Measuring confusion and consensus in social evaluations of PC video games: An empirical test of mechanisms on both sides of the producer-audience interface

Andrew J. Argue

Department of Business Management

National Sun Yat-sen University

Thijs A. Velema

Department of Business Management

National Sun Yat-sen University

Since market positioning by producers is so commonly studied as an antecedent of organizational outcomes from legitimacy to product appeal, it would be useful to know more about why category-spanning products are associated with more negative social evaluations. This relationship is typically understood using either a cognitive or a social approach (Durand, Granqvist, and Tyllstrom 2017). According to the first perspective, producers who span categories are penalized because atypical categorization elicits cognitive confusion (Zuckerman 1999; Phillips and Zuckerman 2001). The second perspective sees such penalties as the result of lower social consensus among the larger and more diverse audiences that atypical offerings attract (Hsu 2006; Pontikes 2012). Yet, while much research has demonstrated that organizations and products with more typical identities are associated with more positive social evaluations, not much empirical study has been devoted to empirically testing the theoretical mechanisms underlying such processes.

Our goal in this paper is twofold: we seek to both empirically validate and disentangle two important tenets of the categorization and social evaluation literature by analyzing the separate effects of confusion and consensus in relation to category spanning. We present a novel approach to measuring confusion and consensus among audiences that accounts for interactions between social actors on both sides of the producer-audience interface—an under-studied area of research (Grodal and Kahl, 2017). The proliferation of information sharing through new technologies such as social networks and online product market platforms (Barlow, Verhaal, and Angus 2019) suggests that a contemporary theory of social evaluations should account for how judgments are formed within an increasingly interconnected collectivity of social actors. With this in mind, we propose that cognitive confusion can be measured as the (dis)similarity between the categorical labels applied by producers and audiences to the same products while audience consensus can be captured using the information entropy of the distribution of audience-applied labels.

The empirical setting for our investigation is the most popular online market platform for PC video games known as Steam. Online platforms and review websites that apply categorical labels to products have recently accelerated research on categories (Kovacs and Hannan 2015) and have been used in a number of recent studies on categorization and product appeal (e.g., Barlow, Verhaal, and Angus 2019; Kovács et al. 2013, Verhaal, Khessina, and Dobrev 2015; Zhao et al. 2018; Zhang and Liu 2017). Steam uses a similar social-tagging system whereby, in addition to specifying which genres they are targeting, producers apply up to 20 specific category tags to their upcoming games. One important difference is that once a game is released on the Steam, audience members can collectively change which tags appear for each game by voting for their own. By leveraging this unique property of Steam’s tagging system, we are able to analyze the relationships between the typicality of producers’ strategic categorizations and characteristics of audiences’ collective categorizations, such as their resulting degree of confusion and consensus.

The results of our analyses support the idea that initial atypical positioning of video games by producers is associated with a greater confusion and less consensus. Due to data limitations researchers have typically been constrained by sets of broad genres and have typically relied on the evaluations of critics as a proxy for audience members (Hsu 2006, SOURCES). In the present study, we present findings based on a much larger vocabulary set (425 possible labels) and use direct measures of categorization behaviors based on the labels applied by end-users themselves. Thus, our results not only provide much needed validation for these commonly relied on tenets of social evaluation literature, but also the more fine-grained nature and temporal ordering of our observations provides strong additional support for a causal argument involving strategic categorization by producers (Pontikes and Kim, 2017). Our research is motivated by recent calls from social evaluation scholars who advocate for more research into discursive approaches that account for the interactions between social actors (Durand, Granqvist, and Tyllstrom 2017). Cognitive confusion and social consensus represent two increasingly diverging perspectives in the categorization and social evaluations literature and here we make some of the first attempts at disentangling these mechanisms. By presenting novel methods to examine the idea that social penalties incurred from category-spanning are due to the effects of confusion and consensus, our study contributes to a better understanding of the construction and maintenance of category meaning in social evaluations.

**Audiences, legitimacy and product appeal**

Audiences control the allocation of market resources and ultimately decide which organizations are legitimate members of market categories. Organizations that are perceived as illegitimate by relevant audiences are not only less appealing but have lower success and survival rates (Dobrev, Kim, and Hannan 2001; Zuckerman 1999). This is because audience members associate greater uncertainty (Zuckerman 1999) devotion of resources (Hsu 2006), lack of skill (Hsu, Hannan, and Polos 2011) or quality (Leung and Sharkey 2014) with illegitimate organizations and their products. Producers that fail to conform to typical patterns of market categorization are therefore subjected to social and economic penalties by audiences, which further reinforces the existing categorical market structures.

Recently, scholars operating at the intersection of organizational theory and strategic management have also begun to acknowledge that producers are not simply passive actors in markets (Pontikes and Kim, 2017). Producers can actively claim membership in certain categories (Vergne 2012) to potentially exert an influence on the formation of audiences’ perceptions about their membership in market categories (Zhao et al. 2018). Such strategic positioning is seen as a signal that producers use to influence the way other social actors categorize them (Durand, Grandqvist, and Tyllstrom 2017). Through strategic positioning within market categories producers also attempt to gain legitimacy through social evaluation (Pontikes and Kim 2017). Affiliation to a category can be claimed through the language a producer uses (Hsu and Hannan 2005) such as through descriptions of their products (Barlow, Verhaal, and Angus 2019) or even naming products (Verhaal, Khessina, and Dobrev, 2015).

Perceptions of an organization or product’s level of category membership are formed through the construction of categorical schemas of characteristics that external audiences associate with individual market categories (Hsu, Hannan, and Kocak 2009). Those that do not clearly conform to a specific category schema will be ignored (Zuckerman 1999; Zuckerman and Philips 2001) or devalued (Hsu, Roberts, and Swaminathan 2012; Hsu, Hannan, and Kocak 2009; Leung and Sharkey 2014; Pontikes 2012). More favorable social evaluations and greater appeal to audiences are associated with social objects whose identities are clearly positioned in one category (Hannan et al. 2019; Hsu 2006; Hsu, Hannan, and Kocak 2009; Negro and Leung, 2013; Leung and Sharkey 2014).

For organizations and their products, the relationship between membership in market categories and the appeal to social evaluators is understood using two main approaches. The first suggests that evaluations by audiences are formed through cognitive comparisons to existing categorical prototypes. From this perspective, category spanning makes it more difficult for audience members to compare offerings to each other, or to category prototypes, increasing the cognitive burden associated with such evaluations (Durand and Boulongne 2017). The cognitive confusion associated with making judgements about such offerings’ category membership leads to more negative evaluations of legitimacy and appeal. Conversely, a recent strand in the literature takes a discursive or social approach to understanding the consequences of atypical market positioning (Durand, Grandqvist, and Tyllstrom 2017; Grodal and Kahl, 2016). From this perspective, evaluations are seen as the result of a negotiation of meanings by a collectivity of social actors. The clarity of an offering’s identity, or category fit, is determined by the level of agreement between multiple evaluators’ perceptions.

**Category-spanning, cognitive confusion, and social consensus**

The first way that category-spanning by producers is understood to lead to more negative social evaluations is that it increases the cognitive burden associated with evaluating offerings. According to Paolella and Durand (2016), “Most research invokes a cognitive confusion mechanism to explain the association between category spanning and lower evaluation” (p. 342). This explanation was first introduced by Zuckerman (1999) in his seminal paper on the categorical imperative. His reasoning is based on the process typically employed by consumers in the social evaluation of producers and their products. Phillips and Zuckerman (2001) describe this in more detail as a selection process consisting of two phases: the construction of a consideration set of perceived legitimate offerings, followed by a comparison between alternatives in the set. Products whose features deviate too much from the category prototype are difficult for audiences to compare to other offerings (Barlow, Verhaal, and Angus 2019).

Organizational theorists have expanded on these ideas to further explain that audiences make sense of markets by classifying producers into market categories by making cognitive legitimacy judgments based on cues such as organizational characteristics or features (Bitektine 2011). Classifications that use multiple labels from different categories introduce organizational-level ambiguity that results in confusion among audiences (Pontikes 2012). Increased confusion may occur since category spanning sets off what Tost (2011) refers to as a social actor’s neural alarm system. This causes them to question legitimacy judgements based on taken-for-granted category membership claims by producers. Thus, when audiences try to make sense of such contradictory membership cues involving multiple labels they are more likely to make active versus passive legitimacy judgments (Bitektine and Haack 2014; Tost 2011) and experience greater cognitive confusion as a result (Leung and Sharkey 2014).

Accordingly, recent research by Zhao et al. (2018) suggests that in markets such as the video game industry, which are characterized by fluid and constantly emerging categories, similarity clustering and pair-wise comparisons involved in the evaluation of offerings puts an especially demanding cognitive burden on evaluators. Because audiences only have limited cognitive resources at their disposal, they are likely to exclude ambiguous offerings from the first phase of consideration and devote more effort to evaluating offerings that are more clearly legitimate category members (Leung and Sharkey 2014). That is, deviation from the category schema fails to evoke a clear market category that consumers perceive an offering fits into (Tan and Roberts 2010) making its categorization and evaluation more confusing.

If strategic categorization behaviors of producers are truly able to exert an influence on the social evaluation behaviors of audiences, then by comparing the categorization behaviors of producers and audiences, i.e., the category labels they use to describe the same products, it should be possible to observe the extent to which category spanning by producers affects the level of audience confusion. Specifically, we argue that when the labels used by producers to signal the market positioning of products are more typical of the target category schema, the audience is more likely to find that such combinations of features evoke a familiar category and that their categorizations will resemble producers. Conversely, if the category labels an audience applies are more dissimilar, it can be taken as evidence that they struggle more with classifying products, indicating a higher degree of confusion about the product’s categorical fit. Thus, we expect that when producers position new video games more typically in relation to the schemas of their targeted genres, they will be more easily comparable, reducing the cognitive burden and producing less cognitive confusion among audience members. Since we represent confusion as a lower similarity between categorization behaviors of two actors our first hypothesis is as follows:

Hypothesis 1: There is a positive relationship between the typicality of a product’s producer-applied category tags in relation to the target category schema and the similarity between the tags that producers and audiences use to categorize the same products.

The social perspective on category-spanning penalties sees the relationship between atypical classification and lower appeal as a result of such offerings producing dissensus within audiences in relation to a product’s identity. Such an approach sees the shared meaning of market category membership as created through discourse among multiple market actors such as consumers, producers, market intermediaries and regulators (Durand, Grandqvist, and Tyllstrom 2017; Grodal and Kahl, 2016; Lehman et al. 2019; Pontikes 2012). A relevant audience is made up of individuals who have competing schemas, i.e., differing views on which features are the most salient signals of category membership, which means that they vary in terms of their assessments of an offering’s level of fit within market categories (Glynn and Navis 2013). Spanning categories attracts a larger, more diverse audience (Hsu, 2006) and thus the constituent members’ perceptions about which qualities a product must possess to be a member of a particular category will become more variegated (Hsu, Hannan, and Kocak 2009).

Audience consensus has a strong positive impact on organizational outcomes, e.g., appeal (Hsu 2006) or brand popularity (Zhang and Liu 2016). Much research has shown that when producers violate audiences’ categorical expectations it creates a lack of consensus among social actors (Carroll and Swaminathan 2000; Granqvist, Grodal, and Woolley 2013; Hsu 2006; Phillips and Kim 2009; Waguespack and Sorenson 2010; Zhao, Ishihara, and Lounsbury 2013). Such violations are caused when the market positioning by producers sends signals that deviate from the target category’s schema. Schemas can be described as a set of expectations about what features a member of that category should exhibit and which they shouldn’t (Hsu and Hannan 2005). By attracting larger, and more diverse audiences (Hsu 2006), the diversity of competing schemas used to evaluate an organization or a product also increases. As the multiplicity of competing category schemas grows there is less agreement between actors about which features are salient for judging membership (Negro, Hannan, and Rao 2011) and this hinders the development of audience consensus.

When producers understand the prototypical category schema of the genres they are targeting, they may strategically position their products by signaling the most salient features of schemas to audiences (Hannan et al. 2007; Pontikes and Hannan 2014; Pontikes and Kim 2017). That is, they may seek to actively shape audience perceptions about the market identities of offerings and influence social evaluations (Cattani, Porac, and Thomas 2017; Cattani, Ferriani, Negro, and Perretti, 2008; Zhao, Ishihara, and Lounsbury 2013). The causal argument involving strategic positioning and audience consensus (Zhang and Liu 2017) is central to how the discursive social perspective explains the negative consequences of category spanning. If video game producers can influence the level of consensus in audience evaluations by strategically positioning products, then those that are more typical of the target category schemas should attract smaller, less diverse audiences and be less likely to violate category expectations, on average. In such cases the relative homogeneity of such audiences should mean that the category labels applied to games should also be more homogeneous, indicating a higher degree of social consensus. Our second hypothesis follows from these considerations:

Hypothesis 2: There is a negative relationship between the typicality of a product’s producer-applied category labels in relation to the target category schema and the information entropy of the distribution of category tags applied by audiences.

**Data and methods**

To test whether producer typicality, i.e., conformity to the schemas of targeted categories, leads to less confusion and greater consensus among evaluating audiences, we study a sample of video games released in 2017 on Steam—the largest online market platform for PC video games. The global market for PC video games is the third largest market, behind mobile and console video games, and is estimated to be worth approximately $39.5 billion (Nesterenko 2020). Steam is the largest online market platform for video games accounting for around 18% of all PC video games sold in 2017 (Bailey 2018). With such rapid growth in the popularity of video games, an increasing number of scholars are now using this research context for scholarship on the categorization behaviors of producers and audiences. Studies on optimal distinctiveness in the console video game market (Barlow, Verhaal, and Angus 2019; Zhao, Ishihara, and Lounsbury 2018), platform market competition (Cennamo and Santalo 2013) and genre classification (Windleharth et al. 2016; Li and Zhang 2020) are some recent examples. Steam tags are similar to labeling systems used on online review websites such as goodreads.com and yelp.com studied recently by Kovacs and Hannan (2015). In another recent study of online social-tagging of digital products by Zhang and Liu (2017) they argue that, “information contained in social tags can reveal information, such as a brand’s niche width (the level of diversity of a brand’s targeted categories) and audience consensus” (p. 106).

The Steam online platform uses a social-tagging system that allows both producers and end-users to apply descriptive labels to the same games. We use historical data on the tags that producers and consumers applied to games to directly observe the categorization behaviors of producers and end-users at two points in time: before games are released and one year after the initial release. Our dataset was constructed using two sources that aggregate data collected from Steam’s own application programming interface: steamdb.info and steamspy.com. Historical data on the tags associated with games obtained from these sources enabled us to observe the tags applied to games before and after their release. Our final sample consists of 1,142 games categorized into 11 major genres, with 425 possible tags.

*Dependent variables*

Following the methods used in recent studies (e.g., Fiss, Kennedy, and Davis 2012; Wang et al. 2019) we conceptualize collections of tags as concept-networks consisting of nodes (tags) connected by edges (co-occurrences among groups of tags). Since we are interested in testing two separate mechanisms, we measure how category-spanning by producers influences audience confusion and audience consensus using two different dependent variables. Our first dependent variable represents audience confusion and is used to test our first hypothesis, that there is a positive relationship between the typicality of producer tags and the similarity between the tags that producers and audience use to categorize the same game. We represent similarity by calculating the Jaccard index (Jaccard 1901) of two sets of tags applied to games 1) those applied by producers before their release and 2) those applied by consumers after one year. The Jaccard similarity index is calculated as:

where indicates tags applied by both producers and audiences, and where indicates tags applied by either producers or audiences. As this variable is intended to represent confusion, it is important to note that higher levels of similarity represent less confusion.

Our second dependent variable is used to test hypothesis 2 and represents the level of audience consensus. Hannan et al. 2019 propose that the entropy (Shannon 1948) of the vector of a social object’s categorization probabilities can measure the conceptual ambiguity of a concept. Following this, we measure the level of consensus among audience members by calculating the unpredictability with which audiences apply tags, i.e., the information entropy associated with the distribution of audience tags at the one-year mark after a game’s release. Shannon entropy is defined as:

where N is the number of possible tags audiences can apply and is the vector of probabilities for each tag based on the distribution of counts associated with each game. Entropy is usually interpreted as the amount of uncertainty in a process or the degree of heterogeneity of a probability distribution, so high entropy would indicate a high degree of heterogeneity (dissensus). Thus when interpreting this measure of consensus, it is important to note that more homogeneity among tags that audiences apply (consensus) is indicated by lower entropy. To demonstrate the relationships between distributions of audience tags and levels of entropy, we visualize the distributions of tags for four different games at low and high levels of entropy below in Figure 1.

---- INSERT FIGURE 1 ABOUT HERE ----

*Independent variable*

To study how category spanning by producers can influence the level of confusion and consensus in social evaluations we treat the typicality of producer-applied tags in relation to the prototypical target category schema as our main independent variable. In addition to a list of tags, Steam allows a producer to indicate up to 10 target genres which we interpret as their target market categories. We measure the level of conformity to targeted genres for each game following the methods of Kovacs and Johnson (2014)[[1]](#footnote-1). For each vector of producer-applied tags we sum the weighted average of Jaccard similarities of tag-category pairs for each game in each genre that they target.

*Control variables*

As an online market platform, Steam provides a number of additional sources of information that allow us to control for genre- and product-level variation. First, similar to films (Zhao, Ishihara, and Lounsbury 2013), there are significant differences in the production budgets of PC video games, and we account for such differences by controlling for price. Second, following Kovacs and Hannan (2019) and Hsu (2006), we use the number of owners for each game as predicted by Steamspy to control for the level of consumer engagement[[2]](#footnote-2). Also, using these ratings of reviews we are able to control for quality and disentangle the commonly co-occurring effects of quality and spanning (Leung and Sharkey 2014).

**Results**

The summary statistics and correlation coefficients for all variables used in our analyses are shown in Table 1. Most notably, our two dependent variables representing confusion and consensus are highly correlated (r = -0.78, p < 0.001). This is in line with our expectations since our hypotheses predict that greater typicality is associated less confusion (smaller distance between producer and audience categorizations) and greater consensus (lower information entropy of audience-applied tag distributions. Table 1 also shows strong correlations between our independent variable, typicality, and both our dependent variables: similarity of producer-audience tags (r = 0.57, p < 0.001), and entropy of audience tags (r = -0.52, p < 0.01). To illustrate the relationships between our variables of interest we plot these bivariate relationships below in Figure 2 and Figure 3.

---- INSERT TABLE 1 ABOUT HERE ----

---- INSERT FIGURE 2 ABOUT HERE ----

---- INSERT FIGURE 3 ABOUT HERE ----

*Multivariate analysis*

Table 2 shows the estimated coefficients of OLS models used to test our two hypotheses about the relationships between the typicality of producer tags, the similarity between producer and audience tags and the information entropy of audience tags. Model 1 and Model 2 test our first hypothesis and use the Jaccard similarity between producer and audience tags as the dependent variable. Model 1 is a baseline model that includes only control variables. Model 2 adds our measure of producer typicality. Adding typicality to the model increases the r-squared from 0.12 to 0.37, indicating that typicality helps to explain more of the variance in producer-audience tag similarity. The estimated coefficient for typicality in Model 2 of 2.34 can be interpreted to mean that for each additional 1% increase in typicality, the Jaccard similarity between producer and audience tags increases (i.e. confusion decreases) by 2.34%. The positive and significant (p < 0.01) coefficient for typicality of producer tags in this model thus supports hypothesis 1.

---- INSERT TABLE 2 ABOUT HERE ----

As an initial test for the effects of typicality on audience consensus we compared the typicality of producer tags to the total number of reviews that games received. Higher consensus about a product’s identity increases demand for a product (Zuckerman, 1999) partly because a producer with a larger niche-width, or product with an atypical identity, tends to attract a larger, more diverse audience (Hsu, 2006: Zhang and Liu 2016). Such a relationship has also been shown to exist in the context of social movements where boundary crossing attracts larger audiences of supporters (Benford 1993; Heaney and Rojas 2014). We found that there is a significant, negative correlation (r = -0.13, p<0.01), between the typicality of tags applied to games by producers and the size of the audience they attract. This is consistent with the relationship between typicality and audience size found in previous literature and provides some initial support for hypothesis 2. The bivariate relationship between these two variables is shown below in Figure 4.

---- INSERT FIGURE 4 ABOUT HERE ----

Model 3 and Model 4 test our second hypothesis and use the entropy of the distribution of audience tags as the dependent variable. Model 3 is a baseline model that includes only control variables and in Model 4 we add our measure of producer typicality. Adding typicality to the model increases the r-squared from 0.22 to 0.39, indicating that typicality helps to explain more of the variance in the entropy of the distribution of tags that the audience applies. The estimated coefficient for typicality in Model 4 is -14.91. This value can be interpreted to mean that for each additional 1% increase in typicality, the value for entropy decreases (or consensus increases) by 0.149. To put this into perspective, a 6% increase in typicality of producer tags would increase audience consensus by about one standard deviation. The negative and significant (p< 0.01) coefficient for the typicality of producer tags in Model 4 thus supports hypothesis 2.

**Discussion and conclusion**

The goal of this study was to examine two fundamental propositions in the literature about how category spanning by producers leads to lower appeal among audiences. The two main approaches to explaining penalties associated with category spanning rely on either a cognitive confusion or a social consensus explanation. We tested for the separate effects of these two social constructs by analyzing how producers and audiences apply categorical tags to video games on the Steam online market platform. Using a novel empirical approach that exploits the unique properties of Steam’s social-tagging system, we measured the degree of confusion and consensus by comparing the categorization behaviors of both producers and audiences for the same products at two different points in time. Our results provide evidence in support of our hypotheses relating to the association between the typicality of the tags that producers apply to their own games and the level of confusion and consensus among audiences who engage with them. Specifically, we find that when producers label their own games using tags that are more typical of the schemas associated with their targeted genres, the categorization behaviors of audience members are characterized by a lower degree of confusion and a higher degree of consensus.

These results make a contribution to the categories and social evaluations literature in a number of ways. First, our results provide additional empirical validation for the argument that it is beneficial for producers to strategically position their own products strategically in order to conform to the targeted categorical schema. Literature on strategic behavior among producers (Pontikes 2017; Porac et al. 1995; Rao, Monin, & Durand 2003) suggests that producers can use categories to frame their offerings (Carroll and Swaminathan 2000; Granqvist, Grodal, and Woolley 2013; Navis & Glynn 2010). In their recent review of the optimal distinctiveness literature, Zhao et al. (2017) suggest that, “more systematic research on how managers manipulate different strategic levers to enhance perceptions of strategic positioning is required.” (p. 98). Our study makes such a contribution by providing additional empirical evidence supporting the argument that producers can shape both the level of confusion and consensus among audiences through the typicality with which they frame their products.

Another contribution of our present study is that by leveraging a new source of categorization data we incorporate behaviors of multiple social actors into our analysis to present new measurements of common social evaluation mechanisms. We believe the measures we use have several advantages and can be added to the toolkit of traditional methods currently employed by scholars. For instance, since Hsu (2006) provided some of the first empirical evidence of the association between producer niche-width and consensus, this social evaluation construct has typically been measured as the similarity between categorizations of products by two or more critics. By contrast, we measure consensus as the entropy of the distribution of caregory labels that an audience uses. Such a measure is based on the unadulterated categorizations of end-users and accounts for the collective perceptions of categorical fit for up to thousands of consumers. The data we use also has the advantage of capturing more unconstrained categorization behaviors since Steam users can apply any novel tag they want rather than choosing from a small set of designated genres.

By both incorporating for a greater number of actors, and accounting for interacitons between actors on both sides of the producer-audience interface, we also respond to recent calls in the literature that encourage increased focus on the social processes involved in categorization and social evaluation (Durand, Granqvist, and Tyllstrom, 2017). The discursive perspective views social actors as engaging in conversations through their mapping of products into categories. It is thus through such communicative exchanges and or discourse that the mental representations that make up market categories emerge and evolve (Grodal and Kahl, 2017). By both paying more attention to the roles of actors involved in evaluations and observing the temporal order of categorizations, our study presents and approach that is thus more in line with a modern, discursive approach to social evaluations that sees the creation and maintenance of categorical meaning in markets as a negotiation between a collectivity of social actors.

Research that uses cognitive confusion to explain category-spanning was originally inspired by work in cognitive psychology on prototypes (Rosch and Mervis 1975). In more recent studies however, cognitive confusion is typically inferred from associations found between atypicality and appeal. One further benefit of our study is that it begins to reconcile this issue by introducing a measure of confusion that is based on a direct measure of similarity between the categorizations of producers and audiences for the same products and tests the effects of typicality on both confusion and consensus as outcomes using the same set of category labels. This takes multiple actors into account, with the additional benefit of incorporating perceptions from both sides of the producer-audience interface, providing a new avenue for future examinations into how negotiations between different social actors occur through categorization in markets.

Finally, our study also contributes to a better understanding of how confusion and consensus relate to one another. While some researchers invoke cognitive confusion and consensus as separate mechanisms, others use them almost interchangeably, or even use one to explain the other. For instance, in their discussion of the results of their analysis of the relationship between niche-width, consensus and brand popularity, Zhang and Liu (2016) say, “although most brands want to target multiple categories to attract more consumers … stretching to too many categories may create confusion among consumers and, thereby, decrease their appeal.”. Whether confusion and consensus are indeed two separate mechanisms or simply “two sides of the same coin” is an important issue to be resolved if scholars continue to rely on these mechanisms to explain category-spanning discounts. While our analysis in the present paper by no means presents conclusive causal evidence to support one of these perspectives in favor of the other, it is one of the first empirical studies we are aware of that attempts to disentangle these two separate effects. This paves the way for future investigations into this important, yet understudied issue.

To be certain, our study has a number of limitations that must be acknowledged. First, although we measure the relationship between typicality, confusion and consensus, due to data limitations, we are unable to relate these measures to performance or appeal. Unfortunately, the unreliability of sales, ownership and critical acclaim data for our sample means that it was not possible to accurately perform such an analysis. Thus, in the present study we were unable to reproduce the findings of previous studies in the literature and construct models with audience appeal as an outcome variable. Future studies could try to match social-tagging data from sources such as Steam with reliable sales or other performance metrics that would allow for a deeper examination of mediation and moderation effects involving our variables of interest.

Another drawback is that, while our measurement methods are novel, there is admittedly, little precedence in the categorization or social evaluations literature for how we operationalize our dependent variables. While Jaccard similarity is well-established as a method of measuring agreement between the categorization behaviors of critics to judge consensus, whether or not the similarity between producer and audience categorizations is an accurate representation of cognitive confusion remains an open question. Future studies could expand on the methods used in the present paper by conducting experiments more in line with cognitive psychological research, but in the context of product categorization in order to perform further tests of the measurements we propose here.

Similar limitations apply with respect to how we operationalize audience consensus and the conclusions that can be drawn from our results. Entropy has been employed in other related contexts to measure constructs such as product diversity (Alexander 1996), heterogeneity of topic coverage (Kuhlen and Preston 2021) or the amount of agreement versus dissention in user ratings (Tastle and Wierman, 2007). One drawback of using an entropy measure, however, is that it uses an unrestricted scale, which means that cross-comparison to other contexts is not possible (Wang, Rao, and Soule, 2019). While Hannan et al. (2019) argue that entropy is a widely used and appropriate measure of cognitive ambiguity associated with the categorization of social objects, there is also no precedence that we are aware of in the extant literature for using entropy to measure the level of audience consensus. Future studies could explore how this measure of consensus using entropy compares to other more traditional methods such as the level of similarity between categorizations of separate market intermediaries.

Notwithstanding these limitations, in this paper we use a novel approach and new data to provide validation for two important and understudied mechanisms in the extant literature. Our results demonstrate that atypical products are associated with both a higher degree of confusion and lower degree of consensus in terms of how audiences categorize them. By presenting additional empirical evidence and illustrating new ways of measuring the relationships between typicality, confusion and consensus our study provides insight into these important social evaluation constructs. This contributes to a better understanding of how category spanning leads to lower appeal and how meaning is created and maintained through interactions between social actors in product markets.

**References**

Alexander, P. J. (1996). Entropy and popular culture: product diversity in the popular music recording industry. *American Sociological Review, 61*(1), 171-174.

Bailey, D. (2018). With $4.3 billion in sales, 2017 was Steam's biggest year yet. PCGamesN. Retrieved from: https://www.pcgamesn.com/steam-revenue-2017

Barlow, M. A., Verhaal, J. C., & Angus, R. W. (2019). Optimal distinctiveness, strategic categorization, and product market entry on the Google Play app platform. *Strategic Management Journal, 40*(8), 1219-1242.

Benford, R. D. (1993). Frame disputes within the nuclear disarmament movement. *Social Forces, 71*(3), 677-701.

Bitektine, A. (2011). Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status. *Academy of Management Review, 36*(1), 151-179.

Bitektine, A., & Haack, P. (2015). The “macro” and the “micro” of legitimacy: Toward a multilevel theory of the legitimacy process. *Academy of Management Review, 40*(1), 49-75.

Carroll, G. R., & Swaminathan, A. (2000). Why the microbrewery movement? Organizational dynamics of resource partitioning in the US brewing industry. *American Journal of Sociology, 106*(3), 715-762.

Cattani, G., Porac, J. F., & Thomas, H. (2017). Categories and competition. *Strategic Management Journal, 38*(1), 64-92.

Cattani, G., Ferriani, S., Negro, G., & Perretti, F. (2008). The structure of consensus: Network ties, legitimation, and exit rates of US feature film producer organizations. *Administrative Science Quarterly, 53*(1), 145-182.

Dobrev, S. D., Kim, T. Y., & Hannan, M. T. (2001). Dynamics of niche width and resource partitioning. *American Journal of Sociology, 106*(5), 1299-1337.

Durand, R., & Boulongne, R., (2016). Advancing research on categories for institutional approaches of organizations. In R. Greenwood, C. Oliver, T. Lawrence, & R. E. Meyer (Eds.), *The SAGE Handbook of Organizational Institutionalism (2nd ed.)*. Oxford UK: Sage.

Durand, R., Granqvist, N., & Tyllström, A. (2017). From categories to categorization: A social perspective on market categorization. In R. Durand, N. Granqvist and A. Tyllström (Eds.), *From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads* (Vol. 51, pp. 3-30): Emerald Publishing Limited.

Fiss, P. C., Kennedy, M. T., & Davis, G. F. (2012). How golden parachutes unfolded: Diffusion and variation of a controversial practice. *Organization Science, 23*(4), 1077-1099.

Glynn, M. A., & Navis, C. (2013). Categories, identities, and cultural classification: Moving beyond a model of categorical constraint*. Journal of Management Studies, 50*(6), 1124-1137.

Granqvist, N., Grodal, S., & Woolley, J. L. (2013). Hedging your bets: Explaining executives' market labeling strategies in nanotechnology. *Organization Science, 24*(2), 395-413.

Grodal, S., & Kahl, S. J. (2017). The discursive perspective of market categorization: Interaction, power, and context. In R. Durand, N. Granqvist and A. Tyllström (Eds.), *From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads* (Vol. 51, pp. 151-184): Emerald Publishing Limited.

Hannan, M. T. (2010). Partiality of memberships in categories and audiences. *Annual Review of Sociology, 36*, 159-181.

Hannan, M. T., Le Mens, G., Hsu, G., Kovács, B., Negro, G., Polos, L., …Sharkey, A. J. (2019). *Concepts and categories: Foundations for sociological and cultural analysis.* New York: Columbia University Press.

Hannan, M. T., Pólos, L., & Carroll, G. (2007). *Logics of organization theory: Audiences, codes, and ecologies*. Princeton, N.J.: Princeton University Press.

Heaney, M. T., & Rojas, F. (2014). Hybrid activism: Social movement mobilization in a multimovement environment. *American Journal of Sociology 119*(4), 1047–103.

Hsu, G. (2006). Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production. *Administrative Science Quarterly, 51*(3), 420-450.

Hsu, G., & Hannan, M. T. (2005). Identities, genres, and organizational forms. *Organization Science, 16*(5), 474-490.

Hsu, G., Hannan, M. T., & Koçak, Ö. (2009). Multiple category memberships in markets: An integrative theory and two empirical tests. *American Sociological Review, 74*(1), 150-169.

Hsu, G., Hannan, M. T., & Pólos, L. (2011). Typecasting, legitimation, and form emergence: A formal theory. *Sociological Theory, 29*(2), 97-123.

Hsu, G., Roberts, P. W., & Swaminathan, A. (2012). Evaluative schemas and the mediating role of critics. *Organization Science, 23*(1), 83-97.

Kovács, B., Carroll, G. R., & Lehman, D. W., (2013). Authenticity and consumer value ratings: Empirical tests from restaurant domain. *Organization Science, 25*(2), 458–478.

Kovács, B., & Hannan, M. T. (2015). Conceptual spaces and the consequences of category spanning. *Sociological Science, 2*, 252-286.

Kuhlen, N., & Preston, A., (2020). *News Entropy*. Available at SSRN: https://ssrn.com/abstract=3820449.

Lehman, D. W., O’Connor, K., Kovács, B., & Newman, G. E. (2019). Authenticity. *Academy of Management Annals, 13*(1), 1-42.

Leung, M. D., & Sharkey, A. J. (2014). Out of sight, out of mind? Evidence of perceptual factors in the multiple-category discount. *Organization Science, 25*(1), 171-184.

Li, X., & Zhang, B. (2020). A preliminary network analysis on steam game tags: Another way of understanding game genres. In *Proceedings of the 23rd International Conference on Academic Mindtrek* (pp. 65-73).

Navis, C., & Glynn, M. A. (2010). How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. *Administrative Science Quarterly, 55*(3), 439-471.

Negro, G., Hannan, M. T., & Rao, H. (2011). Category reinterpretation and defection: Modernism and tradition in Italian winemaking. *Organization Science, 22*(6), 1449-1463.

Negro, G., & Leung, M. D. (2013). “Actual” and perceptual effects of category spanning. *Organization Science, 24*(3), 684-696.

Nesterenko, O. (2020). IDC: video games industry revenue grows 20% in 2020 to nearly $180 billion. Game World Observer. Retrieved from: https://gameworldobserver.com/2020/12/28/idc-gaming-revenue-2020.

Paolella, L., & Durand, R. (2016). Category spanning, evaluation, and performance: Revised theory and test on the corporate law market. *Academy of Management Journal, 59*(1), 330-351.

Phillips, D. J., & Kim, T.-Y. (2009). Why pseudonyms? Deception as identity preservation among jazz record companies, 1920–1929. *Organization Science 20*(3), 481–99.

Phillips, D. J., & Zuckerman, E. W. (2001). Middle-status conformity: Theoretical restatement and empirical demonstration in two markets. *American Journal of Sociology, 107*(2), 379-429.

Pontikes, E. G. (2012). Two sides of the same coin: How ambiguous classification affects multiple audiences’ evaluations. *Administrative Science Quarterly, 57*(1), 81-118.

Pontikes, E. G., & Hannan, M. T. (2014). An ecology of social categories. *Sociological Science, 1*, 311-343.

Pontikes E. G., & Kim, R. (2017). Strategic Categorization. In R. Durand, N. Granqvist and A. Tyllström (Eds.), *From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads* (Vol. 51, pp. 71-111): Emerald Publishing Limited.

Porac, J. F., Thomas, H., Wilson, F., Paton, D., & Kanfer, A. (1995). Rivalry and the industry model of Scottish knitwear producers. *Administrative Science Quarterly, 40*(2), 203-227.

Rao, H., Monin, P., & Durand, R. (2003). Institutional change in Toque Ville: Nouvelle cuisine as an identity movement in French gastronomy. *American Journal of Sociology, 108*(4), 795-843.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379-423.

Taeuscher, K., & Rothe, H. (2021). Optimal distinctiveness in platform markets: Leveraging complementors as legitimacy buffers. *Strategic Management Journal, 42(2), 435-461.*

Tan, D., & Roberts, P. W. (2010). Categorical coherence, classification volatility and examiner-added citations. *Research Policy, 39*(1), 89-102.

Tastle, W. J., & Wierman, M. J. (2007). Consensus and dissention: A measure of ordinal dispersion. *International Journal of Approximate Reasoning, 45*(3), 531–545.

Tost, L. P. (2011). An integrative model of legitimacy judgments. *Academy of Management Review, 36*(4), 686-710.

Vergne, J. P. (2012). Stigmatized categories and public disapproval of organizations: A mixed-methods study of the global arms industry, 1996–2007. *Academy of Management Journal, 55*(5), 1027-1052.

Verhaal, J. C., Khessina, O. M., & Dobrev, S. D. (2015). Oppositional product names, organizational identities, and product appeal. *Organization Science, 26*(5), 1466-1484.

Waguespack, D. M., & Sorenson, O. (2011). The ratings game: Asymmetry in classification. *Organization Science 22*(3), 541–53.

Wang, D. J., Rao, H., & Soule, S. A. (2019). Crossing Categorical Boundaries: A study of diversification by social movement organizations. *American Sociological Review, 84*(3), 420-458.

Windleharth, T. W., Jett, J., Schmalz, M., & Lee, J. H. (2016). Full steam ahead: A conceptual analysis of user-supplied tags on Steam. *Cataloging & Classification Quarterly, 54*(7), 418–441.

Zhang, J., & Liu, R. (2017). Popularity of digital products in online social tagging systems. *Journal of Brand Management, 24*(1), 105-127.

Zhao, E. Y., Fisher, G., Lounsbury, M., & Miller, D. (2017). Optimal distinctiveness: Broadening the interface between institutional theory and strategic management. *Strategic Management Journal, 38*(1), 93-113.

Zhao, E. Y., Ishihara, M., Jennings, P. D., & Lounsbury, M. (2018). Optimal distinctiveness in the console video game industry: An exemplar-based model of proto-category evolution. *Organization Science, 29*(4), 588-611.

Zhao, E. Y., Ishihara, M., & Lounsbury, M. (2013). Overcoming the illegitimacy discount: Cultural entrepreneurship in the US feature film industry. *Organization Studies*, *34*(12), 1747-1776.

Zuckerman, E. W. (1999). The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology, 104*(5), 1398-1438.

Table 1. Descriptive statistics

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mean | SD | Min | Max | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 1. Price | 911.92 | 1067.03 | 0 | 5999 |  |  |  |  |  |  |  |
| 2. Rating | 69.44 | 13.54 | 25.22 | 94.72 | 0.13 |  |  |  |  |  |  |
| 3. Franchise | 1.21 | 0.41 | 1 | 2 | 0.19 | 0.15 |  |  |  |  |  |
| 4. Owners | 106,807 | 362,876 | 1,000 | 6,678,000 | 0.14 | 0.13 | 0.07 |  |  |  |  |
| 5. No. of Genres | 2.72 | 1.16 | 1.00 | 8.00 | -0.26 | -0.05 | -0.06 | -0.03 |  |  |  |
| 6. Typicality | 0.04 | 0.02 | 0.01 | 0.44 | -0.12 | -0.08 | -0.07 | -0.11 | -0.18 |  |  |
| 7. Similarity | 0.11 | 0.11 | 0.03 | 1.00 | -0.10 | -0.19 | -0.07 | -0.12 | -0.24 | 0.57 |  |
| 8. Entropy | 2.60 | 0.84 | 0.00 | 3.91 | 0.19 | 0.25 | 0.16 | 0.22 | 0.25 | -0.52 | -0.78 |
|  |  |  |  |  |  |  |  |  |  |  |  |

Table 2. OLS regression models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Producer-audience similarity | | Audience-audience entropy | |
|  | (1) | (2) | (3) | (4) |
|  | | | | |
| Price | -0.00\*\*\* | -0.00\*\* | 0.00\*\*\* | 0.00\*\*\* |
|  | (0.00) | (0.00) | (0.00) | (0.00) |
|  |  |  |  |  |
| Rating | -0.001\*\*\* | -0.001\*\*\* | 0.01\*\*\* | 0.01\*\*\* |
|  | (0.00) | (0.00) | (0.002) | (0.001) |
|  |  |  |  |  |
| Franchise game | -0.01 | -0.00 | 0.21\*\*\* | 0.18\*\*\* |
|  | (0.01) | (0.01) | (0.06) | (0.05) |
|  |  |  |  |  |
| Number of owners | -0.00\*\*\* | -0.00\* | 0.00\*\*\* | 0.00\*\*\* |
|  | (0.00) | (0.00) | (0.00) | (0.00) |
|  |  |  |  |  |
| Number of genres | -0.03\*\*\* | -0.02\*\*\* | 0.23\*\*\* | 0.16\*\*\* |
|  | (0.00) | (0.00) | (0.02) | (0.02) |
|  |  |  |  |  |
| Typicality of producer tags |  | 2.34\*\*\* |  | -14.91\*\*\* |
|  |  | (0.11) |  | (0.82) |
|  |  |  |  |  |
| Constant | 0.30\*\*\* | 0.16\*\*\* | 0.85\*\*\* | 1.75\*\*\* |
|  | (0.02) | (0.02) | (0.13) | (0.13) |
|  |  |  |  |  |
|  | | | | |
| Observations | 1,144 | 1,144 | 1,144 | 1,144 |
| Adjusted R2 | 0.12 | 0.37 | 0.22 | 0.39 |
| F Statistic | 32.46\*\*\* | 110.80\*\*\* | 64.92\*\*\* | 124.09\*\*\* |
|  | | | | |
| *Note:* | \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 | | | |

Figure 1. Distributions of tag counts at different levels of entropy

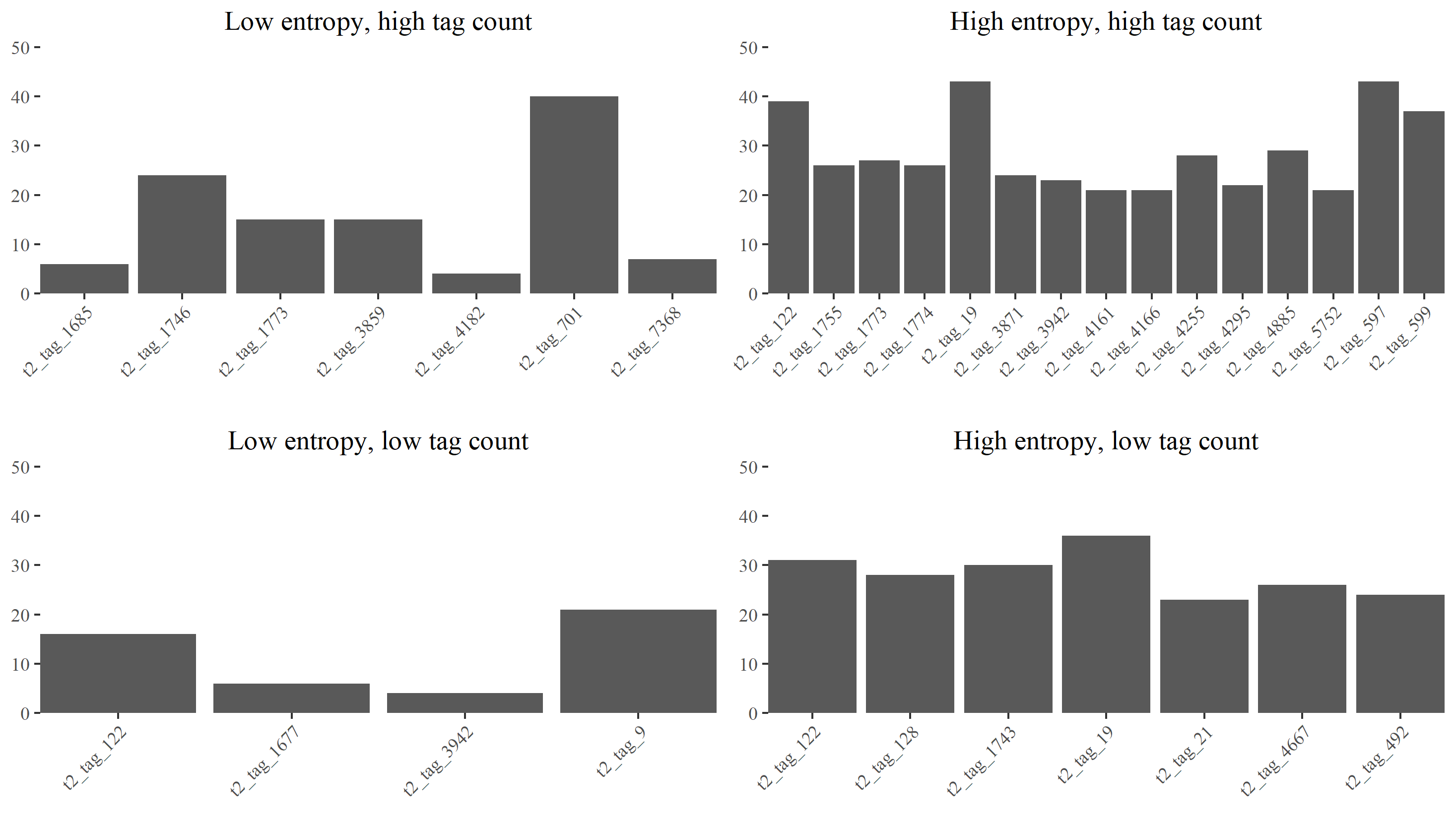


Figure 2. Typicality of producer tags versus similarity between producer and audience tags

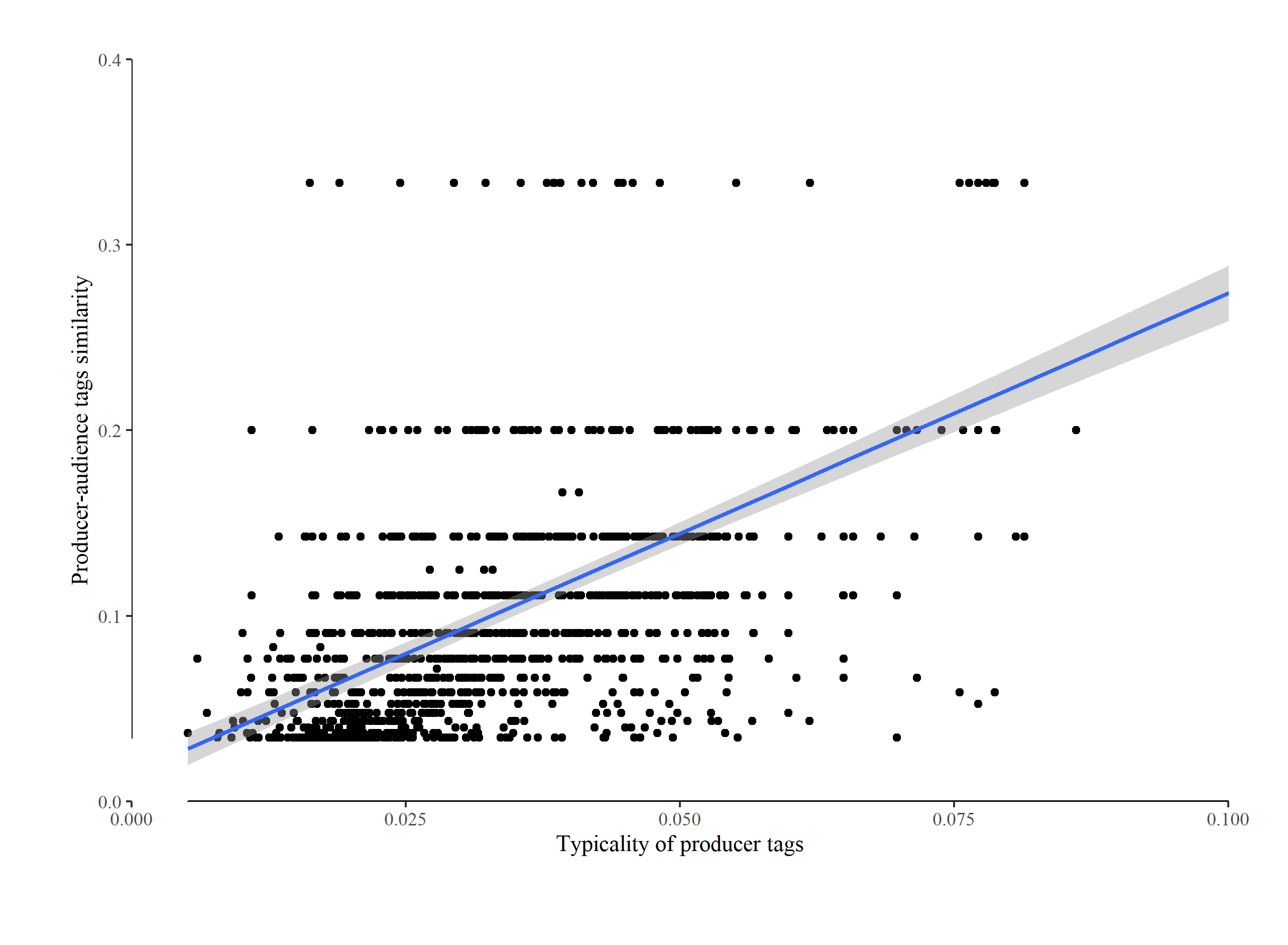


Figure 3. Typicality of producer tags versus entropy of audience tags

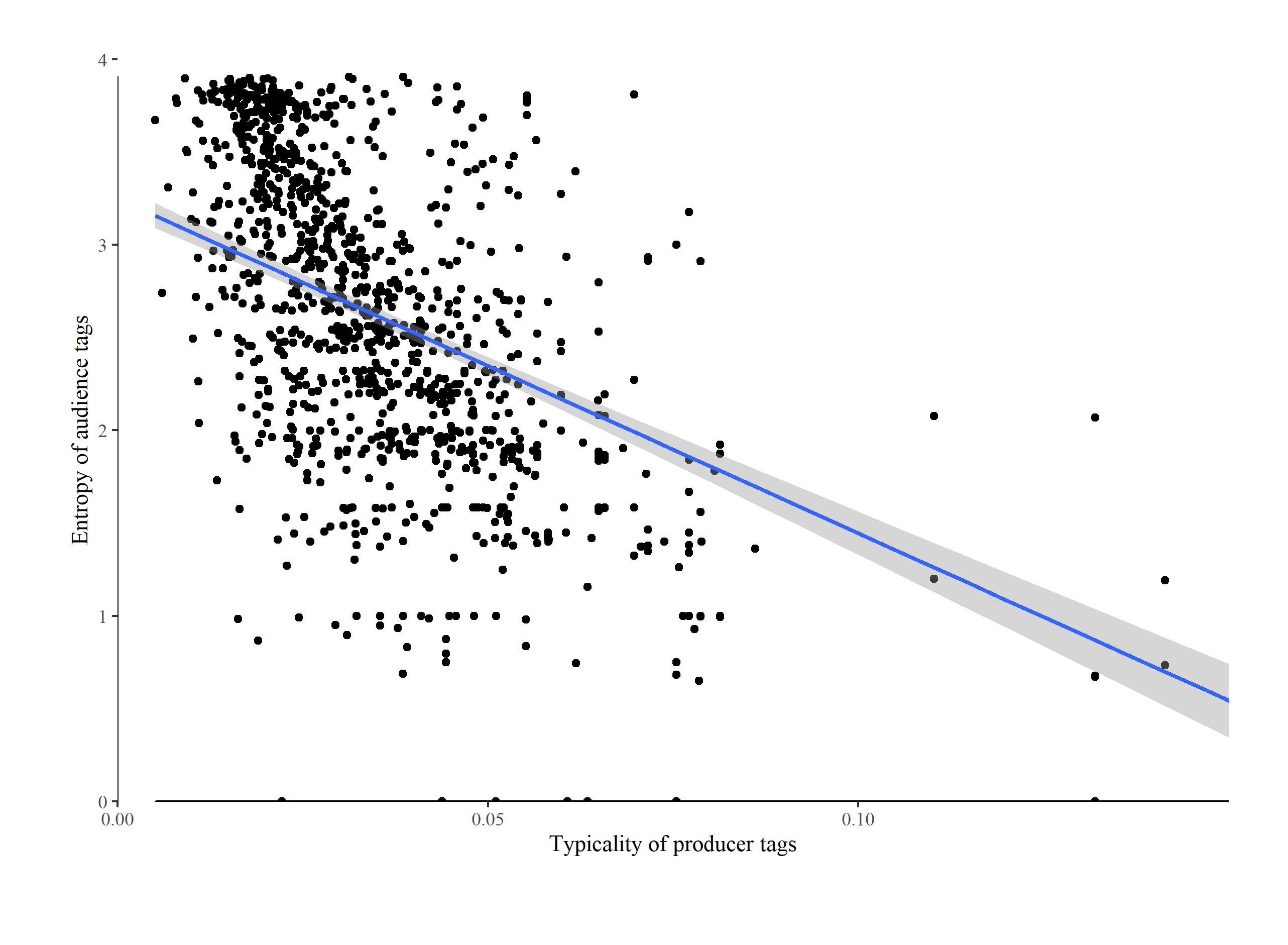
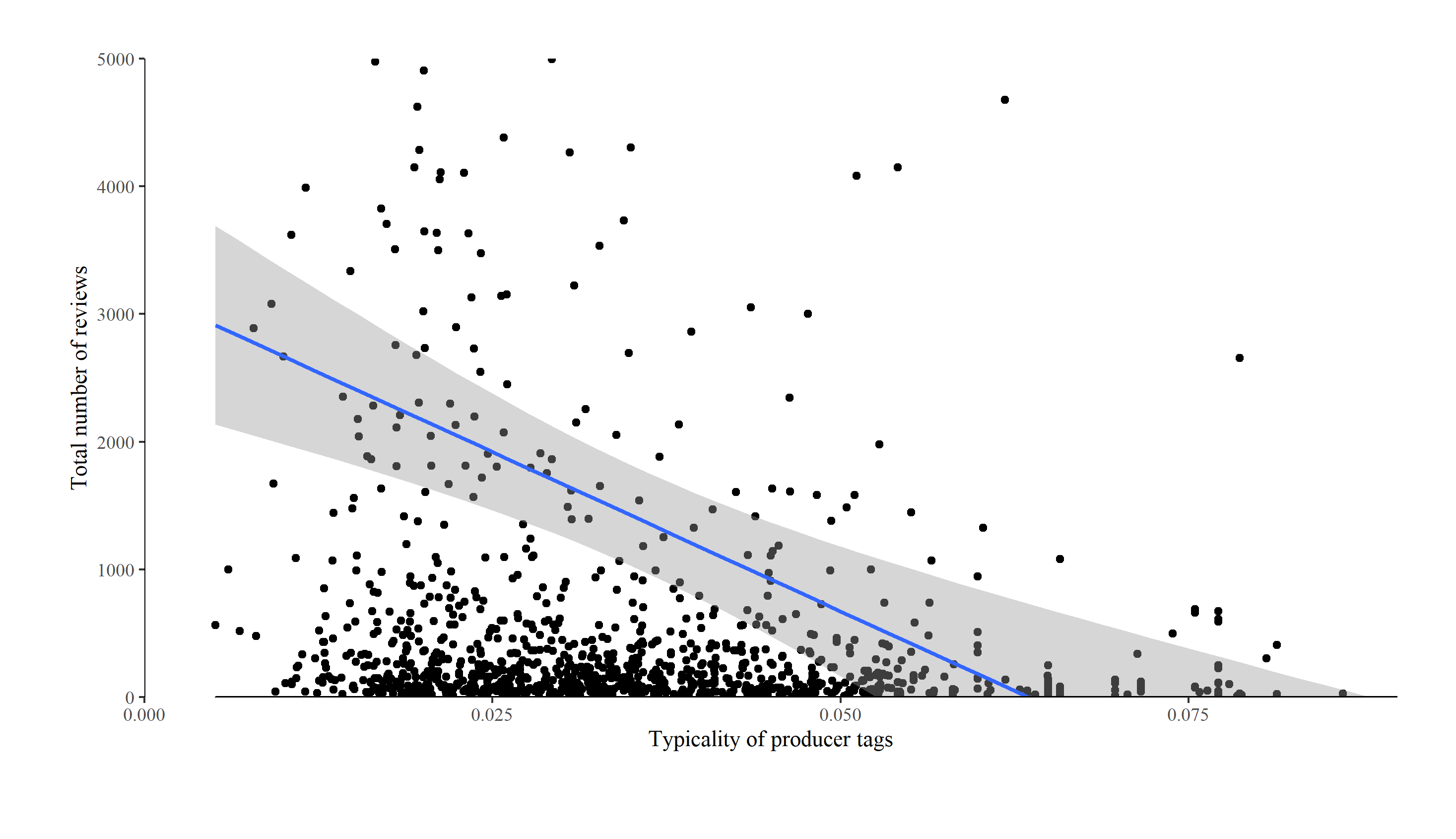


Figure 4. Entropy of audience tags versus number of reviews



1. For a detailed illustration of how this measure is calculated see Appendix I in Kovacs and Johnson (2014). [↑](#footnote-ref-1)
2. According to Steamspy, this estimate includes users who obtained the game in a number of ways, e.g., gifts from other users, giveaways, etc. Unfortunately, this means that this is not a useful measure of performance. [↑](#footnote-ref-2)