(1.070)	(1.069)	(1.186)	(1.101)
2(b) *Delighted*	6.032	6.142	6.161	5.898	5.908
	(1.165)	(1.078)	(1.121)	(1.241)	(1.200)
2(c) *Optimistic*	5.753	5.800	5.855	5.640	5.707
	(1.313)	(1.234)	(1.331)	(1.370)	(1.308)
2(d) *Relaxed*	5.725	5.704	5.853	5.644	5.687
	(1.409)	(1.369)	(1.408)	(1.477)	(1.374)
2(e) *Pleased*	5.837	5.861	5.986	5.738	5.747
	(1.280)	(1.178)	(1.241)	(1.383)	(1.306)
2(f) *Full of energy*	5.230	5.234	5.403	5.157	5.106
	(1.540)	(1.469)	(1.538)	(1.592)	(1.548)
2(g) *Happy*	5.762	5.895	5.853	5.656	5.627
	(1.347)	(1.230)	(1.324)	(1.426)	(1.392)
2(h) *Inspired*	5.171	5.294	5.351	4.984	5.030
	(1.570)	(1.411)	(1.613)	(1.673)	(1.546)
2(i) *Alert*	5.137	5.191	5.296	5.002	5.038
	(1.591)	(1.479)	(1.604)	(1.666)	(1.600)
2(j) *Motivated*	5.401	5.436	5.584	5.283	5.278
	(1.474)	(1.358)	(1.476)	(1.555)	(1.486)
2 *Aggregated*	57.139	56.657	57.523	54.972	55.228
	(11.166)	(10.112)	(11.181)	(11.793)	(11.389)
# of obs.	2101	534	558	508	501

5.4 Survey results

Table 7 presents the estimated effects of music on consumer emotion intensity. The effect on the aggregate measure is presented in column (11). We use OLS for the aggregate measure because the range of integers is much wider when the responses are summed. The coefficient for the *no music* treatment is statistically significant (at least at the 10 percent significance level) for 4 of 10 moods and for the aggregate measure. The effect of *no brand-fit* is even stronger. The *no brand-fit* coefficient is statistically significant for 6 of 10 emotions and for the aggregate measure. They are of equal magnitude, suggesting that they decrease emotion intensity by a similar amount. The *brand-fit B* odds ratio, however, is above one for the emotions for which its coefficient is significant. This suggests that *brand-fit B* increases these emotion intensities compared to *brand-fit A*. The *brand-fit B* coefficients are significant for different feelings than are *no music* and *no brand-fit*, and the *brand-fit B* treatment is not statistically significant for the aggregate variable.

Table 7: Estimates - Emotions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treatment	Satisfied	Delighted	Optimistic	Relaxed	Pleased	Full of Energy	Happy	Inspired	Alert	Motivated	Sum
Brand-fit B	1.196	1.012	1.109	1.317**	1.230*	1.248**	.900	1.192	1.221*	1.319**†	.554
	(0.142)	(0.122)	(0.128)	(0.152)	(0.143)	(0.139)	(0.104)	(0.132)	(0.135)	(0.149)	(0.568)
No Brand-Fit	.755**††	.601***†††	.755**††	.930	.828††	.889	.685***†††	.695***†††	.824*††	.834††	−1.966***
	(0.091)	(0.072)	(0.088)	(0.107)	(0.097)	(0.100)	(0.079)	(0.078)	(0.092)	(0.094)	(0.615)
No Music	1.014	.641***†††	.828	.944	.848†	.865	.690***†††	.721***†††	.841†	.808*††	−1.548***
	(0.121)	(0.077)	(0.096)	(0.108)	(0.098)	(0.096)	(0.080)	(0.080)	(0.094)	(0.091)	(0.600)
Constant											19.042***
											(2.305)
Observations	2,101	2,101	2,101	2,101	2,101	2,101	2,101	2,101	2,101	2,101	2,101
R^2											.262
Estimation	Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit	OLS
Coefficient	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	Odds Ratio	OLS

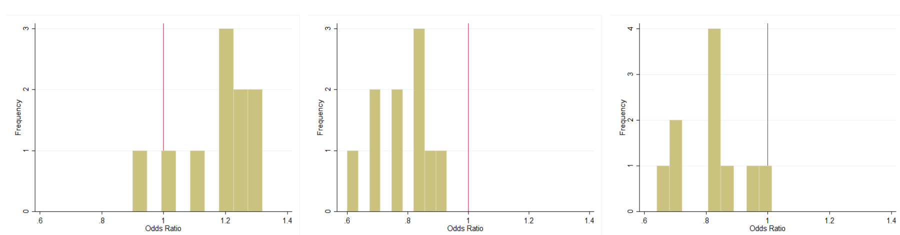
Note: P-values (not Bonferroni adjusted): *** < 0.01, ** < 0.05, * < 0.1 (from robust standard errors). SUR P-values: † † † < 0.01, †† < 0.05, † < 0.1. All estimations apply the control variables.

The aggregated variable tells a general story about the in-store music and emotions, confirming that the presence of music and the congruence between music and brand in general affect mood and emotions. To be conservative about significance, which one should be when estimating many equations, we also multiply the p-values by the number of regressions (10). This procedure is called Bonferroni adjustment and is very conservative. We do not present the recalculated p-values, but it is easy to see that the coefficients that are significant at the 99 percent level (***) are also significant at least at the 90 percent level after the Bonferroni adjustment. In fact, all of these are significant at the 95 percent level after the Bonferroni adjustment. The Bonferroni adjustment assures us that 3 of 10 emotions are affected by *no music* and *no brand-fit*, confirming the results from the aggregate variable.

Another way to analyze a group of similar outcome variables is seemingly unrelated regression (SUR). SUR estimates the equations linearly but allows the error terms to be correlated. We do this in addition to the other methods but only present indicators of p-values (all significant coefficients are of the same sign as in the ordered logit). These results are stronger, showing that *no brand-fit* and *no music* significantly decrease the intensity of 8 of 10 and 6 of 10 emotions, respectively. The coefficient for *brand-fit B* is significant for only one emotion, suggesting that the difference between *brand-fit A* and *B* is less likely to affect moods. This is in line with the hypotheses that expect brand-fit music and the mere presence of music positively affect emotions but that the difference between well-known and less well-known (*brand-fit B* and *brand-fit A*) does not matter.

We also present the distributions of the estimated treatment odds rations in Figure 5. These figures visualize what we observe in Table 7. It is clear that the emotions are overall negatively affected by the absence of music (Panel c) and poor fit between music and brand (Panel b). Nearly all estimated odds ratios of brand-fit B (Panel a) are above 1, indicating that if emotions actually depend on brand-fit A or B, the familiarity of songs increases emotion intensity.

Figure 5: Distribution of Estimates



(a) Brand-Fit B (b) No Brand-Fit (c) No Music

To ensure that the effect of music is unconscious, which the literature suggests, we check whether customers take notice. Table 8 presents the proportion of "yes" responses by music treatment to question 4.1(a): "I noticed the music played in the restaurant today."

Table 8: Question: did you notice the music in the restaurant?

Treatment	**# of obs.**	**Mean YES**
Brand-Fit A	534	.479
Brand-Fit B	558	.629
No Brand-Fit	508	.602
No Music	501	.427
Total	2101	.536

These data indicate that the share of respondents who noticed the music was higher during the *brand-fit B* music treatment (63 percent) than during *brand-fit A* (48 percent), indicating that customers are more likely to notice the music when the songs are well-known. More important, even when there was *no music*, 42.7 percent of the respondents claimed that they noticed the music in the restaurant. They also evaluated the (absent) music in the following sub-questions. This implies that customers are not especially aware of in-store music and that the effect of music is to a great extent unconscious.

6 Conclusions

Field experiments have become increasingly common in economics because they have proven useful for drawing causal inference for different types of interventions. However, as Levitt and List (2009, p. 10) note, field experiments in the private sector "represent a largely untapped opportunity for future research. There are many issues of central economic importance that can benefit from field experimentation, but generally require the partnership of firms to examine. These include certain questions pertaining to consumer choice...".

One question that has attracted considerable attention in the literature on consumer choice is whether and how store atmospherics affect consumers' perceptions and choice. It is often assumed that different atmospheric cues can cause the consumer to either approach or avoid the environment and thereby lead to increased sales and more satisfied customers. In-store music has been identified as one such potentially important atmospheric cue, and previous studies have focused on the importance of certain characteristics of the in-store music, such as tempo and music style. These experiments have generally been limited to one store or restaurant and implemented over a short period. Because the results are, thus, based on small samples and descriptive methods, inference may only be made with great care.

To overcome these limitations, we conducted a field experiment with support from Soundtrack Your Brand, the exclusive provider of Spotify Business. This enabled us to test whether a real-world business solution can increase sales and appeal to consumers' self-perceptions by making in-store music convey brand values. The support from Soundtrack Your Brand also meant that we could test whether the inclusion of less well-known songs can increase sales more than simply playing familiar music, a question that has yet to be studied.

The use of 16 restaurants over a period of 20 weeks made it possible to identify treatment effects that, through randomization and statistical methods, have causal interpretation. It is complicated to ensure statistical significance with such data, and we are keen to not overstate any findings. We believe that our statistical analysis and inference are appropriate given the structure of the data.

The results showed that congruence between in-store music and brand values increased sales. The difference in revenues (quantity sold) between *brand-fit A* and *no brand-fit* was 9.1 (9.2) percent. The results from the drive-through analysis indicate that there is also a difference depending on the familiarity of the songs and the mere presence of music. Furthermore, the effect of brand-fit compared to no brand-fit was stronger in the drive-through analysis. The reason could be that the comparison group in this analysis is better, but it is also possible that the effects of music choice are stronger in road restaurants. To identify the effects of less well-known songs and the mere presence of music more generally, we would need more restaurants or time periods. The survey responses also indicated that the brand-fit music had a positive impact on consumers' emotions. They were, however, often unaware of the music that was played, implying that the music affects consumers unconsciously. It is possible that the chain of causality runs from music to emotions and from there on to consumption, supporting the Donovan and Rossiter (1982) model. The economic research on self-control (e.g., Loewenstein, 1996; Kahneman et al., 1997; Meloy 2000) suggests that increased emotional intensity induces individuals to act more impulsively, leading to increased consumption. It is, however, possible that it was the consumption that induced the positive emotions. Studies such as Bernheim and Rangel (2004) conclude that environmental cues can lead to increased consumption in a more direct way. We were unable to distinguish between direct or indirect effects on sales, but we showed that the emotions are also affected. Future studies should therefore attempt to explicitly demonstrate what role emotions play.

An advantage of our field experiment is that we observe real decisions in a natural situation. The disadvantage is that we study a specific situation, namely a chain restaurant in the Stockholm area. It is possible that our results are generalizable to other contexts, but it is likely that the effect differs depending on the context. The more different the context, the less likely the effect is to be the same. It is possible that consumers are more affected by music in other venues, such as coffee shops or restaurants, where the visit and exposure to the music is longer. We therefore believe that our study of a restaurant chain likely provides a conservative estimate of how brand-fit in-store music affects consumers' perceptions and behavior. Further research is needed to understand the effect in different settings.

There are many other possible avenues for further research. A potentially very important question is how the music influences the employees. In fact, early studies in this research field analyzed how factory employees were affected by music (Wyatt and Langdon, 1937), but the contemporary focus is almost entirely on consumers. This is somewhat surprising since the music might affect the mood of the employees and how they approach their customers, thereby influencing business performance.

7 Appendix

Figure 6: Daily Revenues

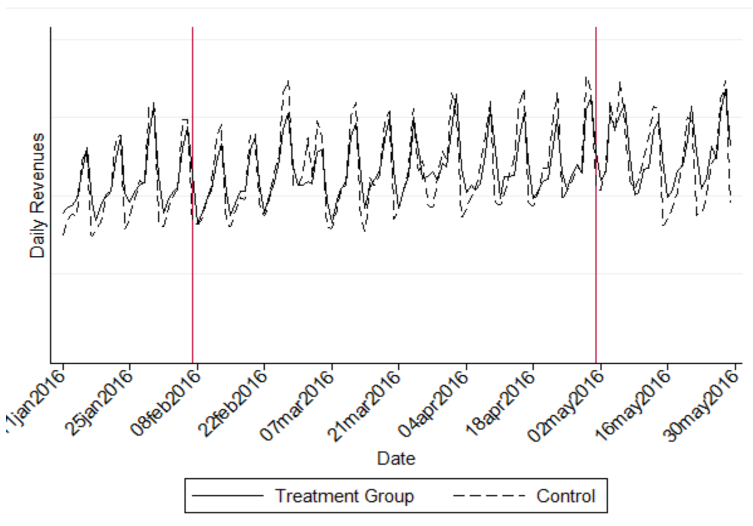


Table 9: Drive-Through - Descriptive statistics

Group	# of obs.	Mean Daily Revenues Full Period	Pre	Experiment	Post
Drive Through	1 118	124.78	114.19	124.37	137.00
		(41.60)	(39.06)	(40.91)	(43.46)
Front Counter	1 118	112.90	100.00	112.55	127.41
		(39.02)	(34.18)	(38.08)	(41.92)
All	2 236	118.87	107.10	118.46	132.20
		(40.76)	(37.34)	(39.94)	(42.92)

Group	# of obs.	Mean Daily Quantity Sold Full Period	Pre	Experiment	Post
Drive Through	1 118	128.96	119.12	129.33	138.23
		(42.07)	(41.06)	(41.78)	(42.22)
Front Counter	1 118	113.69	100.00	113.69	128.10
		(41.49)	(37.07)	(40.84)	(43.37)
All	2 236	121.36	109.56	121.51	133.16
		(42.51)	(40.26)	(42.00)	(43.01)

Table 10: Parts of the day (Log Revenues)

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	Breakfast	Lunch	Snack	Dinner	Evening	Late Night
Brand-Fit B	$-.052^{*}$	$-.038$	$-.040$	$-.014$	.002	$-.009$
	(.030)	(.040)	(.033)	(.035)	(.029)	(.072)
No Brand-Fit	$-.038$	$-.108^{**}$	$-.112^{***}$	$-.080^{*}$	$-.050$	$-.043$
	(.030)	(.054)	(.037)	(.044)	(.035)	(.119)
No Music	$-.027$	$-.051$	$-.030$	$-.021$	$-.014$	$-.099$
	(.038)	(.059)	(.045)	(.050)	(.034)	(.140)
Observations	2,215	2,232	2,232	2,232	2,226	1,602
R^2 within	.346	.201	.478	.236	.300	.478
Number of restaurant	16	16	16	16	16	14

Note: Bootstrapped standard errors in parentheses (number of replications: 1000. Grouped on restaurant). Revenues and quantity sold are in logarithmic values. All estimations include fixed effects for restaurant, week and day of the week. P-values (inferred from degrees-of-freedom adjusted t-values): *** < 0.01, ** < 0.05, * < 0.1.

Acknowledgements

We thank Soundtrack Your Brand, especially Jasmine Moradi, for making this study possible and for genuine support during the process.

References

Aaker, J. (1997). Dimensions of brand personality. *Journal of Marketing Research*, 19, 347-356.

Angrist, J. D. & Imbens, G. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467-475.

Angrist, J. D. & Imbens, G. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 91, 444-455.

Areni, C. S. & Kim, D. (1993). The influence of background music on shopping behavior: classical versus top-forty music in a wine store. *Advances in Consumer Research*, 20(1), 336-340.

Babin, B. J. & Darden, W. R. (1995). Consumer self-regulation in a retail environment. *Journal of Retailing*, 71(1), 47-70.

Bellenger, D. N., Robertson, D. H. & Hirschman, E. C. (1978). Impulse buying varies by product. *Journal of Advertising Research*, 18(December), 15-18.

Bernheim, B. D. & Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 94(5), 1558-1590.

Bruner, G. C. (1990). Music, mood and marketing. *Journal of Marketing*, 54, 94-104.

Buchanan, J. (1975). The Samaritan's Dilemma. In *Altruism, Morality, and Economic Theory,* edited by Edmund S. Phelps. New York: Russell Sage.

Cameron, A. C. & Miller, D. L. (2008). A practitioner's guide to cluster-robust inference. *The Journal of Human Resources*, 50(2), 317-372.

Capra, M. (2004). Mood-driven behavior in strategic interaction. *American Economic Review*, 94(2), 367-372.

Card, D. & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review.* 84(4), 772–793.

Cattell, R. B. & Saunders, D. R. (1954). Musical preferences and personality diagnosis: I. A factorization of one hundred and twenty themes. *The Journal of Social Psychology*, 39(1), 3-24.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2), 215-242.

Donovan, R. J. & Rossiter, J. R. (1982). Store atmosphere: An environmental psychology approach. *Journal of Retailing*, 58(1), 34-57.

Doyle, P. (2002). *Marketing management and strategy*, 3rd ed. 184. Warwick: Prentice Hall.

Duflo, E., Glennerster, R. & Kremer, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, 4, 3895-3962.

Garlin, F. V. & Owen, K. (2006). Setting the tone with a tune: A meta-analytic review of the effects of background music in retail settings. *Journal of Business Research*, 59(6), 755-764.

Gul, F. & Pesendorfer, W. (2004). Self-control and the theory of consumption. *Econometrica*, 72(1), 119-158.

Hinkelmann, K. & Kempthorne, O. (2008). *Design and analysis of experiments. Vol. 1, Introduction to experimental design. 2 ed.* Hoboken, N.J.: Wiley-Interscience

Kahneman, D., Wakker, P. P. & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, 112(2), 375-405.

Kellaris, J. J. (2008). Music and consumers. In Haugtvedt, C.P., Herr, P. M. & Kardes, F. R. (Eds.), *Handbook of consumer psychology* (837-856). New York: Taylor & Francis Group/Lawrence Erlbaum Associates

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114(2), 497-532.

Levitt, S. D. & List, J. A. (2009). Field experiments in economics: the past, the present, and the future. *European Economic Review*, 53(1), 1-18.

Loewenstein, G. (1996). Out of control: visceral influences on behavior. *Organizational, Behavior and Human Decision Processes*, 65(3), 272-292.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2), 426-432.

Mehrabian, A. & Russel, J. A. (1974). *An approach to environmental psychology*. Cambridge, Massachussetts: M.I.T. Press.

Meloy, M. G. (2000). Mood-driven distortion of product information. *Journal of Consumer Research*, 27, 345-359.

North, A. C., Hargreaves, D. J. & McKendrick, J. (1999). The influence of in-store music on wine selections. *Journal of Applied Psychology*, 84(2), 271.

Pavlov, I. P. (1926). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. New York.

Radocy, R. E. & Boyle, J. D. (1997). *Psychological Foundations of Musical Behavior*, 3rd ed. Springfield, IL: Charles C. Thomas.

Rentfrow, P. J. & Gosling, S. D. (2006). Message in a ballad the role of music preferences in interpersonal perception. *Psychological Science*, 17(3), 236-242.

Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

Schelling, T. C. (1978). Egonomics, or the art of self-management. *American Economic Review*, 68(2), 290-294.

Smith, A. *Theory of Moral Sentiments*. London: Millar, 1759.

Thaler, R. H. & Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy*, 89(2), 392-406.

Wilson, S. (2003). The effect of music on perceived atmosphere and purchase intentions in a restaurant. *Psychology of Music*, 7(2), 55-63.

Wyatt, S. & Langdon, J.N. (1937). Fatigue and boredom in repetitive work. *Industrial Health Research Board Report*, 77, Great Britain Medical Research Council, London.

Yalch, R. F. & Spangenberg, E. R. (2000). The effects of music in retail setting on real and perceived shopping times. *Journal of Business Research*, 49(2), 139-147.