Received: 29 September 2017 | Revised: 22 January 2019 | Accepted: 27 January 2019 | Published on: 9 April 2019

DOI: 10.1002/smj.3019

STRATEGIC MANAGEMENT JOURNAL WILEY

RESEARCH ARTICLE

Optimal distinctiveness, strategic categorization, and product market entry on the Google Play app platform

Matthew A. Barlow[1] | J. Cameron Verhaal[2] | Ryan W. Angus[3]

[1]Department of Management, College of Business Administration, University of Nebraska-Lincoln, Lincoln, Nebraska

[2]Department of Management, Freeman School of Business, Tulane University, Goldring/Woldenberg Business Complex, New Orleans, Louisiana

[3]Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, West Virginia

Correspondence
Matthew A. Barlow, Department of Management, College of Business Administration, University of Nebraska-Lincoln, Lincoln, NE 68588.
Email: drmatthewabarlow@gmail.com

***Research Summary*:** New entrants often face uncertainty regarding how to optimally position themselves within product markets. We suggest that new entrants can use two important schemas to strategically categorize themselves to gain a competitive advantage in platform markets: category exemplars and category prototypes. Using a unique dataset of more than 83,000 new Google Play developers and more than 139,000 apps, we find that the optimally distinct entry point is at a high level of exemplar similarity and a low level of prototype similarity. We find that greater alignment of an entrant with the prototype corresponds to a weaker benefit of exemplar similarity. These findings have important implications for understanding competitive dynamics within product markets, strategic positioning at entry, and the interdependence of strategic categorization decisions.

***Managerial Summary*:** Entrepreneurial startups often find it difficult to know how to optimally position their products among a large number of rivals in highly competitive platform markets. Our study suggests that these startups can draw on two reference points to help determine the optimal positioning for their products: category exemplars and category prototypes. Exemplars include the most successful products in a market category while prototypes represent the most common products in a category. Drawing on a large dataset obtained from the Google Play app store, we find that developers can substantially increase the installs of their first app by crafting an app text description that is as similar as possible to the description of a category exemplar and as different as possible from the category's prototypical description.

KEYWORDS

market categories, new market entry, optimal distinctiveness, platform markets, strategic categorization

1 | INTRODUCTION

Scholars have long been interested in the inherent tension that organizations face in balancing the need to conform to consumer expectations with the desire to differentiate themselves from their competitors (Deephouse, 1999; Dimaggio & Powell, 1983; Haans, 2019; Haveman, 1993; Oliver, 1997; Zhao, Fisher, Lounsbury, & Miller, 2017; Zhao, Ishihara, Jennings, & Lounsbury, 2018; Zuckerman, 2016). *Optimal distinctiveness* is typically conceptualized as a balancing act, akin to a zero-sum proposition, where greater differentiation inherently indicates less conformity, and the point of optimal distinctiveness lies at the midpoint between two countervailing forces. However, in line with recent calls, we view the process of achieving optimal distinctiveness as multi-faceted and mutually enabling (Durand & Paolella, 2013; Zhao et al., 2017). As a result, an important challenge for improving our understanding of competitive dynamics related to optimal distinctiveness is the deceptively simple question: to whom or to what do organizations conform to and differentiate from to gain a competitive advantage?

This question is particularly relevant for de novo market entrants that may not have the requisite experience to fully understand the nuances of the conformity/distinctiveness paradox. Moreover, in established market categories, new entrants often struggle to influence or shape how they are perceived vis-à-vis their competitors within the category. Indeed, research on market categories has long suggested that external intermediaries drive perceptions of category fit and audience appeal (Vergne & Wry, 2014; Zuckerman, 1999). Increasingly, however, research on market categorization is identifying conditions under which organizations can assume a more active role in shaping their strategic fit within market categories. Specifically, new entrants can actively engage in a process of strategic categorization to systematically influence their category membership (Cattani, Porac, & Thomas, 2017). While much of this research focuses on strategic categorization across different market categories (Pontikes & Kim, 2017) or how organizations proactively shape the codes and norms that drive category emergence (Khaire & Wadhwani, 2010), we know considerably less about how organizations can strategically position themselves within a single category (for an exception, see Zhao et al., 2018) to achieve optimal distinctiveness. Therefore, we build on these studies and integrate the literatures on optimal distinctiveness (Deephouse, 1999; Zhao et al., 2017) and strategic categorization (Pontikes & Kim, 2017; Vergne & Wry, 2014) to investigate the degree to which de novo entrants should actively align themselves with or differentiate themselves from two interdependent categorical schemas—category prototypes and category exemplars. Following the extant literature, we identify the category prototype as the most representative member in a market category (Rosch & Lloyd, 1978; Rosch & Mervis, 1975) and the category exemplar as the most salient category member or a clear market leader (E. E. Smith & Medin, 1981). By identifying these different positioning points, we suggest that individual market categories are potentially complex and multidimensional competitive spaces in their own right (Cattani et al., 2017). Thus, positioning within a given market category may be as strategically important as positioning across different market categories.

To test our predictions, we focus on two-sided platform markets. Specifically, we draw on a dataset containing more than 22 million app-month observations to identify 139,582 apps published by 83,115 de novo app developers from the Google Play App store between February 2015 and August 2016. Using this empirical context, we measure the degree to which these apps are strategically aligned with prototypical and/or exemplar apps within Google Play's 41 given product market categories. These measures are constructed using natural language processing techniques applied to the text descriptions of all apps on the platform. Through these descriptions, developers can directly communicate to consumers the features or characteristics that they consider most important and position themselves within a given product market category, above and beyond claiming simple category membership. We argue and find support for the notion that in two-sided platform markets, strategic alignment with the category prototype hurts market performance, while conforming with a category exemplar leads to higher performance. We also find that alignment with the prototype dilutes the positive benefits associated with conforming oneself with a category exemplar. In practical terms, our empirical findings suggest that app developers can more than double their expected performance by crafting a text description that is highly similar to a category exemplar and low in similarity to the category prototype. Furthermore, by using this positioning strategy, developers are nine-times more likely to publish an app that generates 100,000 or more installs, a level of performance that applies to less than 4% of the apps in our dataset. These findings provide insight into how organizations can use strategic categorization to achieve optimal distinctiveness and have important strategic implications for the burgeoning literature on platform ecosystems.

2 | THEORY

2.1 | Strategic categorization

Strategic categorization can be understood as explicitly aligning or linking an organization (or aspects of organizational design or identity) to an existing categorical schema for the purpose of gaining a competitive advantage over competitors within that category. Pontikes and Kim identify two fundamental benefits of strategic categorization for organizations: “to communicate information and to position themselves favorably with respect to competitors” (2017: 73). Research has shown that many benefits can accrue for organizations that employ strategic categorization, including the ability to shape categorical boundaries during the nascent stages of category emergence (Kennedy, 2008; Khaire & Wadhwani, 2010; Navis & Glynn, 2010), to reduce the penalty associated with category spanning (Zhao, Ishihara, & Lounsbury, 2013), to reduce the penalties associated with membership in a stigmatized category (Vergne, 2012), and to alleviate the competitive pressures in proximate market categories (Pontikes & Kim, 2017).

Strategic categorization offers a key advantage as it allows organizations to shape the narrative of their positioning in a competitive landscape. Typically, these narratives are left to third-party market intermediaries that serve as a social filter and can help to crystallize the cognitive ordering of a product category and confer legitimacy in the process. These intermediaries often take the form of industry experts or analysts (Zuckerman, 1999), the media (Kennedy, 2008), certification organizations (Lee, Hiatt, & Lounsbury, 2017), or professional critics (Rao, Monin, & Durand, 2003). While third-party participation often aids organizations by clarifying categorical distinctions and helping “to penetrate opaque buyer-supplier interfaces” (Cattani et al., 2017: 78), these third parties can also use categorical membership as a sanctioning mechanism for organizations that deviate from accepted categorical codes or schemas (Zuckerman, 1999).

Much of the prior work on strategic categorization has examined how organizations strategically select the categories to which they belong (Cattani et al., 2017; Pontikes, 2018; Pontikes & Kim, 2017). However, the choice of category membership is not the only choice that organizations can make—organizations can also choose how to position themselves within a given market category. Therefore, we extend the notion of strategic categorization by examining within-category strategic positioning. Given that organizations can use strategic categorization to position themselves within a product category, what are the actual tools that organizations have at their disposal to shape and influence market demand for their goods or services? We identify one potential practice that organizations can leverage—drawing on an existing categorical nomenclature (Cattani et al., 2017). Categorical nomenclatures serve as a semantic tool for "labeling, codifying, and diffusing category-relevant market conversations" (Cattani et al., 2017: 78). Moreover, categorical nomenclatures can aid in transferring the two key benefits of strategic categorization: communicating information and positioning an organization favorably vis-à-vis rivals. Categorical nomenclatures can manifest through direct communication with consumers, advertising, or product descriptions—in effect, anything that the organization uses to communicate information or strategically position itself within the category.

In our empirical context, the Google Play app store, categorical nomenclatures can manifest through the text description that developers are required to provide for each app. These descriptions represent a key mechanism for organizations to proactively position themselves within a market category. As long as apps do not infringe on copyrights or contain malware or offensive material, Google is unlikely to restrict the strategic categorization activities of app developers.[1] Thus, Google Play app developers are largely free to exercise agency and engage in strategic categorization by developing and publishing almost any kind of app, placing these apps into product market categories of the developers' choosing, and describing these apps however the developers see fit.

2.2 | Optimal distinctiveness and strategic positioning

The process of achieving optimal distinctiveness within a market space can be thought of as a two-stage process: to succeed, an organization must first make it into a consumer's consideration set and then distinguish itself from the others in that set to ultimately gain the favor of consumers (Zuckerman, 1999). In other words, an organization should position itself to be "as different as legitimately possible" (Deephouse, 1999: 147).

Given that organizations use strategic categorization to position themselves within a category, an important question remains: what should an organization use as a benchmark or anchor to position against in order to achieve optimal distinctiveness? While early work focused on positioning against the average organization in a market category (Deephouse, 1999), more recent research has argued that no single convergence point exists at which organizations can attempt to achieve strategic balance within a category (Zhao et al., 2017). Thus, while singular organizational attributes or specific environmental conditions are important in understanding organizational performance and heterogeneity, understanding the multidimensionality of a given market category and how organizations are positioned relative to these multiple competitive reference points is critical.

[1]Provided that a proposed app does not contain malware or offensive material and does not infringe on copyrights, it is likely to be approved in a matter of hours (https://android-developers.googleblog.com/2015/03/creating-better-user-experiences-on.html).

Two perspectives have emerged to help explain how organizations might position themselves and how audience members make sense of and interpret different categorical paradigms: prototype- and exemplar-based models. The prototype model builds on early work in cognitive psychology (Rosch & Lloyd, 1978; Rosch & Mervis, 1975) that sought to interpret categorical distinctions through a process of grouping attributes or features of a given category. The features that are considered the most central or representative in the mind of the audience emerge as the prototype for a given category (Hannan, Pólos, & Carroll, 2007; Vergne & Wry, 2014). Prototypes reduce confusion among consumers about an organization's place in the market and whether it belongs there (Hsu, 2006a, 2006b). In contrast, exemplars can be understood as offerings that stand out as particularly salient or exceptional representations of a category (Cohen & Basu, 1987; E. E. Smith & Medin, 1981; Zhao et al., 2018). Often, they can be understood as the most well-known or highest-performing members of a group. Indeed, "some exemplars (e.g., robins as a member of the category "bird") are treated as "better" members of a category" (Durand & Paolella, 2013: 1103), creating a "halo effect" that may confer added legitimacy or direct attention to those that imitate the exemplar. Durand and Kremp (2016) develop this further by labeling conformity to the prototype as "alignment" and conformity to the exemplar as "conventionality." Thus, in their study of symphony orchestras, "when an orchestra's programming choices are close to those of its peers', it exhibits alignment; when it chooses to perform canonical composers more often than its peers, it exhibits conventionality" (2016: 66).

3 | HYPOTHESES

3.1 | Strategic categorization and similarity to the prototype

Prior research suggests that alignment with a category prototype can provide increased legitimacy by reducing confusion in the eyes of audience members and signal that the organization is a valid member of the category. While extensive research has convincingly argued and found empirical support for the benefits of prototypicality (Hannan et al., 2007; Negro, Hannan, & Rao, 2010; Zuckerman, Kim, Ukanwa, & von Rittmann, 2003), certain conditions may exist that render alignment with the prototype a detrimental entry and positioning strategy. For example, recent research suggests that alignment to a category prototype is associated with the desire to look like one's peers and represents a penalty avoidance tactic (Durand & Kremp, 2016). However, not all organizations or individual have the same risk preferences, and prototypicality may create a mismatch between strategic goals and outcomes. Additionally, in the nascent stages of category emergence, a clear prototype may not have had time to emerge (Hiatt & Carlos, 2018; Lee et al., 2017; Zhao et al., 2018). Even if a prototype has emerged, it may change significantly as the product category matures. Furthermore, product categories can be quite broad, with many thousands of competitors offering a wide range of product types within a single category. As categories become more crowded and diffuse, the products in these categories run the risk of being lost in a crowd of disparate offerings. Ultimately, this crowding effect undermines the effectiveness of positioning near the category prototype.

The deleterious effect of positioning near the category prototype is exacerbated in platform markets due to the gatekeeper role played by the platform itself. Many "modern markets for information" are dominated by gatekeepers (Baye & Morgan, 2001: 454). In the case of platform markets, the platform acts as a gatekeeper by controlling the interface between the producer and consumer at the level of the transaction and serving as an important third-party evaluator (Paolella & Durand, 2016) of quality and category membership. Thus, in platform markets such as Google Play, the platform itself bestows a degree of "taken-for-grantedness" (Suchman, 1995) on the products it allows to be sold, reducing the need for organizations to engage in strategic categorization to signal their legitimate membership in a category by strategically categorizing themselves in a manner similar to a category prototype. Therefore, in platform markets, the strategic benefits typically associated with alignment with the prototype (i.e., taken-for-grantedness) are minimized, while strategic challenges (e.g., lack of differentiation in highly crowded markets) are amplified, leading to negative performance outcomes.

To illustrate the potential drawbacks of prototype conformity in these markets, consider a de novo developer entering Google Play's Communication category with an app for delivering personalized communications, news, and notifications to employees of large corporations. During the time of this study, Google Play's "Communication" category had more than 35,000 unique apps and included an incredibly diverse set of web browser, social media, direct messaging, email, video chat, and caller ID apps. The prototypical nomenclature for this category includes words such as *message*, *notify, connect, contact*, *chat*, *inform*, and *share*, all of which are words that this developer might use to describe a personalized corporate communication app. However, these words also apply to a whole host of other Communication apps. While using some of these terms in the personalized corporate communication app's text description is likely inevitable, the more heavily they are used, the more Google Play will group them in the search results with thousands of other apps that also offer these basic features. The result is that consumers are unable to find apps with potentially superior product features because its prototypical categorization undermines its visibility in the market.

Particularly for unique apps (such as the personalized corporate communications app described above), the need to stand out should be acute. One of the key benefits of platform markets is that they facilitate outreach to and communication with an extremely scattered and diffuse target market. In more traditional product markets, entrants may struggle to generate the marketing budget or economies of scale necessary to reach these consumers. However, through the direct interaction inherent in two-sided platform markets, these unique products can potentially reach consumers and thrive. The key becomes strategically categorizing the app in a manner that reaches this broad and diffuse consumer set without getting lost in the crowd in the process. Therefore, in platform markets where strategic categorization is possible, not only will alignment with the category prototype not produce the performance benefits one might expect in more traditional markets, the strategic challenges of operating in a highly competitive platform markets are exacerbated, leading to overall lower performance. This leads to the following hypothesis:

Hypothesis (H1). *The more an app conforms to its category prototype at the time of entry, the lower the performance of that app.*

3.2 | Strategic categorization and similarity to the exemplar

We contend that in platform markets organizations product offerings that differentiate themselves from the category prototype should have a competitive advantage over organizations that align themselves with the prototype. However, whether other categorical schema exist around which conformity can provide a competitive advantage remains unanswered. Indeed, offerings that are excessively distinctive run the risk of incommensurability, or the absence of a comparative baseline with which to make judgments of quality (Durand & Kremp, 2016). We suggest that conformity to a category exemplar (i.e. conventionality) provides a credible strategic advantage when making product market entry and positioning choices. The logic underpinning this argument is that conformity to an exemplar is conducive to a goal-based approach to category positioning that is driven by positive rewards (Pontikes & Kim, 2017). Exemplars stand out, which is important in competitive environments. Moreover, exemplar conformity offers two key strategic benefits. First, by conforming to the salient attributes of a select few actors within a market category, organizations are able to help audiences compare and assess the quality of their offerings (Durand & Kremp, 2016). At the same time, conformity to the exemplar allows organizations to simultaneously differentiate themselves across other dimensions and avoid the hotly contested center of the market.

For example, recent research has identified positive benefits for actors that have product features conforming with the market exemplar, specifically in nascent markets where the prototype may not have had time to fully develop or emerge (Zhao et al., 2018). We extend these findings in an important manner by arguing that, in platform markets, an exemplar-based strategy can still provide a competitive advantage even as the market category evolves and a clear prototype emerges. Importantly, this argument is contingent on the organization focusing its positioning choices on goal- and reward-based strategies (Durand & Paolella, 2013) over strategies built on penalty avoidance. Strategic categorization affords these new entrants the agency to move beyond using positioning simply as a means of signaling category membership, to using positioning to gain a strategic foothold in highly contested markets, accomplished by organizations using current categorical nomenclature (Cattani et al., 2017) to cognitively conform with the category exemplar. The strategic use of categorical nomenclatures that conform more heavily with category exemplars allows firms to stand out from the majority of the market category and gain the attention of the audience, which should lead to better performance.

To examine the exemplar positioning strategy, consider a de novo app developer entering Google Play's Communication category with a new web browser app. This category has exemplar web browser apps with hundreds of millions of downloads, such as Firefox. The nomenclature of these exemplars includes words that are *not* part of the category prototype's nomenclature, such as *bookmark, browse, engine, fast, incognito, intuitive, page, privacy,* and *web.* A de novo app developer that crafts the text description of its web browser app such that it aligns closely with the exemplar nomenclature will be well positioned to capture the attention of consumers and potentially generate a high level of installs for several reasons because the similarity of the new app's text description to the exemplar nomenclature increases the likelihood that the new app will appear in search results when consumers search for web browsers using Google Play's search bar. Moreover, even if consumers first click on one of the exemplar apps, the new app's similarity to the exemplar increases the likelihood that it will appear on Google's list of similar apps displayed within the exemplar app's product details screen. In either case, the new app's similarity to one or more exemplars increases the likelihood that it will be seen by and subsequently downloaded by consumers. Therefore, in platform markets, conforming with a category exemplar (i.e. conventionality) enables entrants to stand out from the crowd and achieve higher performance, thus leading to the following hypothesis:

Hypothesis (H2). *The more an app conforms to a market category exemplar at the time of entry, the greater the performance of that app.*

3.3 | Strategic categorization and optimal distinctiveness in a multidimensional market space

Finally, we suggest a more interdependent perspective where market entry positioning choices should account for alignment with the category prototype and the category exemplar concomitantly. Specifically, any alignment with the prototype can dilute the strategic benefits that products potentially gain from conformity with the exemplar. We argue that this dynamic is likely to be found in markets characterized by low barriers to entry and a lack of shakeout at the later stages of the product life cycle (Klepper, 1997), because it creates a hyper-crowded market—a situation often found in many platform markets (Rietveld & Eggers, 2018). This is exacerbated by the low cost of keeping many products, including failed ones, on the platform for extended periods of time. In these markets, alignment with the prototype dilutes (Vergne, 2012) the process of strategic categorization (Pontikes & Kim, 2017), undermining the effectiveness of directly communicating association with a category exemplar to consumers. Ultimately, the more attributes a product has which aligns with the prototype, the fewer the attributes available to them to stand out (Durand & Kremp, 2016).

The logic underpinning this is derived from economic theories of product positioning which suggest that new entrants will often seek to strategically wedge themselves between other competitors and a mass of customers (Hotelling, 1929). In market categories where numerous competitors have attempted to position themselves between a category exemplar and a mass of customers near the category prototype (at the center of the market), new entrants may end up being high in similarity to both a category exemplar and the category prototype. This can create a largely undifferentiated cluster of competitors around the exemplar in the direction of the prototypical category member. More recent research in density dependence and industry dynamics has found that this clustering at the center of the market can lead to both fiercer competition within markets (Carroll & Hannan, 1989) and exit to adjacent markets (Dobrev & Kim, 2006). As a result, new entrants will not be likely to gain the same benefits they would from exemplar similarity as they would if they were also far from the category prototype. Thus, alignment with the category prototype can substantially, if not fully, dilute the positive benefits of simultaneous alignment with a category exemplar. The optimal positioning strategy, then, is to be simultaneously similar to the category exemplar and far from the category prototype.

This interdependent, interactive effect of exemplar and prototype similarity is depicted in Figure 1. According to the theory developed in this paper, entrant 1, which positions itself close to

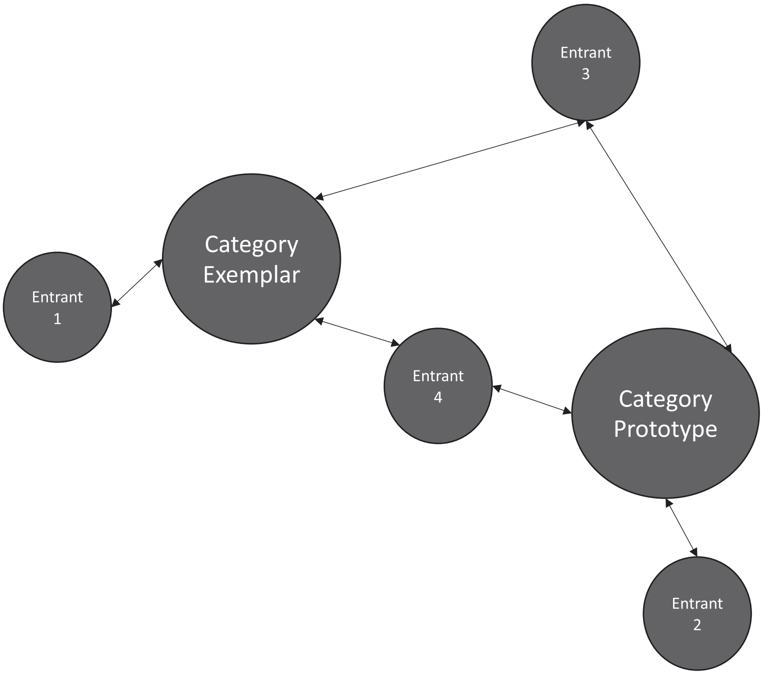

FIGURE 1 Hypothetical market space and entry positioning strategies. *Note:* The arrows represent the distance between the positioning of different offerings within a market category. The longer the arrow, the more dissimilar the positioning of two offerings

the exemplar and far from the prototype, is expected to outperform, on average, the other three entrants. This positioning allows entrant 1 to gain the attention benefits of close association with an exemplar without running the risk of getting lost in a crowd of prototypical competitors. Due to its proximity to the category prototype and distance from an exemplar, entrant 2 may find it difficult to attract audience attention amidst a host of similar, prototypical competitors, resulting in the lowest expected performance of the entrants. Entrant 3 is positioned far from both the category prototype and a category exemplar. While this entrant avoids the costs associated with similarity to the crowd of prototypical offerings, it does not gain the benefits associated with similarity to an exemplar, resulting in lower expected performance than entrant 1. Finally, although entrant 4 is similar to a category exemplar, it is also similar to the category's prototypical offering. Alignment with the prototype in this case dilutes the benefits the entrant would normally derive from close association with an exemplar, again resulting in lower expected performance than entrant 1.

For example, entrant 4 could be a de novo developer entering Google Play's Communication category with a video calling app, which would likely be similar not only to an exemplar with billions of downloads, such as Facebook Messenger, but also to the category's prototype. The text description of the new video call app would likely include nomenclature such as *call, video, chat, contact, message, send,* and *connect.* These words are all part of the nomenclature of an exemplar video call app, but they are also part of the category's prototypical nomenclature used by many thousands of apps. Thus, even if the new app were to appear on Google Play's search ranking list (or the list of apps related to an exemplar), it stands the risk of being lost in a cluster of highly similar apps and ending up so far down the list that it is unlikely to be discovered by consumers. Therefore, the key for this developer would be to either invoke a categorical nomenclature that strategically aligns the app with the exemplar while avoiding (to the extent possible) simultaneously invoking the prototypical nomenclature. Ultimately, the need to stand out in crowded platform market categories is so important that the penalty for alignment with the prototype will dilute the strategic benefits of alignment with an exemplar, thus leading to our final hypothesis:

Hypothesis (H3). *The greater an app's conformity to its category prototype at the time of entry, the smaller the performance benefit of that app's conformity to a category exemplar.*

4 | DATA & METHODS

4.1 | Empirical context

In this paper, we investigate entry strategies and competitive dynamics in two-sided platform markets which are becoming increasingly prevalent in the current economy. Examples of platforms include internet search engines, Amazon, Netflix, Uber, Airbnb, video game consoles, YouTube, eBay, iTunes, and the Google Play app store. Consumers use platforms to search for and acquire products and services developed by producers. Producers use platforms to gain immediate access to huge numbers of consumers who may be interested in acquiring these products and services. Platforms benefit from indirect network effects, becoming more valuable as the number of consumers and producers increases (Cennamo & Santalo, 2013; McIntyre & Srinivasan, 2017; Thomas, Autio, & Gann, 2015; Zhu & Iansiti, 2012). To foster the development of indirect network effects, many platforms adopt open architectures that minimize entry barriers and encourage large numbers of producers to offer their products and services as a means of attracting consumers to the platform (Thomas et al., 2015).

As a result, competition can be particularly intense on successful platforms, with potentially millions of producers competing for the attention of millions (or even billions) of consumers. Indeed, a key benefit for de novo producers entering a successful platform market is that they are immediately exposed to a large number of consumers without having to develop the requisite scope and scale economies that traditional markets require for this same level of exposure. Simultaneously, however, these same de novo entrants face a real possibility of immediately becoming lost in a cluster of similar products from other producers that also operate on the platform.

The platform selected as the empirical context in this study is the Google Play mobile application store (https://play.google.com/store/apps) for phones and tablets running the Android operating system. Google Play generated an estimated $17 billion in revenue in 2016, and this number is projected to reach $42 billion by 2021.[2] In 2016, this platform had more than 1 billion active monthly consumers,[3] 700,000 producers (mobile application development organizations),[4] and 2 million apps.[5] Many Google Play consumers discover apps through an organic search process.[6] For example, a consumer looking for a task management app might enter the text "task manager" into Google Play's search bar. Google Play then executes a search algorithm—which draws on the developer-provided text description for each app[7]—and presents the consumer with a list of apps matching the search criterion. Obtaining a high-ranking position on this list of search results can help an app receive more consumer attention and generate more installs. Google Play publicly provides rich data for each app on its platform, including the number of downloads, the number of reviews received, a category classification, a history of version changes, and—importantly for this study—a complete text description of each app written by the developer. These descriptions represent a key tool for developers to communicate directly with potential consumers and, in our study, serve as a means to identify and measure variation in strategic categorization across developers. Google advises developers that using a strong description to help their app to be found in the market is imperative.[8]

To create the sample, we collected data on more than 1 million apps on a monthly basis between February 2015 and August 2016. In this setting, thousands of new apps are published every month as developers attempt to generate economic value in a single marketplace. This rate of publication allows relatively easy identification of a sample of nascent app development organizations entering the market for the first time and facilitates tracking the performance of the apps published by these developers over time. To test this study's hypotheses, we restrict the sample to include only developers that published their first app between March 2015 and July 2016. We exclude any developers that publish more than 10 apps during our collection period since such developers are more likely to be larger companies or even contract development organizations, which are not the focus of our theorizing and may have very different market strategies compared to new developers. Furthermore, a larger development organization corresponds to a greater likelihood that they have additional resources and capabilities (e.g., the ability to heavily market on social media) that may affect app performance.

[2] App Annie Market Forecast 2016–2021

[3] https://mashable.com/2015/09/29/google-play-1-billion-users/#sS85m8FZsPqq

[4] A total of 707,831 unique app development organizations were present in our dataset.

[5] https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/

[6] https://www.tune.com/blog/app-store-optimization-win-google-play-app-store-search/

[7] https://support.google.com/googleplay/android-developer/answer/4448378?hl=en

[8] http://www.adweek.com/digital/google-discloses-how-search-for-google-play-works-for-the-first-time-12-percent-of-dau-search-for-apps-daily/

We also exclude developers that did not publish apps with an English text description and developers with data collection gaps. Finally, because we include a lagged review score measure to control for app quality, any apps with either no review score in the prior month or that were on the market for only 1 month are dropped from the final analysis. Our final sample consists of 589,681 app-month observations, with 139,582 unique apps and 83,115 unique developers. Our hypotheses argue that app developers achieve better performance at high levels of similarity to a category exemplar and low levels of similarity to the category prototype. Initial analyses of the data reveal that most developers do not appear to follow this strategy. A figure in the supplemental analysis shows a scatterplot of the similarity scores at entry for each app in the dataset. It shows that most apps are clustered in the area where they are low in similarity to both exemplars and prototypes, suggesting that if our hypotheses are correct, then most app developers are not optimally positioning their apps for a competitive advantage.

4.2 | Measures

4.2.1 | Dependent variable

We use *Review Count (ln)*, the logged count of reviews for each app-month observation, as our primary dependent variable. The logged count of downloads would also be a logical measure of app performance. However, the download measure obtained from Google Play is ordered and categorical (0, 1–5, 5–10, 10–50, etc.). We favor the use of the continuous measure of the logged count of reviews, although our results hold when using either measure. This consistency in results is not surprising since the review count and install measures are highly correlated (0.79), and only users who have downloaded an app can give it a review score in Google Play. We selected a logged measure because the distribution of review counts in the Google Play app store is highly skewed, with a considerably lower median number of reviews (8) than the average (487.5). Furthermore, more than 90% of the apps in our sample have less than 200 reviews (while the maximum number of reviews is 5,662,447). We add 0.01 to all review counts before logging to account for apps that have not yet received a review.

4.2.2 | Independent variables

We draw on the text descriptions of the apps in our sample to produce this study's two key explanatory variables. When publishing an app on the Google Play platform, developers must write a text description of their product, offering an opportunity for them to highlight what they view as the key characteristics or features of their app. In our sample, the mean cleansed description length is 486.48 characters. We use natural language processing methodologies to identify the primary language for each app's text description (ensuring that only English apps are included in the sample), remove stop words (e.g., "the," "your," "for"), and stem words to their root forms (e.g., "fish" would be the stem for "fishing," "fisher," and "fished") before producing our explanatory measures. Importantly, our two explanatory variables measure the similarity of each focal app to its category exemplars and category prototype the month prior to publishing the new app. In other words, although our regression sample includes only de novo app developers publishing their first apps, we create these explanatory measures by comparing the text description of each new focal app to *all* extant apps that were available for download in its category by drawing on our larger dataset with more than 22 million app-month observations.

The first explanatory variable, *Prototype Similarity*, measures how similar the focal app in our sample is compared to the representative or the prototypical app in the focal app's category. Accordingly, we identify the 50 most commonly used words contained in the descriptions of *all* the apps (not only the apps within our subsample of de novo developers) in each of Google Play's 41 app categories on a monthly basis. This measure is based on the top 50 words used in the focal app's category the month before it was first published. By looking at the prior month, we can capture the competitive environment at entry. We calculate the prototype similarity measure by dividing the count of the words in a focal app's text description that are also words from the category's top 50 words by the total number of words in the focal app's text description. Thus, a score of 0.0 would indicate that the focal app does not use any of its category's top 50 words and differs significantly from that category's prototypical app. A score of 1.0 would indicate that an app uses only words from its category's top 50 words list in its text description and is highly similar to the prototypical app. In our sample, the mean prototype similarity score is 0.22. Importantly, the prototype of each category is not necessarily an actual app in the category; rather, it is a measure of fit with the characteristics or features that most commonly define that category space. A new entrant can align itself with this prototype by communicating or highlighting these features in its own description.

Second, *Exemplar Similarity* measures how similar the description of the focal app is to its nearest neighbor among its category's list of the top 100 most installed apps (the exemplar apps) in the month prior to the app being introduced. Again, this measure is created by comparing the focal app's text description to *all* extant apps that were available in the focal app's Google Play category the month before the focal app was first published. This comparison ensures that we are capturing the competitive environment as the de novo app developer would have viewed it when positioning its new app on the market. To calculate the measure, we follow prior literature (Hoberg & Phillips, 2010) and calculate the cosine of the angle between two vectors to determine how similar the two vectors are. In this case, we first create a unique vector for each app in our sample. These vectors represent a list of all the words used in a particular app's cleansed text description (which includes only the stemmed words, with stop words removed for apps with English text descriptions). We also create a unique vector for each of the top 100 apps in each of Google Play's 41 categories. We then calculate the basic cosine similarity between the focal app's vector and the vector for each of the top 100 exemplar apps. Finally, we identify the focal app's nearest top 100 neighbor (the single highest cosine similarity score) from this list of 100 scores and use this neighbor as our measure of exemplar similarity. For example, an exemplar similarity score of 0.0 would indicate that the focal app does not contain any of the same words as any of the top 100 apps. An exemplar similarity score of 1.0 would indicate that the focal app has a text description that is identical to at least one of its category's top 100 apps. In our sample, the mean exemplar similarity score is 0.33.

4.2.3 | Control variables

We include several control variables to account for potential omitted variable bias and to control for other factors that may impact an app's performance on the market. We include several variables at the app level. First, we include the number of category name words used in the app's description (i.e., an app from the Books and Reference category that used the stemmed words "book" and "refer" in its description would contain a value of 2 for this variable). Second, we include the order of the app's entry on the market. This variable is measured by examining the "last updated date" for the app. Since these apps are all new on the market in the month in which we began our data collection, this variable not only serves as a proxy for the date when the app was uploaded to the marketplace but also helps control for possible learning effects from the developer. We also include the age of the app, which is the number of days between each wave and the first "last updated date" for each app. A binary indicator is also included to identify whether the app is free (vs. a paid app), and another indicator is used to show whether the app offers in-app purchases. Both of these variables are indicators that the app developer is trying to create value (although in-app advertising is also an important and growing source of revenue). We also control for the total length of each app's description (in characters) and the app size (in megabytes). We also include a binary variable that indicates whether the developer ever changed its description in the future. Approximately 75% of the apps in our sample never changed descriptions; however, controlling for future changes in our models is important. We also include a control for the review score for each app. To calculate this variable, we take the average of all reviews (between 1 and 5 stars) for each app in a given month. We lag this variable by 1 month, as a prior review score should have an impact on future reviews. We also include a binary indicator of whether the developer was considered one of Google's Top Developers since this status could increase the visibility and performance of that developer's apps. Finally, we include a content rating categorical variable that indicates the appropriate age range of the app, which includes the following five categories: everyone, kids, teens, adults, and unrated.

At the developer level, we first control for the percentage of a developer's apps that use category name words. Second, we control for the number of other categories in which the developer has apps for each month. This variable can vary from 1 (no spanning) to 10 (all apps in different categories; it is capped at 10 because we exclude de novo developers that enter Google Play with more than 10 apps). We also include a binary variable indicating whether the developer entered with two or more apps in its initial entry month.

We also include several controls at the category and market levels. These measures are calculated by drawing on our larger dataset and are not restricted to only the subsample of de novo app developers. First, we include the competitive density of each app category, which is the count of the total number of apps in each category in each observation month. Second, we include a measure of category contrast as this has been shown to impact the performance of category members (Kovács & Hannan, 2010; Negro, Koçak, & Hsu, 2010). To measure category contrast, we first calculate the total number of cleansed words for all apps in a given category and then determine what percent of these words (at the category level) are on the top 50 category words list. Higher scores for this variable indicate that the apps within the category are more similar. This variable is lagged by 1 month to better show the contrast of the category at the time that the app entered the market. We also include category-level fixed effects to control for all other factors that do not change within a given category over time. At the market level, we also include month-level fixed effects to account for temporal factors.

4.3 | Analysis technique

Our data consist of repeated measures of app performance over time; thus, we use panel data modeling techniques. Standard fixed effects regression models are inappropriate since our main independent variables do not vary over time (they measure similarity at app entry). Therefore, to test our hypotheses, we utilize time-series generalized estimating equations (GEE) originally developed by Liang and Zeger (1986), which account for potential autocorrelations in the data and calculate population average results. Our data include apps nested within a developer, which may cause our errors to be correlated between different apps from the same developer. Therefore, we also follow Cameron, Gelbach, and Miller (2011) and utilize cluster-robust *SEs* at the developer level as the developer is the highest level of nesting within our data. We further tested our models using the two-way clustered errors described by Cameron et al. (2011) but found similar results to the GEE regression models. To test the fit of our GEE models, we employ the quasi-likelihood under the independence model criterion command created by Cui (2007). With this method, lower values of the outcome indicate a better model fit. We find that the model including all of our controls and predictors (including the interaction) is the best fitting model.

5 | RESULTS

The correlations and summary statistics are provided in Table 1. When examining the correlations, we find that our two predictor variables (exemplar and prototype) are positively correlated (0.57). The scatterplot pattern in the supplemental analysis also highlights this correlation, with a large number of apps positioned low on both of these similarity measures. However, when only these main effects are included, we observe no evidence of multicollinearity (with variance inflation factors below 2). Since we hypothesize a multiplicative interaction, which increases the variance inflation factors (to a maximum of approximately 10), we mean-center our independent variables and include the interaction of the centered variables in our models. Models using the un-centered variables substantively yield the same results as those using the centered variables, but the centered variables lead to much lower variance inflation factors (below 2).

The results for the hypothesized effects are shown in Table 2. Model 1 introduces the control variables, which act in the expected manners. Interestingly, the lagged review score coefficient is negative and significant. One explanation for this finding is that most apps start with high review scores (most likely reviews from friends and family) and tend to drop as more people (who are likely more objective) install and review the app. This result is consistent with the findings of Kovacs and Sharkey (2014), who find that review scores fall over time, potentially because increased notoriety attracts a broader audience.

Model 2 introduces the main effects of prototype similarity, and as shown, the coefficient is negative (b = −0.205, $p = 0.000$). Model 3 adds the main effect of exemplar similarity. The coefficient for exemplar similarity is positive (b = 0.762, $p = 0.000$), while the prototype similarity remains negative. A figure in the supplemental analysis shows the average performance effect results for both similarity measures, holding all other variables constant. To ease interpretation, the dependent variable for this plot is the review count (as opposed to the log of review count). As shown in the supplemental analysis, an app that has a centered prototype similarity score of 0.0 (the mean prototype similarity score) has an average of 36.1 reviews, while an app with a centered prototype similarity score of 0.6 has an average of 25.6 reviews. Thus, a 0.6 increase in the prototype similarity score decreases the expected count of reviews by 10.5, or 29%. In contrast, positioning close to the exemplar leads to higher performance. An app that has a centered exemplar similarity score of 0.0 has an average of 35.1 reviews, while an app with a centered exemplar similarity score of 0.6 has an average of 58.2 reviews. Thus, a 0.6 increase in the exemplar similarity score increases the expected count of reviews by 22.1, or 61%, lending support to H1 and H2.

Turning to H3, we observe that in Model 4, the interaction effect between the two similarity measures is negative (b = −1.338, $p = 0.000$). To explore this effect in greater detail, Figure 2 plots the performance of an app with all control variables held at their means versus both exemplar and prototype similarity measures (interacted) in 3-dimensional space. As before, to facilitate interpretation, the outcome variable is the review count. As shown, the highest levels of performance are at low levels of similarity to the prototype (−0.2) and high levels of similarity to an exemplar (0.6), where an app has 76.7 reviews on average. This is the strategy portrayed as entrant 1 in Figure 1. However, the beneficial effect of similarity to the exemplar diminishes considerably across the range of similarity levels to the prototype. Apps utilizing the strategy shown as entrant 4 in Figure 1, with high exemplar similarity (0.6) and high prototype similarity (0.6), receive 26.4 reviews on average

TABLE 1 Correlations and summary statistics

	Variable	Mean	*SD*	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1)	Review count (ln)	2.97	2.05	−4.61	15.55																		
(2)	Prototype similarity	0.22	0.13	0	1	0.01																	
(3)	Exemplar similarity	0.33	0.17	0	1	0.16	0.57																
(4)	Category density (*10ˆ-3)	72.24	46.37	1.19	173.97	−0.09	−0.23	−0.2															
(5)	Category contrast	0.21	0.05	0.14	0.36	0.07	0.37	0.33	−0.5														
(6)	Developer category spanning	1.88	1.24	1	10	−0.1	0.08	0.08	−0.05	0.07													
(7)	Uses category names	0.49	0.5	0	1	0.07	0.26	0.29	−0.34	0.3	0.11												
(8)	Percent developer apps w/category names	49.26	42.38	0	100	0.08	0.23	0.26	−0.34	0.28	0.13	0.85											
(9)	Review average (lag)	4.29	0.73	1	5	−0.13	0.01	−0.04	0.03	−0.03	0	0	−0.01										
(10)	App age (days)	224.86	129.44	5	571	0.06	−0.02	−0.07	0.1	0	−0.02	−0.03	−0.03	−0.1									
(11)	Free app	0.96	0.19	0	1	0.11	0.03	0.02	0	−0.06	0.07	0.03	0.03	−0.02	−0.02								
(12)	Offer in-app purchases	0.12	0.33	0	1	0.17	0.02	0.06	−0.15	0.11	−0.03	0.15	0.16	−0.01	−0.03	0.03							
(13)	Cleansed description length (char)	482.26	414.22	10	4,636	0.16	−0.09	0.21	−0.05	0.02	−0.08	0.2	0.18	−0.04	−0.05	−0.07	0.12						
(14)	App size (MB)	13.22	33	0	2,355.2	0.09	−0.02	0	−0.1	0.06	−0.04	0.11	0.12	−0.01	−0.01	−0.09	0.16	0.08					
(15)	App entry order	2.18	1.75	1	10	−0.1	0.08	0.09	0.01	0.08	0.44	0.09	0.09	−0.06	−0.15	0.02	−0.03	−0.02	−0.02				
(16)	Changes description	0.49	0.5	0	1	−0.1	0.1	0.1	0	0.09	0.4	0.09	0.1	−0.06	−0.05	−0.01	−0.07	−0.05	−0.04	0.44			
(17)	Developer enters with >1 app	0.34	0.47	0	1	0.15	0	0.04	−0.02	0.03	−0.05	−0.02	−0.01	0.01	0.07	−0.06	0.05	0.08	0.02	−0.06	−0.06		
(18)	Top developer	0	0.04	0	1	0.11	−0.01	0.02	−0.03	0.01	−0.01	0.03	0.03	0	0	−0.01	0.08	0.04	0.07	0	−0.03	0.02	
(19)	Content rating (numerical)	1.3	0.81	1	5	0.06	−0.01	0	−0.1	0.01	−0.01	0.03	0.04	−0.03	0.06	0.01	0.06	0.01	0.06	−0.02	−0.01	−0.02	0.02

TABLE 2 Generalized estimating equation (GEE) regression models with robust SEs clustered at the developer level; the dependent variable is log review counts

Variables	Model 1	Model 2	Model 3	Model 4
Prototype similarity (cen)		−0.205	−0.662	−0.577
		[0.057]	[0.065]	[0.066]
		(0.000)	(0.000)	(0.000)
Exemplar similarity (cen)			0.762	0.794
			[0.053]	[0.054]
			(0.000)	(0.000)
Prototype × Exemplar similarity				−1.338
				[0.274]
				(0.000)
Category density ($\times 10^{-3}$)	−0.000	−0.000	−0.000	−0.000
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.818)	(0.823)	(0.867)	(0.887)
Category contrast	1.011	1.032	0.978	1.004
	[0.377]	[0.377]	[0.377]	[0.377]
	(0.007)	(0.006)	(0.010)	(0.008)
Developer category spanning	−0.096	−0.096	−0.097	−0.097
	[0.005]	[0.005]	[0.005]	[0.005]
	(0.000)	(0.000)	(0.000)	(0.000)
Uses category names	−0.036	−0.025	−0.046	−0.049
	[0.019]	[0.019]	[0.019]	[0.019]
	(0.054)	(0.178)	(0.015)	(0.009)
Percent dev. apps w/category names	0.002	0.002	0.002	0.002
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
Review average (lag)	−0.177	−0.176	−0.176	−0.175
	[0.009]	[0.009]	[0.009]	[0.009]
	(0.000)	(0.000)	(0.000)	(0.000)
App age (days)	0.003	0.003	0.003	0.003
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
Free app	1.590	1.591	1.587	1.588
	[0.033]	[0.033]	[0.033]	[0.033]
	(0.000)	(0.000)	(0.000)	(0.000)
Offer in-app purchases	0.426	0.426	0.424	0.423
	[0.026]	[0.026]	[0.026]	[0.026]
	(0.000)	(0.000)	(0.000)	(0.000)
Cleansed description length (char)	0.001	0.001	0.001	0.001
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
App size (MB)	0.003	0.003	0.003	0.003
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
App entry order	−0.075	−0.075	−0.075	−0.075

TABLE 2 (Continued)

Variables	Model 1	Model 2	Model 3	Model 4
	[0.004]	[0.004]	[0.004]	[0.004]
	(0.000)	(0.000)	(0.000)	(0.000)
Developer enters with >1 app	−0.209	−0.206	−0.220	−0.218
	[0.015]	[0.015]	[0.015]	[0.015]
	(0.000)	(0.000)	(0.000)	(0.000)
Changes description	0.462	0.463	0.459	0.459
	[0.016]	[0.016]	[0.016]	[0.016]
	(0.000)	(0.000)	(0.000)	(0.000)
Top developer	2.614	2.610	2.582	2.576
	[0.444]	[0.444]	[0.439]	[0.439]
	(0.000)	(0.000)	(0.000)	(0.000)
Content rating	*Included*	*Included*	*Included*	*Included*
Category FE	*Included*	*Included*	*Included*	*Included*
Wave FE	*Included*	*Included*	*Included*	*Included*
Constant	1.420	1.403	1.461	1.476
	[0.097]	[0.097]	[0.097]	[0.097]
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	589,681	589,681	589,681	589,681
Number of developers	83,115	83,115	83,115	83,115
χ^2	28,808	28,897	29,106	29,141
p	0	0	0	0
QIC	4,142,051.508	4,145,527.633	4,095,845.709	4,091,937.002

Note. Robust *SEs* clustered on developer ID are in brackets, *p*-values are in parentheses. FE = fixed effects; MB = megabytes; QIC = quasilikelihood under the Independence Model Criterion.

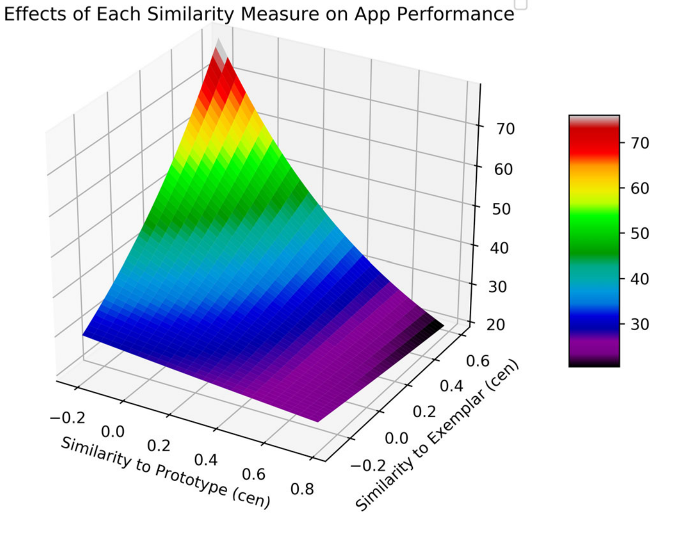

FIGURE 2 Three-dimensional plot showing the app average predicted performance (review count) of new de novo app entrants versus both prototypical and exemplar similarity scores (centered)

(a performance decrease of greater than 66%). In other words, high levels of similarity to the prototype almost completely erase the advantages of alignment with an exemplar. Indeed, the results in Figure 2 show that apps with high centered scores (0.6) for both similarity measures fare worse (26.4 reviews) on average than apps with low centered scores (−0.2) for both similarity measures (depicted as entrant 3 in Figure 1, with an average of 32.8 reviews). This finding highlights the interdependent nature of these two reference points and suggests that attempting to identify optimal positioning based on a single reference point within a market category is not sufficient. New entrants can benefit from developing an understanding of the multidimensional and complex nature of the market category that they are entering. Overall, these results provide support for H3.

5.1 | Robustness checks

We ran several additional models to further test our results. First, since our models predict *average* performance, we sought an empirical test that could predict *superior* performance. In the app store, the distribution of downloads is highly skewed, with only a small number of apps (approximately 4% of our sample) ever generating 100,000 or more downloads. Therefore, we created a binary variable for all apps that reach the 100,000-download level. We then ran a logistic regression model, with robust *SEs* clustered at the developer level. We retain only the last month's observation for each app. Table 3 shows this model, and the results indicate that our main findings still hold. High exemplar similarity coupled with low prototype similarity leads to the highest likelihood of reaching 100,000 or more downloads. Figure 3 shows the predicted probabilities of reaching this level of performance depending on the exemplar and prototype similarity scores. High levels of similarity to the exemplar

TABLE 3 Logistic regression models with robust SEs clustered at the developer level; the dependent variable is a binary indicator for apps reaching 100,000 or more downloads

Variables	Model 1
Prototype similarity (cen.)	−1.301
	[0.172]
	(0.000)
Exemplar similarity (cen.)	2.736
	[0.103]
	(0.000)
Prototype × Exemplar similarity	−1.876
	[0.622]
	(0.003)
Controls	*Included*
Content rating	*Included*
Category FE	*Included*
Wave FE	*Included*
Constant	−4.283
	[1.256]
	(0.001)
Observations	135,207
Pseudo R-squared	0.143
Log likelihood	−22,351

Note. Robust *SEs* clustered on developer ID are in brackets, *p*-values are in parentheses.

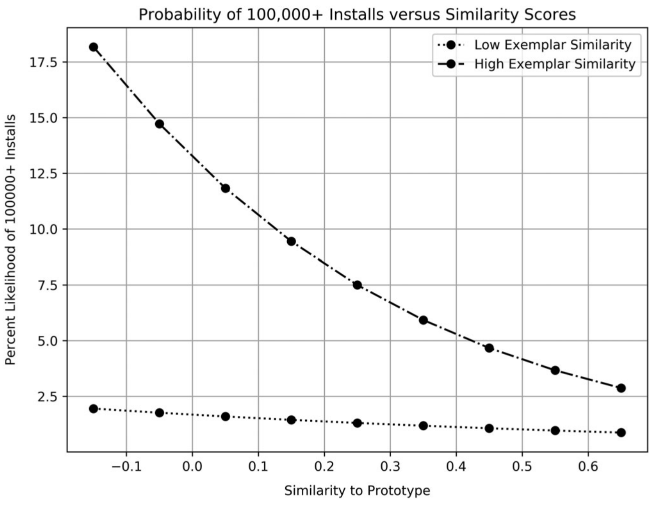

FIGURE 3 Probability of achieving 100,000 or more downloads by prototypical and exemplar similarity scores (centered)

and low levels of similarity to the prototype lead to an 18.1% likelihood of achieving 100,000 or more downloads. This effect is large since a randomly selected app has only a 3.6% likelihood of reaching this level of performance. As a focal app that is highly similar to an exemplar increases its similarity to the prototype, it loses essentially all of its increased probability of superior performance. This effect is consistent with the main models predicting average performance. Therefore, being properly aligned relative to both exemplar prototype categorical schemas not only improves a developer's likelihood of achieving better than average performance but also increases a developer's likelihood of achieving superior performance.

As a second robustness check, we run our original tests with an alternative dependent variable: the logged number of downloads. Consistent with our main models, we find that similarity to the prototype leads to lower downloads, that being more similar to the exemplar leads to an increase in downloads, and that a strong negative interactive effect exists. These results are available in the supplemental analysis. Third, we expanded our sample to include all new apps published by all developers during our data collection period (rather than restricting the sample to include only de novo developers). Due to the computational difficulties associated with running our GEE models on this larger sample with nearly 5 million app-month observations, we ran cross-sectional regression models retaining only the last wave of data for each app and clustering the errors at the developer level. These models produce coefficients that are similar to our main results in magnitude, direction, and significance.

Fourth, we test whether our results hold when looking only at the first month that an app is on the market, thus avoiding the possibility that a few high-performing apps, which are in our sample multiple times (across a number of months), are not driving our results. We run a cross-sectional regression model with the first month that each app is on the marketplace. These models include all of our controls from our main regression models (with the exception of the two lagged variables). These results, which are available in the supplemental analysis, show that our results hold. This finding indicates that being optimally positioned relative to the exemplar and prototype category members has an immediate impact on performance, and coupled with our main models in Table 2, suggests that this positive benefit persists over time.

Finally, we examine subsets of our data based on our category contrast variable since some of our categories are broadly defined while others are narrowly defined. The full results are available in the supplemental analysis. To examine this issue, we first run models on only the top 10 highest contrast categories (those that are most similar). In these results, our main predictions hold, but our interaction term (b $= -1.475$, $p = 0.106$) loses significance. We next ran models on only the bottom 10 lowest contrast categories, where all of our predictions held. We then ran two models with all of our categories except for those in the (a) top 10 or (b) bottom 10 in category contrast. All of our predictions held in these models. Finally, we ran two split-sample models, splitting at the mean for category contrast. In both the above-the-mean and the below-the-mean models, our predictions held. These analyses suggest that despite substantial variation in the contrast, or fuzziness, of Google Play's product market categories, such variation does not necessarily impact the degree to which app developers should position their products.

6 | DISCUSSION AND CONCLUSION

This paper addresses the important strategic question of to whom or to what do de novo entrants conform to and differentiate from to gain a competitive advantage in product market categories. To answer this question, we draw on the strategic categorization literature and identify two categorical schemas that organizations can align with or differentiate themselves from: a category exemplar and the category prototype. Our findings show that these two categorical schemas are interdependent and highlight the importance of considering both in tandem when making entry positioning decisions. We argue and find support for the notion that organizations can strategically categorize themselves and endogenously influence demand for their products in two-sided platform ecosystems such as the Google Play app store. In particular, we find that alignment with the category prototype in a platform market may not confer the same benefits as in more traditional markets. We also find that aligning with a category exemplar leads to higher performance outcomes for organizations, but that this effect is fully negated when alignment with the prototype is also high. Thus, this research investigates the multidimensionality and interdependence of different strategic positions within a product market category at the time of entry and shows the importance of optimizing relative to multiple categorical schemas simultaneously to gain a competitive advantage.

This paper adds to several current research streams. First, we provide contributions to the literature on optimal distinctiveness (Deephouse, 1999; Haans, 2019; Zhao et al., 2017, 2018) and organizational positioning at market entry. Our research shows that an organization's positioning strategy within a market category is not simply a zero-sum game consisting of tradeoffs between conformity to and differentiation from a single reference point. Instead, organizations should consider multiple categorical schemas when positioning their products.

Second, we build on recent research in the conformity literature (Durand & Kremp, 2016; Kim & Jensen, 2011; E. B. Smith, 2011; Zhao et al., 2018) by examining the role of category exemplars in optimal positioning strategies. Furthermore, we distinguish for the first time the concurrent effects of conformity to both the category prototype (i.e., alignment) and the category exemplar (i.e., conventionality), and demonstrate that benefits positioning benefits can vary greatly when taking into account both categorical reference points. In doing so, we extend recent findings in an important manner, directly answering their call to “continue to explore the relationship between prototype and exemplar models in other contexts” (Zhao et al., 2018: 19). However, future research could examine the evolution of positioning strategies across the growth and development of market categories to further our understanding of this phenomenon.

Third, we add to the emerging literature on strategic categorization (Cattani et al., 2017; Durand & Khaire, 2017; Khaire, 2018; Pontikes, 2018; Pontikes & Kim, 2017; Vergne & Wry, 2014). Past research argues that categories are used by audiences to gain information and that organizations can strategically leverage semantic cues and categorical nomenclature to communicate with and help shape how consumers view them. In our empirical setting, app developers choose how to describe their app, highlighting the features or characteristics that they perceive to be most important or salient, ultimately signaling to consumers how they fit within the competitive framework of a given market category. Our results show that this process has a tangible impact on an app's level of competitive advantage with respect to both average performance and superior performance. This finding supports the notion that organizations can use strategic categorization "to communicate information and to position themselves favorably with respect to competitors" (Pontikes & Kim, 2017: 73).

Finally, we add to the nascent and growing literature on platform ecosystems (McIntyre & Srinivasan, 2017; Rietveld & Eggers, 2018). To date, this literature has primarily has adopted the platform as the unit of analysis and has tended to focus on either the strategies that a platform can adopt to achieve a competitive advantage vis-à-vis rivals or the strategies that a platform can enact to solve the "chicken-or-the-egg" problem associated with two-sided business models (Cennamo & Santalo, 2013; McIntyre & Srinivasan, 2017; Rietveld & Eggers, 2018; Zhu & Iansiti, 2012). Our study is among the first to specifically examine intra-platform competition among producers. We show that in these intensely competitive environments, the entry positioning choices of de novo producers can have a substantial effect on their performance. Specifically, we show that de novo producers can increase their likelihood of success by positioning themselves close to the category exemplar while simultaneously positioning far from the category prototype.

This research also has important managerial implications. In our sample, most new entrants do not position themselves close to an exemplar. In other words, most new entrants employ a suboptimal positioning strategy. Managers would be well advised to attempt to understand the overall competitive context of a given market category prior to entry. Focusing on a single positioning strategy relative to one categorical schema is likely to be less effective than adopting a strategy that considers multiple categorical schemas. We provide evidence suggesting that managers can take control of how information about their organizations' products are communicated to consumers and how this information can be used to influence demand for these products and achieve superior performance. These tasks can be performed by utilizing semantic cues and adopting categorical nomenclatures to engage consumers and to optimally position product offerings within a given category structure. Understanding how to position products such that they appear higher in the platform's search rankings increases the probability that a product will gain attention from and ultimately be selected by consumers who may tend to satisfice and to stop searching once they find a suitable product that fits their needs.

This research can likely be generalized beyond the Google Play app store. In particular, our findings should be generalizable to other two-sided platforms. As noted above, platform ecosystems are becoming an increasingly important part of the current economy. As competition on these platforms becomes fiercer, paying close attention to positioning choices to gain attention and avoid getting lost in the crowd will become increasingly important for producers.

As with all studies, there are limitations inherent in our study that provide opportunities for future research. For example, we cannot directly observe Google Play's search algorithms. Consequently, we cannot control for "top apps" or apps listed as "similar to" major exemplars. Additionally, although we test the robustness of our findings on high- and low-contrast subsamples, we cannot entirely eliminate the alternative explanation that category contrast may have partially driven our findings on the negative effect of prototype similarity. While we posit that our results are generalizable beyond the Google Play app store, more research is needed to substantiate our findings in other contexts. Future research could test these same processes in other markets where strategic categorization is possible. Additionally, it is unclear whether our theory is generalizable to nascent market categories or categories that are thinly populated, in part because category exemplars and prototypes may not have had time to develop. While we show how app developers can align their apps relative to category exemplars and prototypes, we do not fully know the extent to which developers actually consider these categorical schemas when publishing a new app. Future research could build on these findings by qualitatively gauging how strategic organizations are in their categorization and positioning decisions. Finally, prior research has shown that different audiences can view the same category differently (Durand & Paolella, 2013); thus, the optimal position for critics or other audiences may be different from the optimal position for consumers. Future research should examine this question.

ACKNOWLEDGEMENTS

The authors would like to thank associate editor Rodolphe Durand and two anonymous reviewers for their helpful comments in improving this article. We would like to thank Leif Lundmark, Filippo Wezel, and Özgecan Koçak, for helpful comments on previous versions of this article. We would also like to thank seminar participants at the University of Utah, participants at the 2017 Strategic Management Society annual meeting, and participants at the 2018 Western Academy of Management annual meeting for their thoughtful comments. Finally, we wish to thank Marcello Lins, without whom this paper would not exist.

ORCID

Matthew A. Barlow https://orcid.org/0000-0002-7186-571X

REFERENCES

Baye, M. R., & Morgan, J. (2001). Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *The American Economic Review*, *91*(3), 454–474.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Carroll, G. R., & Hannan, M. T. (1989). Density dependence in the evolution of populations of newspaper organizations. *American Sociological Review*, *54*(4), 524–541.

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

Cennamo, C., & Santalo, J. (2013). Platform competition: Strategic trade-offs in platform markets. *Strategic Management Journal*, *34*(11), 1331–1350.

Cohen, J. B., & Basu, K. (1987). Alternative models of categorization: Toward a contingent processing framework. *Journal of Consumer Research*, *13*(4), 455–472.

Cui, J. (2007). QIC program and model selection in GEE analyses. *Stata Journal*, *7*(2), 209–220.

Deephouse, D. L. (1999). To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*, *20*(2), 147–166.

Dimaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Dobrev, S. D., & Kim, T.-Y. (2006). Positioning among organizations in a population: Moves between market segments and the evolution of industry structure. *Administrative Science Quarterly*, *51*(2), 230–261.

Durand, R., & Khaire, M. (2017). Where do market categories come from and how? Distinguishing category creation from category emergence. *Journal of Management*, *43*(1), 87–110.

Durand, R., & Kremp, P.-A. (2016). Classical deviation: Organizational and individual status as antecedents of conformity. *Academy of Management Journal*, *59*(1), 65–89.

Durand, R., & Paolella, L. (2013). Category stretching: Reorienting research on categories in strategy, entrepreneurship, and organization theory. *Journal of Management Studies*, *50*(6), 1100–1123.

Haans, R. F. J. (2019). What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories. *Strategic Management Journal*, *40*(1), 3–27.

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

Haveman, H. A. (1993). Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly*, *38*, 593–627.

Hiatt, S. R., & Carlos, W. C. (2018). From farms to fuel tanks: Stakeholder framing contests and entrepreneurship in the emergent U.S. biodiesel market. *Strategic Management Journal*. https://doi.org/10.1002/smj.2989

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, *23*(10), 3773–3811.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, *39*(153), 41–57.

Hsu, G. (2006a). 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. (2006b). Evaluative schemas and the attention of critics in the US film industry. *Industrial and Corporate Change*, *15*(3), 467–496.

Kennedy, M. T. (2008). Getting counted: Markets, media, and reality. *American Sociological Review*, *73*(2), 270–295.

Khaire, M. (2018). Entrepreneurship by design: The construction of meanings and markets for cultural craft goods. *Innovation*, *21*(1), 13–32.

Khaire, M., & Wadhwani, R. D. (2010). Changing landscapes: The construction of meaning and value in a new market category—modern Indian art. *Academy of Management Journal*, *53*(6), 1281–1304.

Kim, B. K., & Jensen, M. (2011). How product order affects market identity: Repertoire ordering in the U.S. opera market. *Administrative Science Quarterly*, *56*(2), 238–256.

Klepper, S. (1997). Industry life cycles. *Industrial and Corporate Change*, *6*(1), 145–182.

Kovács, B., & Hannan, M. T. (2010). The consequences of category spanning depend on contrast. In *Categories in markets: Origins and evolution* (pp. 175–201). Bingley, UK: Emerald Group Publishing Limited.

Kovacs, B., & Sharkey, A. J. (2014). The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative Science Quarterly*, *59*(1), 1–33.

Lee, B. H., Hiatt, S. R., & Lounsbury, M. (2017). Market mediators and the trade-offs of legitimacy-seeking behaviors in a nascent category. *Organization Science*, *28*(3), 447–470.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

McIntyre, D. P., & Srinivasan, A. (2017). Networks, platforms, and strategy: Emerging views and next steps. *Strategic Management Journal*, *38*(1), 141–160.

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. (2010). Categorical contrast and audience appeal: Niche width and critical success in winemaking. *Industrial and Corporate Change*, *19*(5), 1397–1425.

Negro, G., Koçak, Ö., & Hsu, G. (2010). Research on categories in the sociology of organizations. In *Categories in markets: Origins and evolution* (pp. 3–35). Bingley, UK: Emerald Group Publishing Limited.

Oliver, C. (1997). Sustainable competitive advantage: Combining institutional and resource-based views. *Strategic Management Journal*, *18*(9), 697–713.

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.

Pontikes, E. G. (2018). Category strategy for firm advantage. *Strategy Science*, *3*(4), 620–631.

Pontikes, E. G., & Kim, R. (2017). Strategic categorization. In *From categories to categorization: Studies in sociology, organizations and strategy at the crossroads* (pp. 71–111). Bingley, UK: Emerald Publishing Limited.

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.

Rietveld, J., & Eggers, J. P. (2018). Demand heterogeneity in platform markets: Implications for complementors. *Organization Science*, *29*(2), 304–322.

Rosch, E., & Lloyd, B. B. (1978). *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Smith, E. B. (2011). Identities as lenses: How organizational identity affects audiences' evaluation of organizational performance. *Administrative Science Quarterly*, *56*(1), 61–94.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, *20*(3), 571–610.

Thomas, L. D., Autio, E., & Gann, D. M. (2015). Architectural leverage: Putting platforms in context. *The Academy of Management Perspectives*, *3015*(1), 47–67.

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.
Vergne, J.-P., & Wry, T. (2014). Categorizing categorization research: Review, integration, and future directions. *Journal of Management Studies*, *51*(1), 56–94.
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). An exemplar-based model of proto-category evolution: Strategic differentiation and performance in console video game industry. *Organization Science*, *29*, 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.
Zhu, F., & Iansiti, M. (2012). Entry into platform-based markets. *Strategic Management Journal*, *33*(1), 88–106.
Zuckerman, E. W., Kim, T.-Y., Ukanwa, K., & von Rittmann, J. (2003). Robust identities or nonentities? Typecasting in the feature-film labor market. *American Journal of Sociology*, *108*(5), 1018–1073.
Zuckerman, E. W. (1999). The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology*, *104*(5), 1398–1438.
Zuckerman, E. W. (2016). Optimal distinctiveness revisited: An integrative framework for understanding the balance between differentiation and conformity in individual and organizational identities. In M. G. Pratt, M. Schultz, B. E. Ashforth, & D. Ravasi (Eds.), *The oxford handbook of organizational identity* (pp. 183–199). Oxford, UK: Oxford University Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Barlow MA, Verhaal JC, Angus RW. Optimal distinctiveness, strategic categorization, and product market entry on the Google Play app platform. *Strat Mgmt J*. 2019;40:1219–1242. https://doi.org/10.1002/smj.3019