Dislikes: Results

Klaus Frieler

17.01.2022

## Participants

We collected data from a total of 639 participants. After inspecting the rating behavior for anomalies, we excluded 12 participants, who showed 90% of all their ratings on the extreme end of the ratings scales (low or high). After this step, 627 participants (402 female, 225 male) stayed in the study, with a median age of 26 (range: 18-75, IQR = 12). The age distribution appeared to be a bimodal distribution, which is rather typical for this kind of online study with a convenience sample. Fitting a Gaussian Mixture Model (using the MClust package for R) corroborated this visual impression. We found one age group with participants up to 30 years (N = 417, 66.5%), and another one with participants older than 30 years (N = 210, 33.5%).

The participants were mainly highly educated, with 48.6% having a university or college degree (Bachelor’s, Master’s, or Ph.D.), and 94.7% having a school degree of A-levels (“Abitur”), with 56.6% of the sample being students. Nevertheless, the older age group was a little bit less educated than the younger, as measured by the proportion of participants having at least A-levels (97.6% for the younger vs. 89% for the older, Chi-squared(1) = 20.50, p = 0.000, Cramer's V = 0.18).

## Construction of a Dislike Scale

In order to reduce the number of variables and thus the complexity of the analysis, we first applied an exploratory factor analysis to the rating scales, across all four conditions. This resulted in an oblique factor solution with eight factors (explaining a total of 48% of variance, see Supplementary Material, Sect. *Orignal Factor Solution*). After inspecting this factor solution, it became clear that the factors did not neatly represent distinct psychological constructs (or explanatory strategies), but that different constructs had been merged into some of the factors due to strong correlations. As using these impure factors would have made the interpretation of results blurry, we decided to “purify” the original factors manually by constructing a scale with a conceptually disentangled set of subscales. This led finally to a “virtual dislike scale” with nine subscales with a clearer semantic separation. The definition of the subscales in terms of the original items can be found in Tab. 1. (A list of item names and item wording as well asa full correlation matrix of all single items, can be found in the Supplementary Material, Tab. S1). Subscale scores were calculated as the mean value of all involved items after using the R package mice for imputing missing values. The subscales were named **Too Niche**, **Too Complex**, **Too Emotional**, **Too Simple**, **Not Authentic**, **Too Mainstream**, **Social Incongruence**, **No Impact**, and **Displeasure**. The **Not Authentic** subscale consists of only one item, which is basically suboptimal for a subscale, but this choice can be justified by the item’s unique semantic position within all items. Authenticity is a very specific cultural concept, which cannot be grouped *a priori* with any of the other items in our view.

The subscales presented themselves as one set of reasons regarding the “expression” of the music and the other regarding the impression on the listener in the form of emotional “reactions” to the disliked music. The latter set comprises the subscales **No Impact** and **Displeasure**. Either the music has no effect or causes displeasure, and both are (conceptually disjunctive) reasons for disliked music. This also implies that people expect an impact from music, and when this impact is not provided, then it is a reason for disapproval (though more often, lack of impact might, in fact, cause indifference, but since the participants in our study were only asked for disliked music, we cannot make any statements about this from our data). The “expression” of the music regards the musical content (**Too Niche**, **Too Complex**, **Too Simple**, **Too Emotional**) and social aspects (**Too Mainstream**, **Not Authentic**, **Social Incongruence**).

**Table** : Definition of the virtual Dislike scale with subscales

| Subscale | No. Items | Items |
| --- | --- | --- |
| Too Niche | 10 | body.missing\_danceability, music.bad\_vocals, music.disliked\_instruments, music.too\_chaotic, music.too\_disharmonic, music.too\_fast, music.too\_little\_melodious, music.too\_loud, music.too\_niche, music.too\_unrhythmic |
| Too Complex | 5 | lyrics.too\_complex, lyrics.too\_realistic, music.too\_complex, music.too\_much\_change, music.too\_variable |
| Too Emotional | 5 | emo.too\_emotional, music.too\_melodious, music.too\_rhythmic, music.too\_slow, music.too\_soft |
| Too Simple | 6 | lyrics.too\_simple, lyrics.too\_unrealistic, music.too\_little\_tension, music.too\_schematic, music.too\_simple, music.too\_uniform |
| Not Authentic | 1 | social.not\_authentic |
| Too Mainstream | 3 | music.too\_little\_change, music.too\_mainstream, social.too\_often\_heard |
| Social Incongruence | 4 | social.incongruent\_ideology, social.no\_identification, social.not\_peer\_approved, social.reject\_fanbase |
| No Impact | 3 | emo.expressionless, emo.no\_feelings, emo.no\_impact |
| Displeasure | 4 | body.displeasure, emo.bad\_feelings, emo.bad\_mood, social.bad\_experiences |

## Correlations of subscales

The scale distributions and correlations across all conditions can be seen in the panel plot in Fig. 1 (see Supplementary Material Fig. SP2-SP5, for panel plots for all four conditions). All scales differ visually from normality, which was also the case for most original items. They show either strong (mostly left) skewness or bimodality. Only **Social Incongruence** and **No Impact** are unimodal with a mode not at an extreme end of the scale. The two scales **Too Complex** and **Too Emotional** are rarely used. The scale **Not Authentic** shows a clearly bimodal distribution, indicating that artists and styles are typically considered as either authentic or not; there does not seem to be a strong middle ground for authenticity in the view of the respondents.

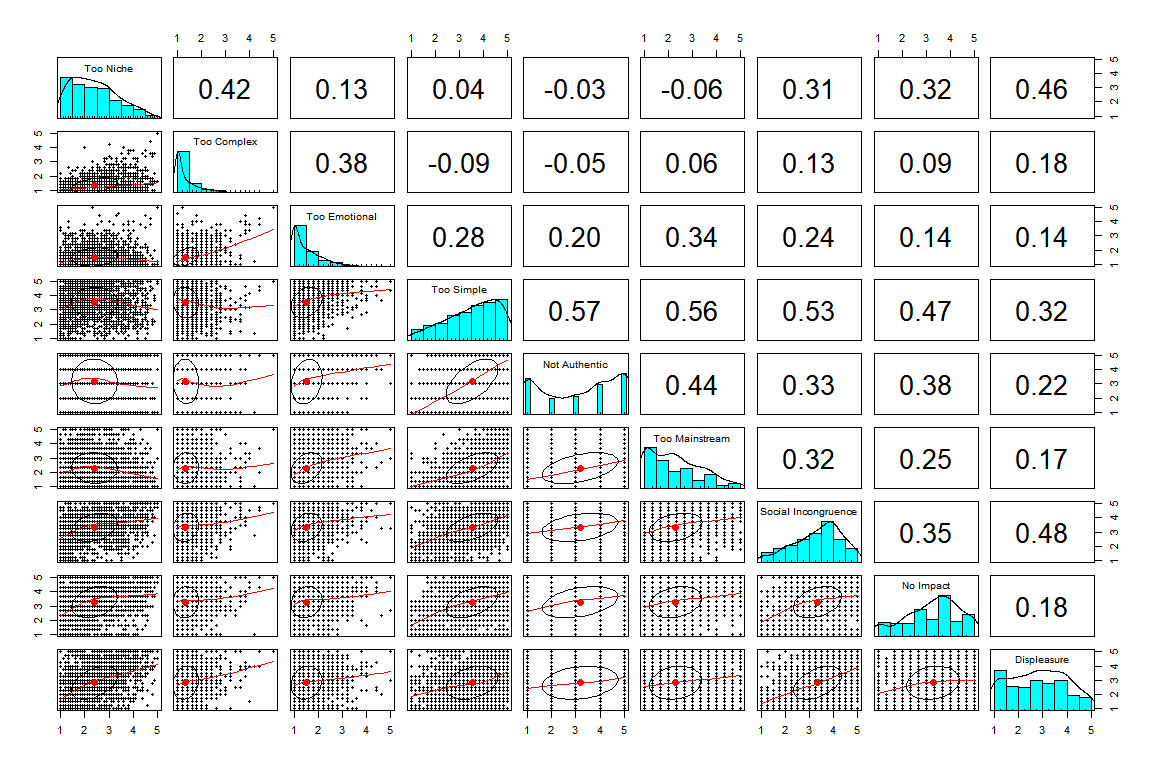


Fig. 1. Panel plot of Dislike subscales

However, the correlation matrices of the subscales differ significantly between the four main conditions (Jennrich’s test for correlation matrices: *p < .001* for all condition pairs). The distributions are partly also different, but the general observations from above still holds up (see Supplementary Material, Fig. SP1-5).

In order to find the most stable correlations, we used the correlation matrices for the complete data set and the four conditions and selected all pairs of correlations that were highly significant in all four correlation matrices (using Holm adjustment and an significance level of *.002*, because there was one adjusted p-value of *p = .00125*). This left us with 14 correlations, which can be found in Tab. 2

**Table** : Stable correlations across all conditions of the virtual Dislike subscales

| Subscale Pair | Range | Mean |
| --- | --- | --- |
| Too Simple - Too Mainstream | 0.527-0.620 | 0.566 |
| Too Simple - Not Authentic | 0.501-0.591 | 0.559 |
| Too Simple - Social Incongruence | 0.326-0.612 | 0.496 |
| Too Simple - No Impact | 0.292-0.619 | 0.465 |
| Not Authentic - Too Mainstream | 0.416-0.470 | 0.436 |
| Too Niche - Too Complex | 0.409-0.460 | 0.435 |
| Too Niche - Displeasure | 0.308-0.527 | 0.435 |
| Social Incongruence - Displeasure | 0.242-0.530 | 0.422 |
| Not Authentic - No Impact | 0.289-0.457 | 0.382 |
| Too Complex - Too Emotional | 0.311-0.436 | 0.374 |
| Too Emotional - Too Mainstream | 0.298-0.383 | 0.346 |
| Too Mainstream - Social Incongruence | 0.259-0.431 | 0.326 |
| Not Authentic - Social Incongruence | 0.209-0.443 | 0.309 |
| Too Emotional - Too Simple | 0.256-0.315 | 0.284 |

## Latent Profile Analysis

In order to further investigate the explanatory strategies of the participants, we looked for latent profiles in the rating behavior. To this end, we used the R package tidyLPA to extract two latent profiles for each of the four conditions (Artist/Style, Strong/Slight). Two profiles are not the optimal solution according to standard model estimators (AIC, BIC), but it is the most simple solution which produced the most balanced group sizes and also has the easiest interpretation in accordance with the stable correlations.

As presented in Fig. 3, latent profiles had different characteristics regarding the subscales. Profile 1 (red) contained those participants with high levels of **Too Simple**, **Not Authentic**, **No Impact**, and **Social Incongruence**, medium levels of **Too Mainstream**, **Displeasure**, and low levels of **Too Niche**, **Too Emotional**, and **Too Complex**. LP1 included 60-70% of the participants, depending on the condition. Profile 2 (black) contained those participants with high levels of **Too Niche** and **Too Complex**, but in the strong dislike condition only.

The profiles specify two groups that use different explanatory strategies for their disliked music which reflect their general attitude toward music and their overall musical taste. Participants in LP1 were dubbed “highbrow” and participants in LP2 “lowbrow.” The concept of a highbrow and a lowbrow in musical taste is well established in the literature, where it refers to people disliking certain musical styles which are associated with these attributes (see also below for the comparison of style and LPs). It is of note that only in two subscales in two conditions, the lowbrows show a higher judgement than the highbrows. They do not use the same explanatory strategy for all conditions. (This should be prepared in the Introduction and this part can be shortened.)

It is of note that the highbrows show a clear/similar profile for all conditions, i.e., these participants use the same explanatory strategies for all conditions. The lowbrow profile is similar within the strong degree of dislike, but not for the slight degree of dislike. Hence, these participants show a more diverse/flexible pattern in their explanatory strategies, where in the slight degree of dislike, the judgments do not seem to be strong, i.e., nothing seems to be that bad. Taken together, we show for the first time that the highbrow/lowbrow is not only associated with certain musical styles, but also with the associated attributes/features of the music. (Rework and move to Discussion)

Overview statistics of **Highbrow** and **Lowbrow** profiles can be found in Tab. 2. Mean values of the Dislike subscales can be seen in Fig. 3.

**Table** : Latent Profile Classes by type and degree

| Type | Degree | Cases | No. Lowbrow | No. Highbrow | Lowbrow (%) | Highbrow (%) |
| --- | --- | --- | --- | --- | --- | --- |
| artist | slight | 348 | 130 | 218 | 37.4 | 62.6 |
| artist | strong | 489 | 150 | 339 | 30.7 | 69.3 |
| style | slight | 591 | 260 | 331 | 44.0 | 56.0 |
| style | strong | 582 | 215 | 367 | 36.9 | 63.1 |

Interestingly, not all participants belong to the same LPA class across the four conditions. We defined a total LPA class by investigating the profile statistics of each participant, defining five classes “lowbrow”, “mainly lowbrow”, “mixed”, “mainly highbrow”, and “highbrow”. The extreme classes are assigned if a participant uses the same profile in all conditions, the mainly highbrow/lowbrow classes are used when the majority but not all profiles are present, and the mixed class is applied to participants with an equal split of **Lowbrow** and **Highbrow** profiles. The distribution can be seen in Tab. 3. Highbrow and Mainly Highbrow profiles are prevalent, only 14% of all participants did not show at least one Highbrow profile. This might be due to our overall well-educated sample.

**Table** : Overall Latent Profile Classes of participants

| Overall LPA class | N | Percentage (%) |
| --- | --- | --- |
| Lowbrow | 87 | 13.9 |
| Mainly Lowbrow | 89 | 14.2 |
| Mixed | 96 | 15.3 |
| Mainly Highbrow | 142 | 22.6 |
| Highbrow | 213 | 34.0 |

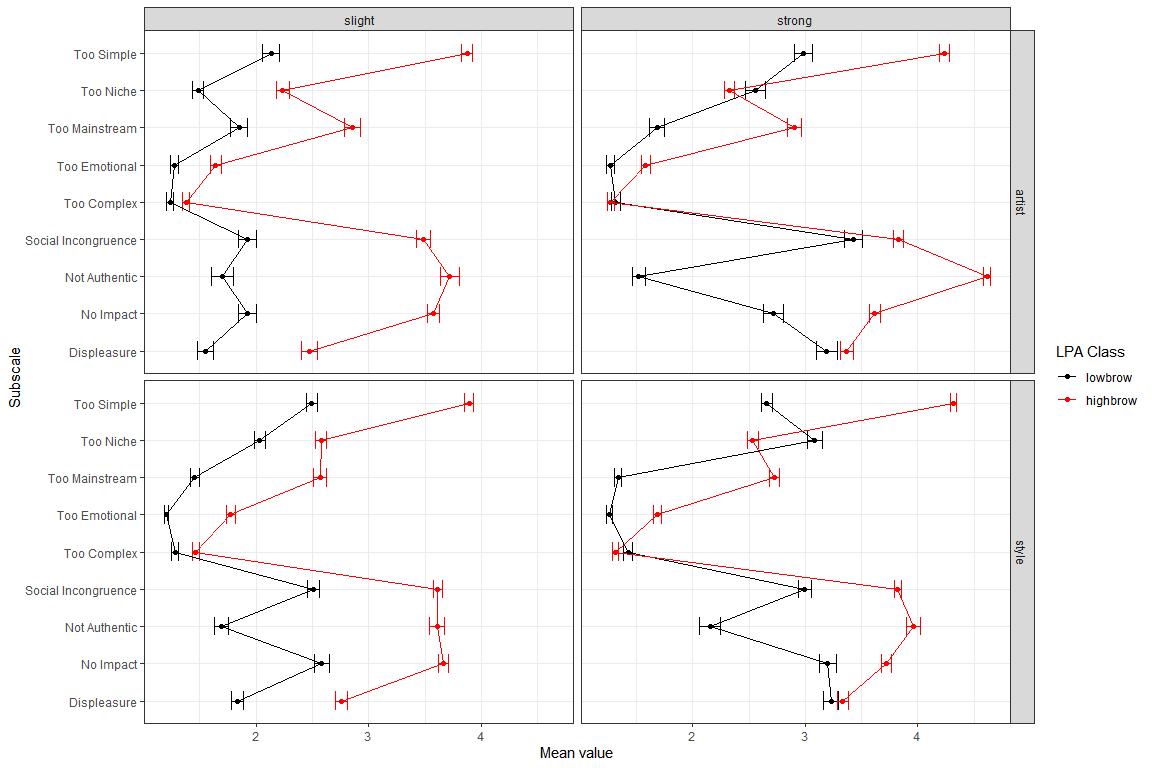


Fig. 3. Latent Profile Class means by conditions. Errorbars represent standard errors.

## Causal analysis [Discussion]

Because of the specific semantic structure, some cautious and preliminary causal statements can be derived from these stable correlations and the LPA. The reason is that emotional reactions to music cannot cause the structure of music (though it could partly influence the perception or judgment of musical properties). The same holds for the perceived **Social Incongruence** between oneself and the fans of this music (even though people might believe in an implicit homology between musical and social/personal characteristics of its fans). We summarized these ideas in some sort of causal network, that can be found in Fig. 2. Nodes represent the subscales, shape encodes the classification into subscales regarding the “expression” and emotional “reaction” variables. Node color further classified the nodes into interpretations. Arrows indicate causal influence, if the arrows goes both ways between nodes, no causal identification is available. Distances between nodes roughly represent mean correlation between variables. There is a cluster of three (or four) subscales, **Too Simple**, **Not Authentic**, **Too Mainstream**, and, possibly, also **Too Emotional**, at the center of the network. We interpret this as a typical “highbrow” reasoning complex of disliking music. The results of this stance is a lessened impact of the music as well as a perceived **Social Incongruence** to the audience of the disliked music. In our data set, this is a very common attitude toward *Schlager* and *Traditional* music, which are the top disliked styles. On the other hand, there is another reasoning complex, consisting of **Too Niche** and **Too Complex**, which can be dubbed “Lowbrow”, Here, only **Too Niche** leads to **Displeasure**, while **Too Complex** is linked to **Too Emotional**. (The latter correlation seems to be genuine and not an artifact of the extremely left-skewed distributions of the two variables and reveals that complexity is not limited to musical features, but can refer to the emotional expression as well.)

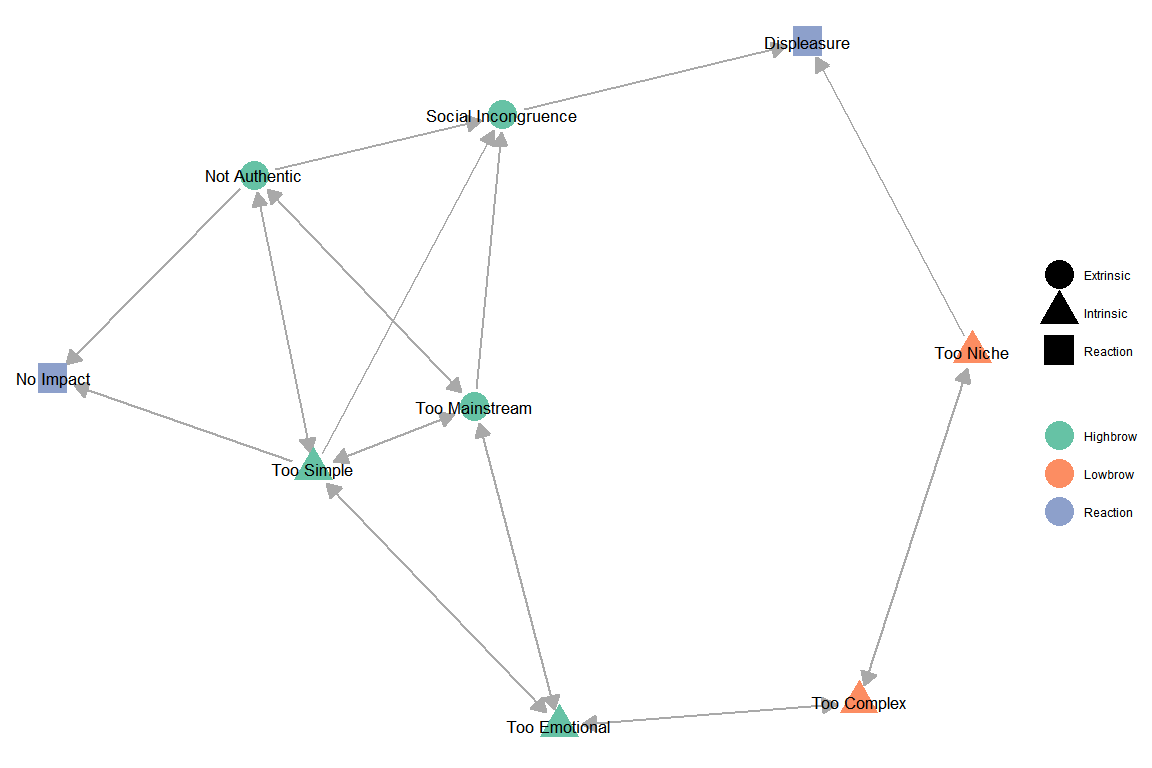


Fig. 2. Network of stable correlations between subscales

It also interesting to have a short look at the correlations that are nearly fully stable, i.e., those which occur in 4 of the 5 correlation matrices. There are five such correlations, neither of which did not appear in the full correlation matrix. The correlation **Too Niche** - **Social Incongruence** (mean *r = .35*), **Too Niche** - **No Impact** (mean *r = .346*), **Too Mainstream** - **No Impact** (mean *r = .294*) did not show up in the Style/Strong condition. The correlation **Too Emotional** - **Social Incongruence** (mean *r = .268*) was not present in the Artist/String conditions, whereas **Social Incongruence** - **No Impact** (mean *r = .375*) did not make it into the Artist/Slight condition (*KF: Interpretation anyone?*).

## Differences in style selections

Next, we analysed the distribution of styles and their assessment. In the two Style conditions, participants had to explicitly select disliked styles from a list with 15 suggestions. In the two Artist conditions, participants were also asked to provide a style descriptor for the chosen artist, even though sometimes they were not able to do so (42 out of 837). The distribution of styles is rather imbalanced, see Tab. 4. The top disliked styles were *Schlager*, *Traditional* and *HipHop*. *Schlager* occurs about twice as often as the next styles, either *HipHop* or *Tradtional*, which highlights the special role *Schlager* plays in the German music culture.

There are clear differences between the Artist and Style conditions. In the Artist condition, there were much more mentions of *Pop* and *Rock* artists than in the Style condition, whereas *Schlager*, *HipHop*, and *Classical* were about equally often selected, and all other styles appeared more frequently selected in the Style conditions. This can be explained by the fact, that *Pop* and *Rock* are very broad and prominent styles that are rarely disliked in its entirety, but single artists from these styles can easily draw very negative judgments. It is also more likely that people actually know artists from these styles due to their higher popularity, whereas for more generally disliked styles such as *Metal*, *Techno*, and *Traditional* people are less likely to know names of specific artists, because these are (1) disliked styles, and (2) they are not (anymore) strongly present in public media, or operate generally without a prominent star system such as *Techno*.

Comparing the degrees of dislike, one finds that *Schlager*, *Traditional*, and *Metal* are always more frequent in the Strong category than in the Slight category. Techno is slightly more often mentioned in the Strong subcondition of the Style condition, but about equally often in both Artist conditions. All other styles tend to occur more often in the Slight than in the Strong degree. This is particularly true for **Pop** in the Artist/Slight condition. All in all, the contrast in style distribution between the degrees are less clear than in the comparison of Artist and Style conditions.

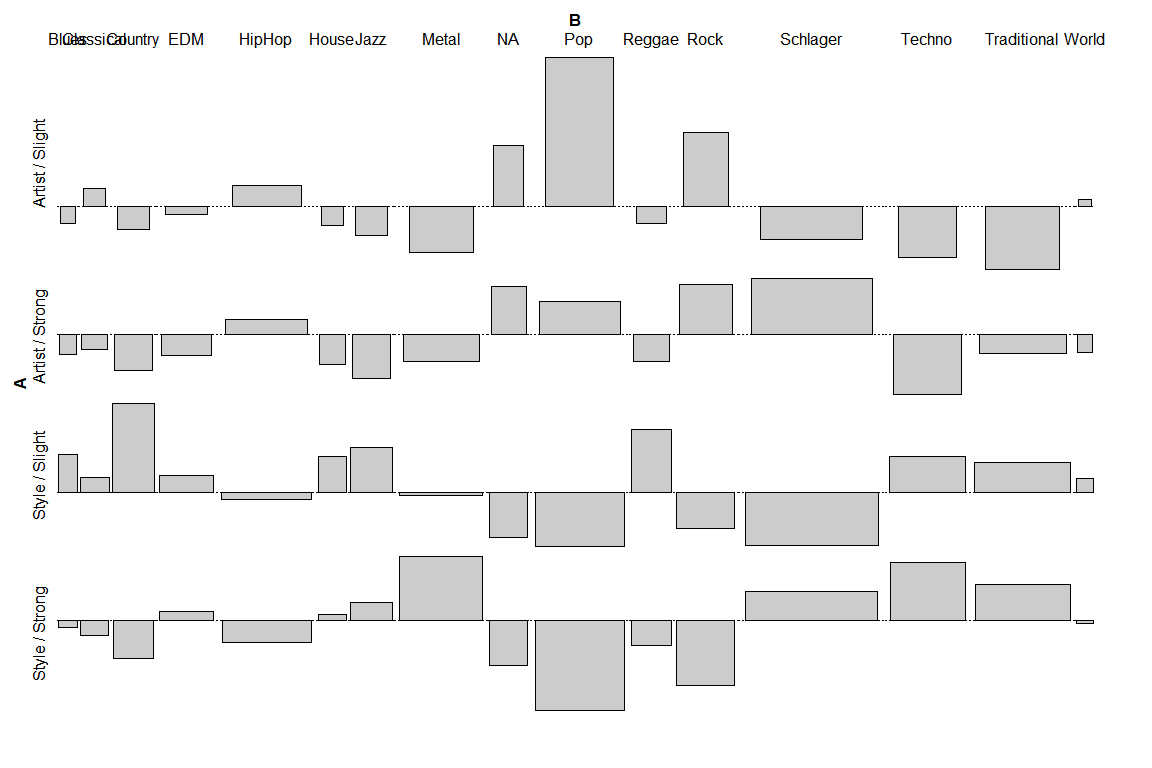
**Table** : Style counts and percentages by main condition (Style and Artist)

| Style | Count | Perc. (Artist) | Perc. (Style) | Perc. (Artist/Slight) | Perc. (Artist/Strong) | Perc. (Style/Slight) | Perc. (Style/Strong) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Schlager | 503 | 27.8 | 23.0 | 18.1 | 34.8 | 16.4 | 29.7 |
| Traditional | 260 | 7.5 | 16.8 | 3.4 | 10.4 | 16.4 | 17.2 |
| HipHop | 231 | 13.7 | 9.9 | 14.4 | 13.3 | 10.7 | 9.1 |
| Pop | 227 | 22.2 | 3.5 | 32.2 | 15.1 | 5.4 | 1.5 |
| Metal | 197 | 5.5 | 12.9 | 3.7 | 6.7 | 9.5 | 16.3 |
| Techno | 162 | 2.0 | 12.4 | 2.0 | 2.0 | 11.3 | 13.4 |
| Rock | 96 | 9.8 | 1.2 | 11.5 | 8.6 | 2.2 | 0.2 |
| EDM | 84 | 3.0 | 5.0 | 3.4 | 2.7 | 5.2 | 4.8 |
| Jazz | 50 | 0.2 | 4.1 | 0.6 | 0.0 | 4.7 | 3.4 |
| Country | 48 | 0.6 | 3.7 | 0.9 | 0.4 | 6.8 | 0.5 |
| Reggae | 45 | 1.0 | 3.2 | 1.1 | 0.8 | 5.2 | 1.0 |
| NA | 42 | 5.0 | 0.0 | 5.7 | 4.5 | 0.0 | 0.0 |
| Classical | 24 | 1.2 | 1.2 | 2.0 | 0.6 | 1.7 | 0.7 |
| House | 23 | 0.1 | 1.9 | 0.3 | 0.0 | 2.4 | 1.4 |
| Blues | 10 | 0.0 | 0.9 | 0.0 | 0.0 | 1.4 | 0.3 |
| World | 8 | 0.2 | 0.5 | 0.6 | 0.0 | 0.7 | 0.3 |
| Sum | 2,010 | 99.8 | 100.2 | 100.0 | 100.0 | 100.0 | 100.0 |

Participants used a large variety of styles to describe their dislikes: 48.8% of the participants selected a different style in all four conditions (or only took part in one condition), whereas nobody mentioned only one style (but note that participant had to chose two different styles in the Style condition).

To check whether the style distributions showed differences between the type and degree of dislike, we used a bootstrap Chi-squared test (with 100 samples) to alleviate the repeated measurement problem. This became highly significant (all *p < .001*, mean *Cramers V = .365*) with a strong effect. The same held true for comparing Style vs. Artist conditions (all *p < .001*, mean *Cramer’s V = .502*) and also for Strong vs. Slight degrees (all *p < .001*, mean *Cramer’s V = .282*).

The main differences in the degree of dislike is that *Pop*, *Country*, and *Reggae* artists were chosen more frequently in the Slight conditions, whereas *Metal*, *Schlager*, and *Traditional* occurred more often in the Strong conditions. Hence, *Metal*, *Schlager* and *Traditional* are the most and *Pop*, *Country*, and *Reggae* belong to the least disliked styles in the current study.

 ## Rating differences in Type and Degree of dislike

Besides the clearly different distributions of styles in all conditions, we were also interested in differences in ratings in the Type and Degree of dislike. As the distribution of styles was rather different between the Types (cf. Tab. 4), we treated both cases separately and stratified also for styles. Furthermore, as the style distributions are very skewed and the Dislike subscales clearly differ from normality, we used an permutation test (independence\_test from the coin package for R) as an omnibus test and then resorted to a battery of Kruskal-Wallis tests to check for differences in Type and Degree on all subscales while using Holm adjustment of p-values to account for multiple testing.

The results for comparing the Degrees of dislike, can be found in Tab. 7, where only significant differences at the 5% level (adjusted) are shown (in the Artist condition the cases with no assigned style are omitted, for a complete table please refer to the Supplementary Material, Tab. 1). Both omnibus tests yielded highly significant differences (*p < .001*). A graphical display of mean values and 95% confidence intervals can be seen in Fig. 6 and 7.

The main differences are, expectedly, to be found on the subscale **Displeasure**, where values in the Strong degree are higher than in the Slight degree, with overall larger differences in the Artist main condition. Other differences pertain to **Social Incongruence** (Artist: *Pop* and *Rock*, Style: *Schlager*) and **Not Authentic** (Artist: *Pop* and *Schlager*). All in all, the main differences between Type and Degree lie mostly in the selection of styles as well as in the higher **Displeasure** ratings in the Strong vs. the Slight degree.

Both omnibus tests for comparing the Type across the same conditions of degree, resulted in highly significant differences (*p < .001*), but no Kruskal-Walis test survived correction for multiple testing. This is corroborated by the mean value plots in the Fig. SM1 and SM2 in the Supplementary Material, which show either similar values throughout for more frequent styles or wide and overlapping confidence Intervals for the less frequent styles.

**Table** : Significant differences in style ratings on the Dislike subscales. Last column, p-values are Holm adjusted.

| Condition | Style | Dislike Subscale | d (Strong - Slight) | Statistic | N (slight) | N (strong) | p (adjusted) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Artist | HipHop | Displeasure | 1.207 | 706.0 | 50 | 65 | 0.000\*\*\* |
| Artist | Pop | Displeasure | 1.168 | 1,797.0 | 112 | 74 | 0.000\*\*\* |
| Artist | Rock | Displeasure | 1.923 | 161.5 | 40 | 42 | 0.000\*\*\* |
| Artist | Schlager | Displeasure | 0.813 | 3,274.5 | 63 | 170 | 0.001\*\*\* |
| Artist | Pop | Not Authentic | 0.705 | 2,838.5 | 112 | 74 | 0.015\* |
| Artist | Schlager | Not Authentic | 0.622 | 3,798.5 | 63 | 170 | 0.020\* |
| Artist | Pop | Social Incongruence | 0.639 | 2,667.5 | 112 | 74 | 0.004\*\* |
| Artist | Rock | Social Incongruence | 1.588 | 242.0 | 40 | 42 | 0.000\*\*\* |
| Artist | Pop | Too Niche | 0.335 | 2,702.5 | 112 | 74 | 0.006\*\* |
| Artist | Schlager | Too Simple | 0.360 | 3,589.0 | 63 | 170 | 0.010\* |
| Style | HipHop | Displeasure | 0.832 | 852.5 | 63 | 53 | 0.001\*\*\* |
| Style | Metal | Displeasure | 0.673 | 1,685.0 | 56 | 95 | 0.021\* |
| Style | Schlager | Displeasure | 0.988 | 3,774.0 | 97 | 173 | 0.000\*\*\* |
| Style | Techno | Displeasure | 0.774 | 1,558.5 | 67 | 78 | 0.004\*\* |
| Style | Traditional | Displeasure | 0.766 | 2,933.0 | 97 | 100 | 0.000\*\*\* |
| Style | Schlager | Social Incongruence | 0.432 | 5,508.5 | 97 | 173 | 0.000\*\*\* |
| Style | Schlager | Too Simple | 0.461 | 5,311.0 | 97 | 173 | 0.000\*\*\* |

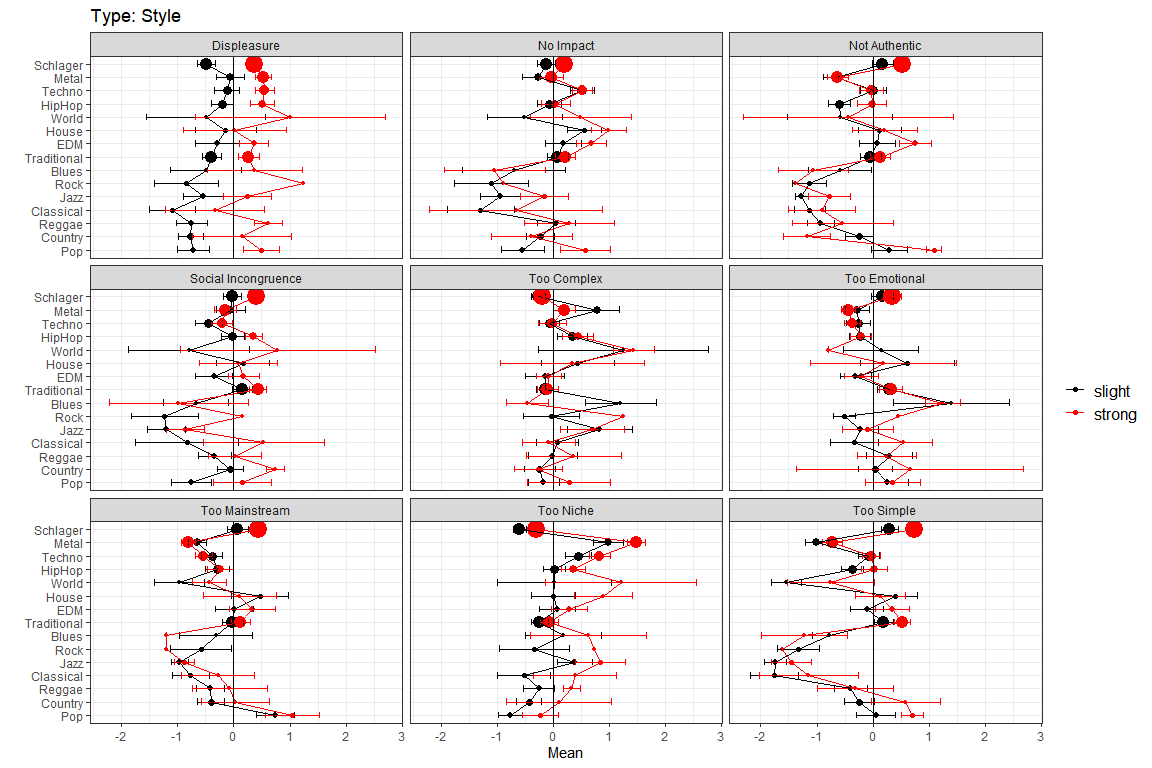


Fig. 6. Mean value of style ratings in the Style main conditions for the Strong/Slight subconditions. Error bars are 95% CI, dot size is proportional to total count of style.

### supplement plot artist

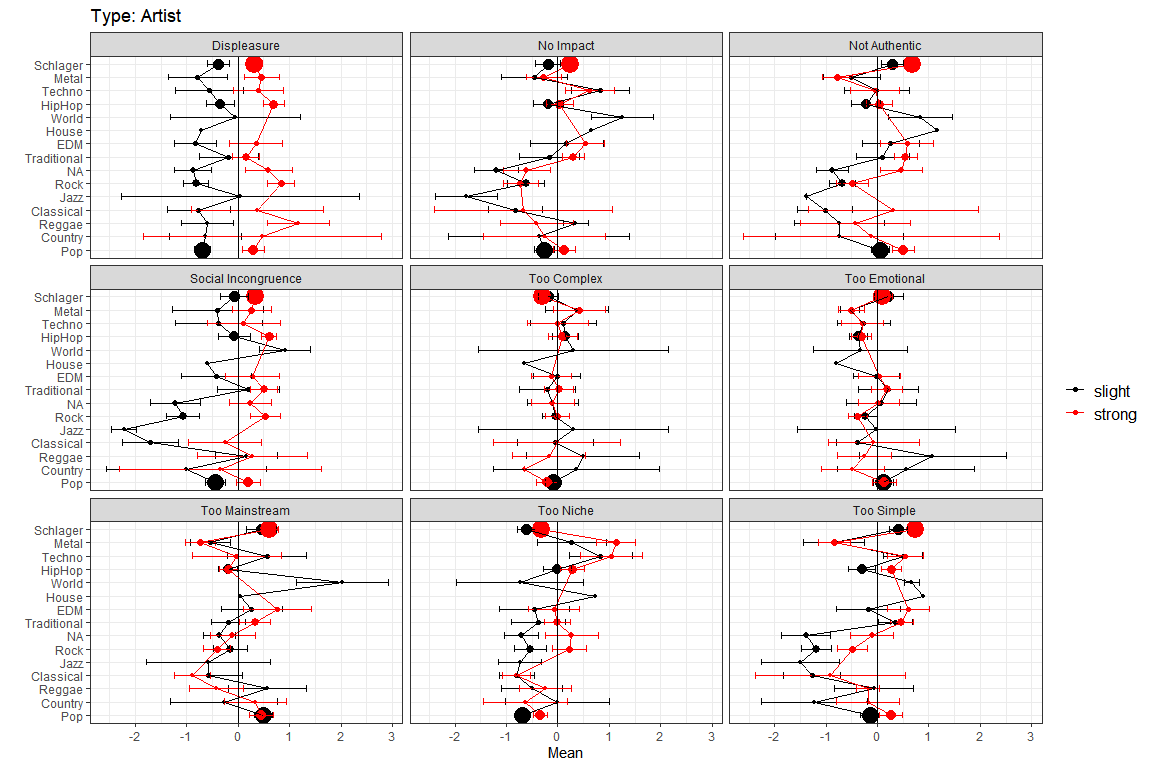


Fig. 7. Mean value of style ratings in the Artist main conditions for the Strong/Slight subconditions. Error bars are 95% CI, dot size is proportional to total count of style.

## Aggreement in style ratings across conditions

We were also interested in how consistent the participants judged the different styles in the different conditions on the nine Dislike subscales. To this end, we binarized the Dislike subscales, by cutting along the midpoint of the scale, i.e., for values above 3, we interpreted the subscale as an attribute that applies, otherwise as an attribute that not applies. We then looked a the distribution of the attributes and generated an overall judgment by a majority vote. If, however, the votes for this attribute (per style and conditions) was not significantly different according to a chi-squared test with a significance level of 1%, we assigned a “mixed” vote. We then checked for every combination of style and attribute whether it was applied in all four conditions. The results can be found in Fig. X.

For example, the ratings for World were consistently discordant (grey, mixed) or not available (white), the majority of the participants dislikes Traditional because of being too simple (red, yes), but does not dislike it because of too complex (black, no). The decision to binarize the data suggested itself through the rating behavior. The bimodal distribution in some items such as not authentic or the skewed distributions in others such as too complex showed that participants already seemed to use the scale in a manner that reflects a binary use that describes an agreement or disagreement rather than a continuous/gradual rating. (Discussion)

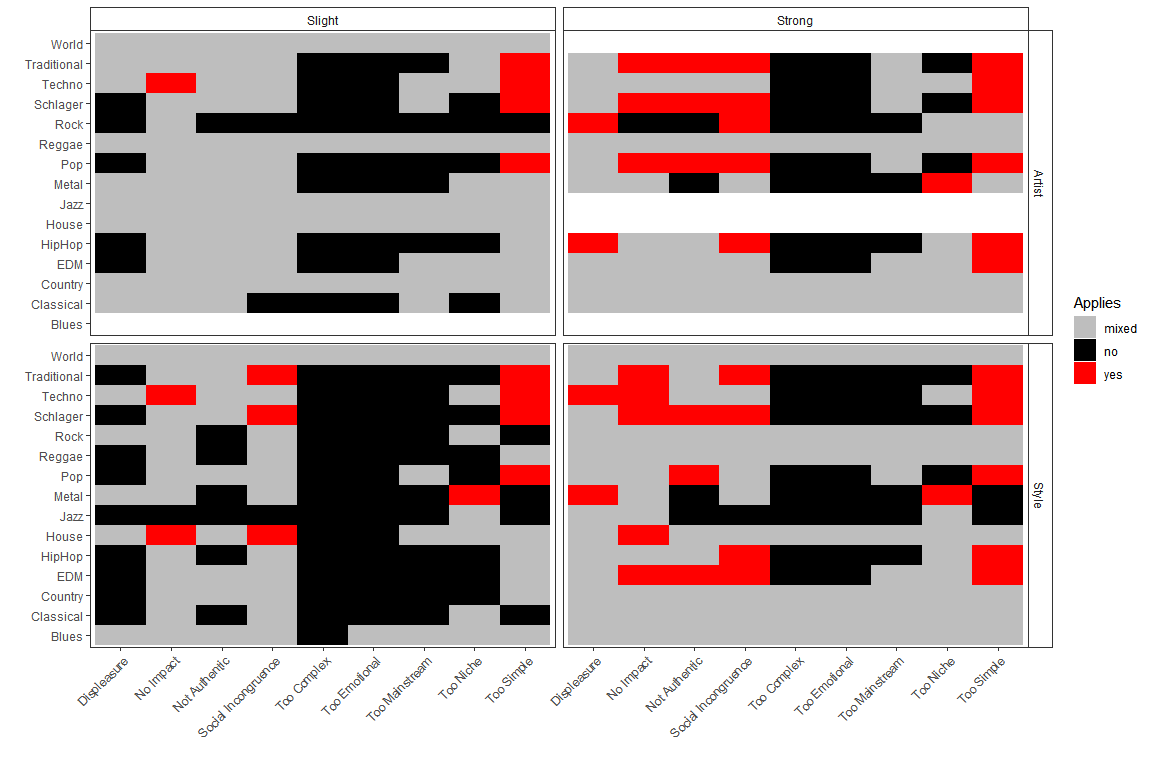


Fig. y, Agreement for the four conditions of styles attributes judgments.

This clearly shows that the judgment of the style in the four conditions was mostly not congruent (white squares), only *Traditional*, *Techno*, *Schlager* and *Pop* were consistently judged as **Too Simple**. Similarly, *Schlager* and *Pop* were never attributed as **Too Niche**, and *Metal* and *HipHop* were never judged as **Too Mainstream**. The rarely used **Too Complex** and **Too Emotional** were never consistently attributed to a style, only for *World*, where the judgements were always mixed on both attributes, and *Blues*, which always received a mixed judgment of **Too Emotional**. This reflects the generally consistently mixed judgement of *World* and *Blues*, which also stems from the fact that these were the least frequently selected styles.

Looking at both, the LPA and the agreement, the explanatory strategies for disliked music are dependent on the condition and the musical style. The profiles and rating patterns are diverse for most styles, but similar for some well-known ones. The question arises whether participants have different conceptions/ideas/impressions about the specific musical styles or dislike them for different reasons. In the case of less common/popular styles such as World or Reggae, the participants refer to different examples or substyles of the music. In more popular/well-known styles such as Schlager, Techno or Metal, the ratings become more similar. (Discussion)

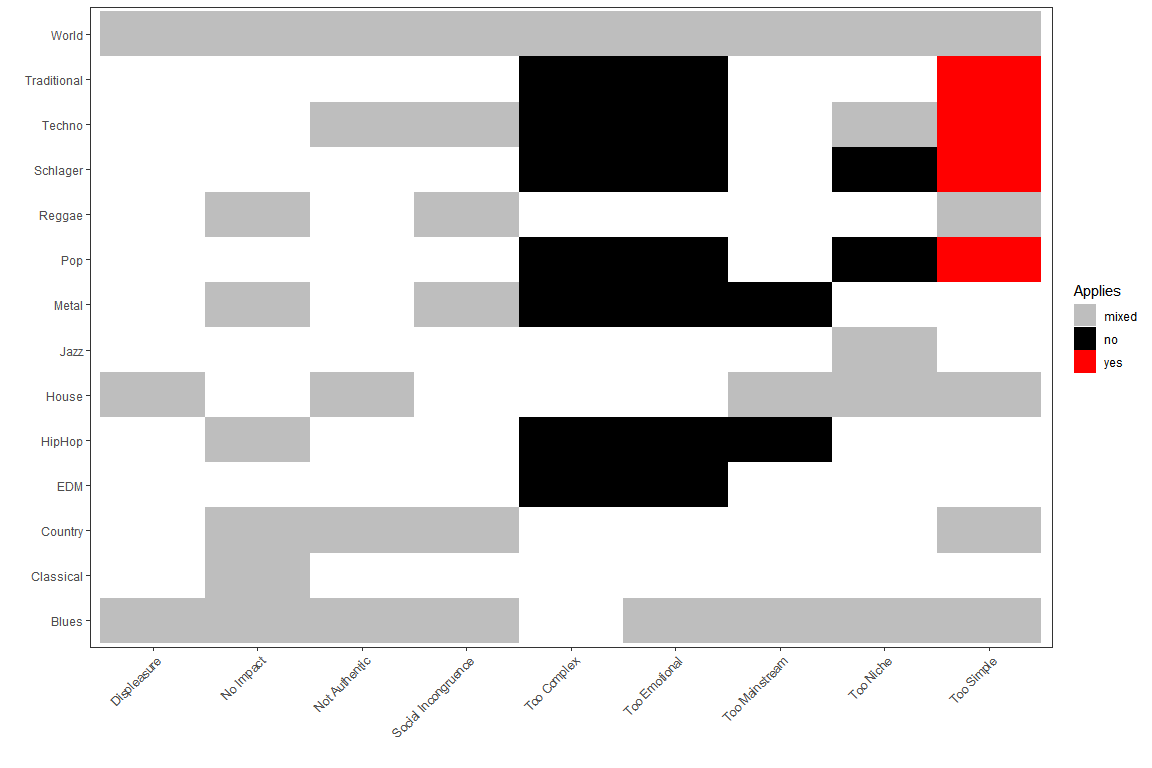


Fig. X, Agreement across conditions of style attributes.

As an alternative route of analyzing inter-rater agreement, we calculated Krippendorff’s Alpa for each style in each of the four conditions, across the nine Dislike subscales. Results can be found in Fig. Y. Styles and conditions are sorted according to the largest mean inter-rater agreement. This basically corroborates the preceding analysis. Only very few values pass the usual .6667 threshold of acceptable agreement. (KF: Very sketchy text… To be improved).

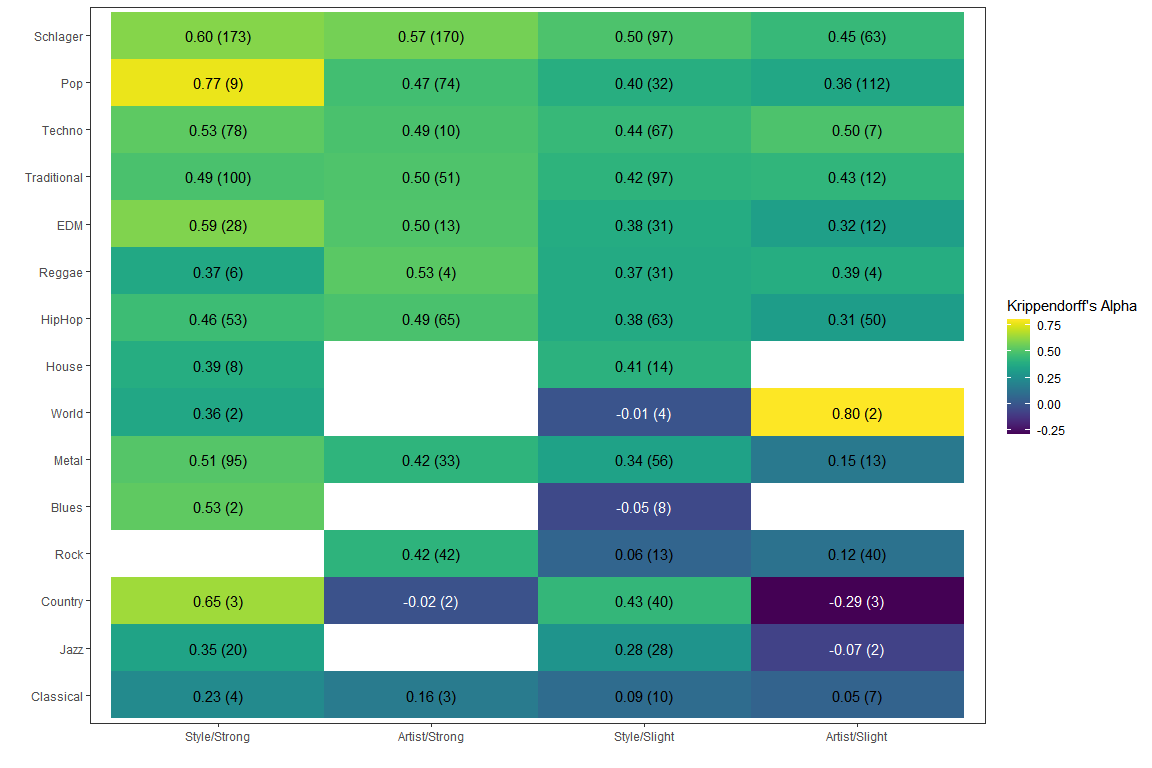


Fig. Y, Krippendorff’s ALpha for Agreement of style ratings across the 9 Dislike subscales, by conditions. Labels are alpha values, number of raters in brackets.

## Differences by covariates

We further checked for differences in style distributions with respect to the demographic variables **Age Group** and **Gender** (dropping the “diverse” gender group) as well as for assigned **LPA Class**. Results of the Chi-Squared tests can be found in Tab. 6. There are strong differences for **Age Group** and **LPA Class** in all four conditions, but only one for **Gender** (in the Style/Strong condition). Chi-Squared residual plots of styles for the Style/Strong conditions according to the demographic variables can be found in Fig. 5a, b and c, from which some noteworthy observations can be made.

With respect to **Age Group**, the younger (less than 30 years) group clearly dislike *Schlager*, *Metal*, and *Pop* more, whereas the elder group dislikes *Traditional*, *Jazz*, *Country*, and *Techno* more. On the other hand, the elder age group are more open to *Schlager*, *Metal*, and *Pop*, whereas the younger group has fewer reservations against *Traditional*, *Techno*, and *Jazz*.

There are some stereotypes to be corroborated in the style preferences of males and females. Males dislike more strongly and often *Schlager*, *Pop*, *HipHop*, and *EDM*, but are more fond of *Jazz*, *Metal*, *Techno*, and also *House*. The female respondents show the clear opposite: more dislikes for styles traditionally often regarded as “virile”, such as *Jazz* and *Metal* as well as *Techno*, and fewer reservations against *Schlager*, *EDM*, *HipHop*, and *Pop*.

The differences for **LPA Class** justify their denomination as *Highbrow* and *Lowbrow* reasoning styles. **Highbrows** have more dislikes for *Schlager*, *Traditional*, *EDM*, and *Pop*, whereas *Lowbrow* profiles more strongly dislike *Jazz*, *Metal*, and *Techno* (see Fig. SC1a-d, SC2a-d, SC3a-d in the Supplementary Material; all chi-squared tests in all four conditions with *p < .001*).

The distribution of **Total LPA Class** is independent from **Age Group** (Chi-squared(1) = 2.93, p = 0.087, Cramer’s V = 0.07), as well as from **Gender** differences (Chi-squared(1) = 0.35, p = 0.551, Cramer’s V = 0.02).

The changes in disliked musical styles with age need cultural/historical considerations, which can here only be speculated about. Overall, the polarity and the discourses about music change over time and with the current findings, we get a glimpse of these changes and can only hint about the reasons. First, Traditional music used to be very present in the German media about 10-15 years ago, which has changed and might explain why people under 30 years of age do not dislike Traditional music as much anymore than people over 30, who grew up with this medial exposure. Second, Jazz seems to have become more hip in recent years, i.e., younger participants dislike it less than older ones. Hence, the perception and exposure of certain styles change, dependent on the media and therefore the familiarity of the styles. (Discussion)

**Table** : Chi-Square tests of differences in style distribution for condition by age group, gender (with diverse), and LPA class. All p-values Holm adjusted.

| Type | Degree | Group | Statistic | Df | Cramers V | P Value Adj |
| --- | --- | --- | --- | --- | --- | --- |
| artist | slight | age\_group | 37.2 | 14 | 0.33 | 0.001\*\*\* |
| artist | strong | age\_group | 45.2 | 11 | 0.30 | 0.000\*\*\* |
| style | slight | age\_group | 25.6 | 14 | 0.21 | 0.029\* |
| style | strong | age\_group | 48.8 | 14 | 0.29 | 0.000\*\*\* |
| artist | slight | gender | 17.4 | 14 | 0.22 | 0.236 |
| artist | strong | gender | 16.3 | 11 | 0.18 | 0.129 |
| style | slight | gender | 6.8 | 14 | 0.11 | 0.942 |
| style | strong | gender | 36.6 | 14 | 0.25 | 0.001\*\*\* |
| artist | slight | lpa\_class | 51.5 | 14 | 0.38 | 0.000\*\*\* |
| artist | strong | lpa\_class | 99.1 | 11 | 0.45 | 0.000\*\*\* |
| style | slight | lpa\_class | 77.7 | 14 | 0.36 | 0.000\*\*\* |
| style | strong | lpa\_class | 199.8 | 14 | 0.59 | 0.000\*\*\* |