Overcoming the Illegitimacy Discount: Cultural Entrepreneurship in the US Feature Film Industry

Eric Yanfei Zhao
University of Alberta School of Business, National Institute for Nanotechnology, Canada

Masakazu Ishihara
New York University Stern School of Business, USA

Michael Lounsbury
University of Alberta School of Business, National Institute for Nanotechnology, Canada

Abstract
How can organizations spanning institutionalized categories mitigate against the possibility of reduced attention by audiences? While there has been a good deal of research on the illegitimacy discount of category spanning, scant attention has been paid to how organizations might strategically address this potential problem. In this paper, we explore how the strategic *naming* of products might enhance audience attention despite the liabilities associated with category spanning. Drawing on a sample of films released in the United States market between 1982 and 2007, we analyze different naming strategies and show that names that simply signal familiarity are not potent enough to offset the illegitimacy discount, while names imbued with known reputations serve as a symbolic device that enhances audience attention to genre-spanning films.

Keywords
attention, categories, cultural entrepreneurship, film industry, legitimacy, naming, reputation

Gaining attention of audiences such as customers, rating agencies, potential employees, and investors is an important task for organizations, and has critical consequences for organizational survival and performance (e.g., Lounsbury & Glynn, 2001; Oliver, 1991; Pfeffer & Salancik, 1978). Yet, this task is difficult because multiple organizations and products tend to compete for the limited attention of audiences. Faced with many alternatives with unclear inherent quality, audiences have been shown to rely on institutionalized category schemes to simplify thought and

Corresponding author:
Eric Yanfei Zhao, University of Alberta School of Business, 2-24 Business Building, Edmonton, Alberta, T6G 2R6, Canada.
Email: eric.zhao@ualberta.ca

allocate attention (Kennedy, 2005). Categories help audiences by lumping similar organizations and products together, demarcating them from different ones, and creating shared understandings and expectations for those entities contained within the same category (Zerubavel, 1997). Categorization has been a prominent theme in institutional approaches that claim that categories influence perception, interpretation, and action, and provide the default conditions for making sense of the social world (e.g., DiMaggio, 1997; Scott, 2001; Weber & Glynn, 2006). Institutionalists have argued that categories shape imitation (DiMaggio & Powell, 1983), circum-scribe the range of appropriate strategic actions for firms (Kraatz & Zajac, 1996), and restrict variety (Powell, 1991).

Drawing on this work, Zuckerman (1999) prominently demonstrated that entities that span mul-tiple categories suffer from reduced audience attention—a phenomenon he termed the "illegiti-macy discount." As a result of this persuasive finding, many scholars have stressed the need for organizations and products to be unambiguously situated in a particular category, and to operate in accordance with the norms and expectations associated with that category (Glynn, 2008; Navis & Glynn, 2011; Pedersen & Dobbin, 2006). While evidence in support of the illegitimacy discount has amassed across a range of contexts (Hannan, 2010), some studies have highlighted that, under certain circumstances, organizations and products can successfully expand their offerings across categories (e.g., Pontikes, 2012), or (re)combine elements of different categories to expand market niches and gain broader audience attention (e.g., Jensen, 2010; Karthikeyan & Wezel, 2010). Others have argued that the illegitimacy discount may become attenuated in emergent categories, as categories become de-institutionalized, or as boundaries segregating categories become blurred and eroded (e.g., Kennedy, Lo, & Lounsbury, 2010; Kovács & Hannan, 2010; Rao, Monin, & Durand, 2005; Rosa, Porac, Runser-Spanjol, & Saxon, 1999; Ruef & Patterson, 2009; Wry & Lounsbury, 2013).

However, virtually no attention has been paid to how actors might act strategically to mitigate against the illegitimacy discount. In reviewing the literature on institutions and identity, Glynn (2008) highlights that while the pressures to conform to institutional expectations are profound, it is equally critical for organizations and products to gain some level of distinctiveness to stand out from the crowd. Lounsbury and Glynn (2001) argue that legitimacy can be attained and resources acquired by strategic differentiation that remains within the confines of institutional appropriate-ness—what Brewer (1991) refers to as "optimal distinctiveness" (see also Deephouse, 1999, on balance theory). While Lounsbury and Glynn (2001) initially focused on storytelling to illustrate processes of cultural entrepreneurship, the cultural entrepreneurship literature has expanded to address how a wide variety of cultural resources such as vocabularies, logics, naming, and dis-course could be strategically deployed to shape audience attention in order to achieve desirable ends (Jones, Maoret, Massa, & Svejenova, 2012; Loewenstein, Ocasio, & Jones, 2012; Phillips, Lawrence, & Hardy 2004; Thornton, Ocasio, & Lounsbury, 2012; Weber & Dacin, 2011). We seek to explore how cultural entrepreneurship, involving the strategic use of such cultural resources, might circumvent the illegitimacy discount and enable increased audience attention (Ocasio, 2011).

As such, our paper speaks to the more general interest in organization studies of how sensemak-ing and sensegiving processes can resolve ambiguity (Weber & Glynn, 2006) or normalize devi-ance (Lounsbury & Crumley, 2007), as well as how extant institutional resources can be drawn upon to legitimize something that is new or not well understood (Jones, 2001). We study how the illegitimacy discount can be mitigated in a context where it has been documented to operate—the US feature film industry. Like previous studies, our level of analysis is the film because "expectations and evaluations of the film-going experience are generally formed at the product level, in the sense

that the characteristics of a film tend to be more salient than the characteristics of the producing and distributing organizations" (Hsu, Hannan, & Koçak, 2009, p. 157). That is, audiences allocate attention at the level of the film in this particular setting (Zuckerman & Kim, 2003).

In accordance with the findings of extant studies, we suggest that genre-spanning films will in general suffer an illegitimacy discount as reflected in reduced attention by consumers (see Hsu, 2006; Hsu et al., 2009). However, we suggest that consumer inattention to films can be overcome by the strategic use of a particular cultural element—*name*. The name of a movie is especially important in shaping audience perceptions of it. For instance, Wallace Wang notes that:

> A good title hints what the movie is about and makes you want to see it. A bad title just leaves you scratching your head, wondering whether the movie is a western, a chick flick, an action movie, or a science fiction thriller. If you don't know what you're going to get, you probably aren't going to bother trying it. (www.15minutemoviemethod.com/Blog/files/94ffdf4df99b5673aa4aba6906a881d5-13.html)

Grounded in our empirical context—the US feature film industry—we examine the naming of movies. We argue that there are two different kinds of naming strategies that can increase audience attention by either enhancing *familiarity* or tapping into an established *reputation*. We expect that films with names that simply share some elements with prominent, preexisting films, even without direct connections in contents, can invoke familiarity and attract consumer attention. Thus, compared with a genre-spanning film that adopts a relatively unique name unconnected to previous films, genre-spanning films that draw upon such recognized name elements are less likely to be ignored by consumers (see Glynn & Abzug, 2002).

By contrast, genre-spanning sequels that invoke the names of their predecessors can enhance audience attention by invoking a more particularistic reputation connected to a film series (e.g., Fombrun & Shanley, 1990). Compared with those films that share similar names with preexisting but unrelated films, film sequels named after their predecessors can resonate with audiences in specific and more predictable ways since films in a series may share the same or similar characters, actors, plots, and themes. In addition, we argue that since film series may have variable reputations, audience attention may be further enhanced as a result of the higher market reception of predecessor films.

Based on an analysis of all movies released in the US market between 1982 and 2007, we find that names that simply signal familiarity are not potent enough to offset the "illegitimacy discount," while names that encode more particularistic reputations serve as a symbolic device and interpretive lens that increases audience attention, and can enable a film to overcome the "illegitimacy discount." We conclude the paper by discussing the implications of our findings for the literature on categorization dynamics and cultural entrepreneurship.

Categories, Category Spanning, and the Illegitimacy Discount

Studies have shown that category systems enable the smooth functioning of complex fields and markets because they segment organizations and products into comparison sets. Through this act of grouping, categories provide a cognitive infrastructure which conveys the appropriate attributes and features of its members, and help audiences to make quick distinctions about "types" of products and organizations (Douglas, 1986; Zerubavel, 1997; Zuckerman, 1999). For instance, producers in Scottish knitwear focused their limited attention on a restricted range of competitors

by grouping them into distinct categories such as "handknitters making traditional knitwear," "handknitters making designer knitwear," and "mass-market contract knitters" (Porac, Thomas, Wilson, Paton, & Kanfer, 1995). Categorization thus simplifies thought by delimiting the locus of audience attention and enables them to process vast amounts of information more quickly and with reasonable efficiency (Rosch & Lloyd, 1978; Simon, 1947).

Because of their critical role in shaping audience attention, categories provide a baseline for understanding processes related to legitimacy—defined as the general perception that an organization or product is proper or appropriate within a "socially constructed system of norms, values, beliefs, and definitions" (Suchman, 1995, p. 574)—of the various organizations and products that populate a field. The general argument of Zuckerman (1999) is that entities which span multiple categories are perceived as less legitimate because audiences have difficulty comprehending their identities and, as a result, allocate their attention elsewhere. Zuckerman termed this phenomenon the "categorical imperative," noting that entities which:

> do not exhibit certain common characteristics may not be readily compared to others and are thus difficult to evaluate. Such offers stand outside the field of comparison and are ignored as so many oranges in a competition among apples. It is this *inattention* that constitutes the cost of illegitimacy. (Zuckerman, 1999, p. 1401)

In many studies, including Zuckerman's, *attention* is often used as a concrete dependent variable to infer legitimacy or lack thereof.

The cost of illegitimacy—the illegitimacy discount—has been validated across a wide range of empirical settings, with studies consistently finding that organizations or products are viewed as legitimate when they convey membership in a specific category, while those with attributes that span categories suffer from reduced attention, discounted evaluations, and poorer performance (see Hannan, 2010, for a review). For example, Zuckerman (2000) showed that firms which straddled industry categories hindered efforts by security analysts to evaluate their shares and thus faced pressure from analysts to de-diversify. Phillips and Zuckerman (2001) found that middle-status law firms in Silicon Valley were less likely to diversify into family law practice to avoid the risk of being screened out of consideration by corporate clients and elite law schools. Ruef and Patterson (2009) found that hybrid organizations received negative credit coverage and evaluation when they crossed institutionalized industry boundaries. Likewise, Zuckerman, Kim, Ukanwa, and von Rittmann (2003) observed that novice actors without a distinct genre-based identity faced difficulties in securing future acting jobs.

One context where category effects have been shown to be particularly robust is in cultural industries. In these contexts, genres provide categories that convey important information about the underlying features of different offerings (DiMaggio, 1987). This plays an important role in consumption decisions because each genre is associated with a specific audience and their particular tastes. For example, films that are classified under the same genre share a variety of features such as "the nature of the protagonist and antagonist, the structure of dramatic action, the catalytic event, narrative style and structure, and tone" (Hsu, 2006, p. 427). In contradistinction, films classified under different genres present different dramatic plots and resonate with audiences in distinct ways. Evidence suggests that genre is one of the most important factors that audiences consider when deciding which films to attend (Austin, 1989) because genres allow audience members to make inferences about different offerings as a base step in their selection process (Hsu, 2006; Hsu et al., 2009). In line with more general research on the "illegitimacy discount," studies have documented that genre-spanning films receive lower audience ratings and box office revenue (Hsu, 2006; Hsu et al., 2009).

Naming and Audience Attention

As implied by the work of Zuckerman (1999) and others, to overcome the illegitimacy discount, solving the problem of audience *inattention* is paramount. Thus, gaining the attention of audiences is a critical first step upon which further quality distinctions and evaluations are typically made (Hsu & Hannan, 2005; Phillips & Zuckerman, 2001). Past studies have suggested that it is not just certain well-established categories that shape audience attention (Porac et al., 1995); audiences perceive and afford attention to organizations, products and events based on a variety of stimuli (Ocasio, 2011). Given this premise, we argue that when category spanning is associated with the prospect of reduced audience attention—and its attendant economic consequences—organizations can act strategically to offset such inattention.

In particular, cultural elements may be strategically deployed to signal their appropriateness, gain audience attention, and garner support (Scott, 2001; Suchman, 1995). Cultural elements include names, stories, elements of logics, narratives, rhetoric, symbols, rituals, and vocabularies (e.g., Lounsbury & Glynn, 2001; Loewenstein et al., 2012; Taylor & Whittier, 1992). These cultural elements provide resources which can be strategically deployed by organizations to engage audience attention (Rao, Monin, & Durand, 2003). For example, Jones (2001) found that in the content era of the American film industry, entrepreneurs wittingly imitated high cultural symbols and formats of Broadway theaters such as uniformed ushers, plush chairs, two-hour shows, and elaborate buildings. By tapping these cultural resources, they evoked cognitive heuristics from consumers, gained their attention, and ultimately legitimated the emerging industry. Similarly, Jones, Livne-Tarandach, & Balachandra (2010) found that architects' vocabularies can persuade clients of their competence and gain opportunities, and that the most effective architects deployed vocabularies strategically.

Given that we are studying films, we focus on name as a key symbolic marker that can be strategically devised to trigger audience attention. Movie names can suggest images and set expectations of what a movie is about:

> When I hear a magnificently evocative real title—like this week's documentary about the genocide in Darfur, *The Devil Came On Horseback*—I can't help imagining what might be the perfect movie under that title…when I hear those words, they conjure up the authentic, atmospheric tang of a 1940s Warner Bros proto-noir western, like the Raoul Walsh films *Pursued* and *Colorado Territory*. (Patterson, 2008)

Thus, different names signal information with more or less visibility, more or less credibility, and more or less favorableness to audiences. In general, audiences find it easier to notice an organization or product and isolate it from the rest of the social world when it has a name that they recognize (Zerubavel, 1997). It has been argued that names can increase familiarity (e.g., Glynn & Abzug, 2002) or signal reputation (e.g., Fombrun & Shanley, 1990)—two discrete mechanisms that can affect audience attention. With regard to familiarity, names gain audience attention when they are notable and audiences have been previously exposed to them (Glynn & Abzug, 2002; Pollock & Rindova, 2003; Suchman, 1995). For instance, Glynn and Abzug (2002) found that organizations changed their names in terms of name length, ambiguity, and domain specificity to conform to the prototypical name styles in the industry, and this conformity contributed to the public understandability of and attention to these names. Likewise, Lee (2001) argued that announcements of name changes to include ".com" gained audience attention by identifying a firm with the internet and sending a clear signal about the universe in which the firm competed. This signal in turn led to increased investor interest and positive stock price reactions. Phillips and Kim (2009) suggested that organizations sometimes employed pseudonyms in order to preserve identity

and seek alignment with the institutional environment, even though the underlying production was decoupled from such an identity.

Given that prominent naming elements can increase audience attention, we should expect that audiences are less likely to ignore category-spanning organizations and products in the first place if they utilize well-understood and appreciated name elements. Prominent naming elements thus provide a means for enabling a "deviant innovation" to become recognized and to dampen the potential penalty associated with that innovation (Loewenstein et al., 2012). In the context of films, it is clear that copycat naming abounds. For example, a variety of films invoked the naming element "lethal" after the very successful movie, *Lethal Weapon* (1987), starring Mel Gibson; these include *Lethal Combat, Lethal Force, Lethal Ninja, Lethal Lolita, Lethal Seduction, Lethal Justice, Lethal Tender, L.E.T.H.A.L., Barely Lethal, Lethal Eviction*, and *Lethal Obsession*.[1] Although these later films rarely share any direct connections (in themes, plots, etc.) with *Lethal Weapon*, the common naming element "lethal" helped them gain audience attention. Hence, we expect that genre-spanning films will have a better chance of being attended to and thus have higher box office revenue when their name invokes similarities with prominent, preexisting film names. Note that our focus on an outcome of attentional processes (box office revenues) is consistent with the stream of literature on "attentional selection" (Ocasio, 2011; see also Greve, 2008; Hansen & Haas, 2001; Hoffman & Ocasio, 2001). Therefore, we propose:

> *Hypothesis 1 (H1)*: The negative effect of genre spanning on audience attention (a film's box office revenue) will be reduced when a film's name uses familiar name elements from prominent preexisting film names.

Compared with names that provide pure familiarity, some names can also be infused with reputational capital if they are connected to a preexisting entity such as when a newly created subsidiary maintains the name of a highly regarded parent firm (see, e.g., Fombrun & Shanley, 1990). Names that invoke a distinct reputation may resonate with audiences in specific ways and help them recognize and understand the named entity with more certainty. Reputation is important because it differentiates between the "qualified" and the "outstanding" (Lawrence, 1998) and affects an organization's favorableness to audiences (Deephouse, 2000). Reputation can also serve as a sensemaking device that triggers attention and helps audiences to make value-based distinctions (Fiske & Taylor, 1991; Gioia & Thomas, 1996; Weick, 1995). As a result, audiences may draw more favorable interpretations of an organization or product with a clearly reputable name.

Past studies have shown that organizations strategically use name as a reputation marker to gain audience attention and appreciation. For example, Ingram (1996) suggested that US hotel chains used the same name across branches to signal their connections and relatedness, which are recognized and rewarded by consumers. Similarly, Zuckerman and Kim (2003) indicated that independent film distributors strategically kept their label affiliations with their major studio owners in order to break out into the mainstream film market. A consistent message out of these studies is that names that signal authentic connections with reputable preexisting entities consequently shape audiences' attention (Baron, 2004; Hsu & Hannan, 2005; Peterson, 1997).

In the movie industry, film sequels typically rely on the reputation they have established among filmgoers. The same name used across films in a series draws attention to a specific theme or type (e.g., "Harry Potter" or "007" films). Compared with films that simply share some name elements with prominent preexisting but unrelated films, film sequel names provide unambiguous linkages to past movies, and thus are more likely to attract filmgoers' attention. As a result, sequel naming is a strategy that may mitigate the illegitimacy discount associated with genre spanning by enhancing the perceived reputation of a film. Therefore, we propose:

Hypothesis 2 (H2): The negative effect of genre spanning on audience attention (a film's box office revenue) will be reduced when reputation is embedded in a film's name by clearly identifying it as part of a film series.

Of course, reputations of organizations and products can vary, enabling entities to be compared and rank ordered (Power, Scheytt, Soin, & Sahlin, 2009). Thus, we believe that the reputational effect of sequel naming will be enhanced if the predecessor movie in a series was a bigger box office hit. While performance is not the only dimension that affects reputational judgments, it has been shown to be an especially important one (Benjamin & Podolny, 1999; Bitektine, 2011; Podolny & Phillips, 1996). Accordingly, names identified with highly successful predecessors should better mitigate against the illegitimacy discount, and may even foster favorable interpretations of category spanning.

This may occur because a highly reputable name increases the perceived value of the focal organization or product through a "halo effect" (Crane, 1965; Perrow, 1961). Systemic findings in the past have confirmed that positive biases often creep into the process of evaluating highly reputable entities (Blank, 1991; Cole, Cole, & Simon, 1981; Goldin & Rouse, 2000). In addition, a highly reputable name increases the credibility of innovative behavior (Fombrun, 1996; Merton, 1968). With a highly reputable name, a category-spanning organization or product is more likely to be considered as innovative rather than deviant. The Matthew effect maintains that highly reputable actors not only are shielded from penalties, but also receive more credit for doing tasks identical to those of less reputable ones (Merton, 1968). Highly reputable actors thus have more leeway in how much they should conform to the conventions of a category, and can innovate through cross-category hybridization. Nonetheless, one would expect category-spanning organizations or products with a highly reputable name to garner heightened audience attention.

Thus, in the film industry, we expect that the reputation associated with a sequel film name will be enhanced if the predecessor film in the series performed well. As a result, audience attention will increase even despite the potential "illegitimacy discount" associated with genre spanning:

Hypothesis 3 (H3): The negative effect of genre spanning on audience attention (a film's box office revenue) will be reduced when a better reputation is embedded in a film's name by clearly identifying it as part of a film series that has a more successful preceding film in the series.

Note that in all hypotheses we theorize interaction effects. In H1, we examine the interaction effect between genre spanning and familiarity (similarity with prominent, preexisting names). We argue that such naming helps gain audience attention, reducing the illegitimacy discount of genre spanning. However, we are agnostic as to whether or not the positive moderating effect of name fully offsets the illegitimacy discount caused by genre spanning. In H2, we argue that names tied to film series will increase audience attention by signaling reputation. We argue that sequel names make clearer linkages to past films and are thus more likely to attract consumer attention than those names that are similar to preexisting but unrelated films. In H3 we further argue that sequel names may have different levels of reputation depending on the predecessors' performance. Compared with the naming effects theorized in H1 and H2, names linked to a higher-performing previous film in a series are more likely to reduce the illegitimacy discount associated with genre spanning. As a result, in addition to a positive interaction effect, in H3 we also expect a more positive overall effect of genre spanning for a movie with a more reputable sequel name.

Data and Methods

Our hypotheses are grounded in our empirical setting—the US feature film industry. We chose the film industry as our empirical setting for three reasons. First, the film industry, like other creative industries such as video games and music recording, is highly volatile and unpredictable in terms of audience demands (De Vany & Walls, 1996; Faulkner & Anderson, 1987). This uncertainty is exacerbated by the intense competition and rapidly increasing investment in making a film. Delivering a film to the market is an increasingly complex, lengthy, and costly process (Baker & Faulkner, 1991). This type of uncertain environment makes it critically important for films to attract audience attention by standing out from competing films.

Second, there is a widely studied genre categorization system in the film industry. When a film is released to the market, its genre becomes a valuable tool for audiences to consider as they construct their choice set (Austin, 1989). There is also evidence for demographic variation in the audiences for different genres (Hsu, 2006). Therefore, genre is a critical medium through which producers communicate with audiences and signal the distinctiveness of their offerings.

Third, name is an important identifier of a film. This is particularly true in the name of a sequel. Producers normally employ the same name as the original film to carry over its reputation with audiences. Film sequels named in this way help audiences recognize them and link them to the main cast or storyline of the predecessor film(s). Name and genre thus constitute two critical stimuli that audiences use to recognize and evaluate a film. The film industry, therefore, represents precisely the kind of empirical setting required for testing the ideas proposed in this paper.

Data sources

We collected data covering all films released in the US market between 1982 and 2007 as listed in *Variety*. Previous studies have shown that the *Variety* listings are comprehensive in covering films released in the US market (Sorenson & Waguespack, 2006). As noted by Sorenson and Waguespack (2006), films with sales of as few as 100 tickets could qualify for inclusion on the *Variety* lists. Internet Movie Database (*IMDb*) and RottenTomatoes (*RT*) were the other two main data sources and provided the following information: film name, release date, genre, sequel, MPAA rating, run time, full cast, director, production budget, distributor, number of opening screens, critics ratings, and user ratings. Specifically, film sequels were identified through information collected from IMDb where sequel was indicated by the keyword "sequel" under the section of plot keywords of each film. This information was double-checked with the sequel data listed on Wikipedia and inconsistencies were resolved by referring to a third source.

We selected 1982 as the starting year of analysis because we used the number of opening screens as a proxy of marketing expenditure for a film. This proxy emerged when distributors experimented in the early 1980s with the "wide release" strategy in which films open on the same day across all domestic markets.

Dependent variable

The unit of analysis of this study is a film. We focus on the general public—the normal filmgoers—as an audience and examine the attention they pay to a film as reflected in its box office revenue. Our focus on box office revenue as an outcome-based indicator of attention goes beyond mere awareness or noticing to understand how the mechanisms we posit relate to the willingness to experience a product vis-a-vis competing offerings—in our case, attending a film. Ocasio (2011)

reviews disparate approaches to attention, highlighting the importance of studying what he labels "attentional selection" or the outcome of attentional processes (see also Greve, 2008; Hansen & Haas, 2001; Hoffman & Ocasio, 2001). When filmgoers purchase tickets, they have yet to see the film and thus their purchase is influenced by factors other than their real experience with it (e.g., name familiarity or reputation). Previous studies on the film industry have also examined post-consumption ratings by professional critics and filmgoers. In our context, such ratings are less relevant because critics and filmgoers offer their ratings only after they have watched the movie. Therefore, their ratings are more likely to be influenced by real consumption experience than by the naming effects discussed in this paper.

Our dependent variable is a film's *opening week box office revenue*. The naming effects on audience attention are more pronounced in the opening week than in later weeks when alternative information on movies becomes available through word-of-mouth and online reviews (Liu, 2006). Thus, our focus on opening week revenues helps us rule out these alternative influences on our dependent variable. Empirically, we first made inflation adjustments of box office revenue based on the average price of a movie ticket reported annually by the National Association of Theatre Owners (year 2000 is the base year). We then transformed it using a natural log function to adjust for its skewed distribution.

Independent variables

The first key independent variable of interest is *genre spanning*. Genre spanning was calculated by following the steps employed by Hsu et al. (2009). First, we gathered information about genres assigned to each film from two archival sources: IMDb and RT.[2] Variations in the specificity of genre assignments across the two sources were reconciled according to Hsu et al.'s (2009) criteria. For example, while RT regards "action/adventure" as one genre, IMDb considers "action" and "adventure" as distinct genres. In such hybrid cases, we reclassified the films under two distinct genres.[3] Based on this procedure, films in our dataset were classified into 16 genres: action, adventure, adult, animation, comedy, documentary, drama, family, fantasy, horror, musical, romance, science fiction, sport, thriller, and western.

Second, we measured a film's grade of membership (GoM) in a genre as the proportion of the two archival sources IMDb and RT that classified it under that genre. Then, we used Simpson's (1949) index of diversity to calculate genre spanning of each film:

$$GS_i = 1 - \Sigma_{g \in G} \, \mu_{ig}{}^2$$

where i is the focal film, G is the set of genres (16 genres in this study), and

$$GoM_{ig} = \Sigma_{s \in S} \, I_{sg} / N_S$$

$$\mu_{ig} = GoM_{ig} / (\Sigma_{k \in G} \, GoM_{ik})$$

where μ_{ig} is the weighted GoM of genre g for film i. S denotes the set of classification sources, N_S is the number of sources in the set S ($N_S = 2$ in this study), and I_{sg} is an indicator function that takes 1 if source s classifies film i as genre g, and 0 otherwise. The weighted GoM, μ_{ig}, takes values between 0 and 1, and increases as film i is classified by sources as having fewer genres and as more sources classify film i as genre g. Accordingly, genre spanning, GS_i, takes values between 0 and 1, and increases as film i is classified by sources as having more genres.

The second primary independent variable is *familiarity*, capturing non-sequel similarity to a prominent preexisting name. This variable is inspired by the studies of Glynn and Abzug (1998, 2002), and aims to tap into the familiarity of name elements. We constructed this variable

as follows. For each non-sequel film in our sample, we searched for past films that have a title "similar" to it. As there is no universal way of defining similarity between two film titles, we decided to use the following procedure which combines computer and manual coding. First, for each non-sequel film, we searched for past films that share at least a word in titles that is not generic (we excluded words such as I, you, we, us, a, the, prepositions, movie, etc.). If a focal non-sequel film does not share any non-generic word with any of the past films, we coded *familiarity* as 0 for the focal film. If it shares a word with a past film, two coders then independently examined their full titles and verified if the two titles are indeed similar. If both coders agree that the titles are similar (or dissimilar), we coded *familiarity* as 1 (or 0) for the focal film. When there were disagreements, the two coders reconciled their coding after a discussion.

If there are multiple past films whose titles are judged as similar to that of the focal film, we chose a past film whose release date is the closest to the focal film in constructing the variable that controls the potential impact of the past film's success attached to its name.[4] Finally, as robustness checks, we limit the sample of past films to those which were shown in the US after 1982 and among the top 5%, 10%, or 20% most successful in terms of the total box office revenues. The idea behind this is to capture the most prominent past films that are relevant to non-sequel copycat naming. We find that the results remain qualitatively unchanged under these alternative specifications.

The third independent variable is *reputation* as captured by films that use the same sequel name in a film series. Before coding this variable, we first generated one dummy variable *sequel*, coded as 1 for film sequel and 0 otherwise. We defined "sequel" broadly to include all related films in a series, which may include a prequel or a film with an unrelated storyline but with a common character (e.g., the James Bond series). The key for us is whether the sequel is identified with predecessors in the same series by a resonating name. Nine percent of films in our sample are sequels. *Reputation* was then coded as 1 when the film sequel is named after its predecessors and 0 otherwise. A subset of film sequels in the data (9%) were named differently from their predecessors— that is, without sharing any common phrases. One example is *Staying Alive*, which is a sequel of *Saturday Night Fever*. This portion of sequels provided us with a rare opportunity to test whether simply being a sequel, no matter whether the film takes a preexisting name or not, attracts audience attention. To rule out this rival explanation, we created the variable *different sequel name*, which was coded 1 for film sequels with completely different names from their predecessors' and 0 otherwise.

To differentiate sequel names with different levels of reputational capital, we weighted the *reputation* and *different sequel name* variables by the logged total box office revenue of the immediate predecessor and generated two new name variables: *weighted reputation* and *weighted different sequel name*.

Control variables

Many other factors may influence the opening week box office revenue of a film. Accordingly, we controlled for a detailed baseline model that included a broad range of attributes at the film and distributor level. We also included time variables to account for the potential effects of seasonality and macro environmental changes.

Film-level controls. We controlled for the rating of each film by the Motion Picture Association of America (MPAA)—*MPAA rating*—whose aim is to aid audiences in determining a film's content and suitability. Although the MPAA rating system is voluntary, most mainstream US film producers

submit their films for rating before releasing them on the market to avoid potential negative impact on film box office without such a rating. MPAA uses five rating symbols: G (general audiences), PG (parental guidance suggested), PG-13 (parents strongly cautioned), R (restricted), and NC-17 (no children 17 and under admitted). The MPAA's decision on film rating may impact on the box office performance because it may restrict the potential market size as well as the opportunity for advertising. We coded these five rating symbols as dummy variables, and set R as the base case.

Previous literature has suggested that production budget plays a role in explaining variation in box office performance (Sorenson & Waguespack, 2006). Thus, we included the *production budget* as a control variable in the estimations. Marketing efforts of each film may also affect audiences' choices and thus film box office revenue. Since most films spend a short period in the theaters, audiences may have limited time to learn about the availability and quality of new film offerings (Sorenson & Waguespack, 2006). Therefore, strong marketing efforts are used to alert audiences to the presence of a new film in the theater and help match potential viewers with films of their tastes. We used the number of *opening screens* on which a film played in its first week of exhibition as a proxy for the marketing efforts that distributors allocated to promoting a film. Given that systematic information of the exact dollars devoted to marketing of a film is not available, previous studies have found that the number of *opening screens* is a valid and reliable proxy for marketing expenditures (Sorenson & Waguespack, 2006). We took natural log of both variables to adjust for their skewed distribution.

Existing studies that used film box office as the dependent variable normally included star power and director power as covariates, but have found mixed results on the extent to which stars and famous directors drove box office success (Albert, 1998; Hsu, 2006). We generated two variables—*star power* and *director power*—to control for the potential star and director effects on film box office. We defined the lead actors and actresses of each film as stars.[5] We calculated *star power* at the time of a specific film as the average box office revenue of films in which the actor/actress was a starring cast member in the previous three years (Moul, 2007). We took the natural log of this variable and, in the case of multiple-star movies, we calculated the average of the *star power* score for each star before taking the natural log. We calculated *director power* as the cumulative box office revenue of films that the director directed in the previous three years (Moul, 2007). Similarly we took the natural log of this variable to adjust for its skewed distribution.

Films are often not completely new products, but based on *preexisting creations* such as novels, comics, TV series, etc. Such films may receive a high degree of attention from audiences who appreciate the preexisting creations. To control for this, we generated a dummy variable that equals 1 if a film is based on preexisting creations and 0 otherwise. IMDb provides information on whether a film is based on a novel, comic book, TV show, play, etc., under the section of plot keywords.

We also included a film's grade of membership in each of the 16 genres to control for genre-level fixed effects, such as genre popularity and potential audience size (Hsu et al., 2009). To capture the competitive pressure faced by each film, we also generated the *genre crowding* variable to capture the degree to which the genres of a film overlap with those of all other films exhibited during the period of a focal film's exhibition. This variable was calculated by summing up the genre overlap of a film with all others in that period, where genre overlap was operationalized as the fraction of the genres of the focal film that are also assigned to the alter (Hsu et al., 2009).

One important factor that may affect the box office performance of a film is its perceived quality. While there is no perfect measure of how audiences perceived the film quality at the time of a movie-going decision, we attempted to control for it by including the *average user rating* of a film obtained from RT (10-point scale). RT computes this variable based on all user reviews posted on

their website over time. Our measure was collected long enough after a film dropped out of theaters, and thus it captured the overall level of user satisfaction with a film.[6]

The two films sharing the same name elements may also have different genres. There may be concerns that such genre differences may affect audiences' perception of the focal film and drive its box office performance. To control for this, we created two variables to capture the genre differences: *number of genres changed* and *percentage change in genres*. The first variable was calculated as the sum of the number of genres added and dropped from the predecessor. The second variable was computed by dividing the first variable by the total number of distinct genres of a sequel and its predecessor. As an example, for a sequel (action, adventure, and science fiction) and its predecessor (science fiction and romance), *number of genres changed* is 3 (action, adventure, and romance), and *percentage change in genres* is 0.75 (3/4). We only included number of genres changed in our estimation models because the two variables are highly correlated at 0.95. We calculated this variable for both sequels and non-sequel films that have a name similar to a past film and included both as controls.

Finally, we included the *logged total box office revenue of the predecessor* to control for the impact of a predecessor film's financial success on the focal film's performance. We interacted this variable with the two naming dummies—*reputation* and *different sequel name*—to allow for a different impact of predecessor's financial success on film sequels with different naming strategies. Operationally, this is equivalent to the inclusion of *weighted reputation* and *weighted different sequel name* variables in our estimation equations. We also computed a similar variable for non-sequel films that have a name similar to a past film using the past film's logged total box office revenue, and named this variable as *weighted familiarity*.[7]

Distributor-level controls. We included distributor fixed effects to control for potential distributor heterogeneity which may affect film box office. Based on the distribution of total box revenue during the observation period, we classified film distributors into major and non-major distributors by the dummy variable *major distributor*. Distributors classified as a major distributor included Buena Vista Pictures Distribution, Paramount Pictures Corp., Sony Pictures Entertainment, Inc., Twentieth Century Fox Film Corp., Universal City Studios LLLP, and Warner Brothers. This list is consistent with the "big six" major US film distributors classified by MPAA.

Time controls. To control for the potential effect of changes in the macro environment of the film industry on film box office, we included the *release year* of each film in the estimation. With regard to the potential within-year seasonality in film box office, Vogel (2007) argued that there are two low seasons for movie-going: pre-summer low (May) and autumn plateau (September, October, and November). Thus we included in the estimation a dummy variable, *low season*, and coded it as 1 if a film was released in May, September, October, or November, and 0 otherwise.

Table 1 displays the descriptive statistics on key variables used to test the hypotheses. Correlations among the control and independent variables are reasonably small in magnitude except for interaction terms which are high by construction. To assess the risk of multicollinearity problems, we computed the variance inflation factors (VIFs) and confirmed that no severe multicollinearity exists (all VIFs were less than 10 in all estimation models).

Estimation strategies

Since our dependent variable—the natural log of the opening week box office revenue—is left-censored, previous literature has employed Tobit regression in the analysis (Hsu et al., 2009).

However, one complication in our estimation is that all hypotheses are tested through the interaction effects, which are difficult to estimate and interpret in nonlinear models such as Tobit (Norton, Wang, & Ai, 2004). Therefore, we used OLS estimation with robust standard errors instead. To assess the severity of data censoring, we compared the estimation results between OLS and Tobit estimations of the baseline model. The results turned out to be quite similar, indicating that OLS estimation will generate results that are comparable to Tobit regression. Moreover, the Tobit regression revealed that there were actually no censored observations in our dataset, further justifying our use of OLS estimation. Therefore, we report results based on the set of OLS models.

Results

Model 1 of Table 2 is our baseline model. Notably, results in model 1 replicate the negative impact of *genre spanning* on film performance as suggested in the past literature. Production investment (*production budget*) and marketing efforts (*opening screens*) significantly increase the opening week box office revenue. *Star power* and *director power* appear to be positively and significantly related to the opening week box office revenue. We also found that opening week box office revenue is higher for films (1) *based on preexisting creations*, (2) with higher *average user rating*, (3) distributed by a *major distributor*, and (4) with higher *number of genres changed (for sequel)*, but lower for films located in more crowded genres.

Model 2 adds the effects of *weighted familiarity, weighted reputation*, and *weighted different sequel name* variables to control for past films' financial success on the focal film's opening week performance. We found both *weighted reputation* and *weighted different sequel name* have a positive and significant impact, suggesting that reputation built by predecessors in a film series continue to benefit the market performance of film sequels. However, the good performance of a past film that shares a similar name element with a focal film (familiarity) does not contribute to the focal film's opening week box office revenue. Model 3 adds the interaction term: *genre spanning × familiarity*. The results show that the interaction term is not significant, indicating that pure similarity to the name of a previous film does not significantly mitigate the negative effect of genre spanning. Our supplementary analyses using three different cut-offs of previous films' success levels (top 5%, 10%, and 20%) further suggested that simply copycatting even the most prominent past films is not sufficient to overcome the illegitimacy discount associated with genre spanning. Therefore, we find no support for H1. Model 4 tests two interaction terms: *genre spanning × reputation* and *genre spanning × different sequel name*. We found that the coefficient of *genre spanning × reputation* is positive and marginally significant, providing partial support for H2. By contrast, the coefficient of *genre spanning × different sequel name* is positive but not significant, indicating that, without a same name, simply being a sequel does not mitigate the illegitimacy discount due to genre spanning.

To quantify the reduction in the illegitimacy discount of genre spanning due to reputation effects, we computed the marginal effects of genre spanning on the opening week box office revenue for non-sequels vs. sequels with same sequel name (i.e., those signaling reputation). We fixed all controls at their mean value and examined the change in the opening week box office revenue when genre spanning increases from 0 to 0.5.[8] We found that the marginal impact is −$475,855.81 for non-sequels and −$125,291.66 for sequels with a reputable name. Thus, reputable sequel names reduce the illegitimacy discount of genre spanning by $350,564.15.

Model 5 replaces the two interaction terms in Model 4 with another two interaction terms: *genre spanning × weighted reputation* and *genre spanning × weighted different sequel name*. The results show that the coefficient of *genre spanning × weighted reputation* is positive and significant. This

Table 1. Descriptive Statistics and Correlation Matrix of Key Variables.

Variable	N	Mean	S.D.	1	2	3	4	5	6	7	8	9	10
1. Opening week box office revenue (logged)	2827	14.49	2.74										
2. Genre spanning	2827	0.49	0.25	.35									
3. Familiarity	2827	0.03	0.17	.05	.04								
4. Reputation	2827	0.08	0.27	.24	.15	−.05							
5. Different sequel name	2827	0.01	0.09	.02	.04	−.02	−.03						
6. Weighted familiarity	2827	0.51	2.80	.05	.04	.99	−.05	−.02					
7. Weighted reputation	2827	1.31	4.06	.24	.15	−.05	.99	−.06	−.05				
8. Weighted different sequel name	2827	0.12	1.32	.03	.04	−.01	−.06	.99	−.01	−.09			
9. Genre spanning × Familiarity	2827	0.00	0.04	−.02	−.03	.22	−.04	−.01	.23	−.04	−.01		
10. Genre spanning × Reputation	2827	0.01	0.06	.04	−.06	−.04	.51	−.03	−.04	.51	−.05	−.05	
11. Genre spanning × Different sequel name	2827	0.00	0.02	.01	−.03	−.01	−.03	.59	−.01	−.05	.63	−.01	−.03
12. Genre spanning × Weighted reputation	2827	0.15	0.97	.16	.29	−.03	.54	−.03	−.03	.56	−.04	−.06	.93
13. Genre spanning × Weighted different sequel name	2827	0.01	0.28	.05	.08	−.01	−.03	.63	−.01	−.05	.68	−.02	−.07
14. Opening screens (logged)	2827	5.57	2.70	.92	.36	.06	.21	.01	.06	.21	.02	−.02	.00
15. Production budget (logged)	2827	16.38	1.65	.71	.29	.04	.14	.01	.04	.15	.01	−.01	.03
16. Star power (logged)	2827	11.96	7.64	.39	.12	.02	.03	−.01	.02	.03	−.01	.01	−.02
17. Director power (logged)	2827	5.96	8.04	.25	.07	−.00	.03	.04	−.00	.03	.04	.00	.03
18. Based on preexisting creations	2827	0.36	0.48	.11	.04	−.01	−.04	−.02	−.01	−.04	−.01	.02	.01
19. Genre crowding	2827	26.86	13.98	−.41	−.60	−.03	−.22	−.03	−.03	−.22	−.02	.03	−.01
20. Average user rating	2827	5.58	1.93	.05	−.02	−.02	−.05	.05	−.02	−.05	.05	.03	.01
21. No. of genres changed (non-sequel)	2827	0.10	0.65	.04	.04	.89	−.05	−.01	.87	−.05	−.01	.25	−.04
22. No. of genres changed (sequel)	2827	0.08	0.41	.14	.13	−.04	.57	.34	−.04	.55	.31	−.03	.39
23. Major distributor	2827	0.67	0.47	.51	.17	.03	.10	−.00	.03	.11	.00	.03	.01

Variable	11	12	13	14	15	16	17	18	19	20	21	22
1. Opening week box office revenue (logged)												
2. Genre spanning												
3. Familiarity												
4. Reputation												
5. Different sequel name												
6. Weighted familiarity												
7. Weighted reputation												
8. Weighted different sequel name												
9. Genre spanning × Familiarity												
10. Genre spanning × Reputation												
11. Genre spanning × Different sequel name												
12. Genre spanning × Weighted reputation	−.06											
13. Genre spanning × Weighted different sequel name	.98	−.06										
14. Opening screens (logged)	.00	.13	.04									
15. Production budget (logged)	−.02	.13	.02	.65								
16. Star power (logged)	−.03	.03	−.02	.34	.51							
17. Director power (logged)	.01	.05	.02	.20	.32	.26						
18. Based on preexisting creations	−.01	.03	−.00	.05	.22	.14	.09					
19. Genre crowding	.01	−.21	−.06	−.39	−.29	−.06	−.10	−.03				
20. Average user rating	.03	.01	.02	−.10	.08	.06	.07	.10	.08			
21. No. of genres changed (non-sequel)	−.01	−.02	−.01	.05	.03	.02	−.01	−.01	−.03	−.01		
22. No. of genres changed (sequel)	.21	.41	.21	.13	.07	−.01	−.00	−.02	−.15	−.01	−.03	
23. Major distributor	.01	.07	.03	.46	.51	.30	.16	.11	−.15	.01	.01	.05

Note: Interaction terms are calculated based on mean-centered variables.

Table 2. Effects of Genre Spanning and Naming on the Opening Week Box Office Revenue.

Variable	Model 1 OLS	Model 2 OLS	Model 3 OLS	Model 4 OLS	Model 5 OLS	Model 6 2SLS
Genre spanning	−0.45**	−0.43**	−0.43**	−0.41**	−0.45**	−0.45**
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Genre spanning × Familiarity			0.05			
			(0.30)			
Genre spanning × Reputation				0.32+		
				(0.23)		
Genre spanning × Different sequel name				0.71		
				(0.67)		
Genre spanning × Weighted reputation					0.03*	0.03*
					(0.02)	(0.02)
Genre spanning × Weighted different sequel name					0.04	0.04
					(0.05)	(0.05)
Weighted familiarity		−0.00	−0.00	−0.00	−0.00	−0.00
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Weighted reputation		0.04**	0.04**	0.04**	0.04**	0.04**
		(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Weighted different sequel name		0.03**	0.03**	0.02*	0.02*	0.02
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Opening screens (logged)	0.81**	0.81**	0.81**	0.81**	0.81**	0.81**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Production budget (logged)	0.18**	0.18**	0.18**	0.18**	0.18**	0.18**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Star power (logged)	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Director power (logged)	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Based on preexisting creations	0.12**	0.13**	0.13**	0.13**	0.13**	0.13**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Genre crowding	−0.01**	−0.01**	−0.01**	−0.01**	−0.01**	−0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Average user rating	0.16**	0.16**	0.16**	0.16**	0.16**	0.16**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
No. of genres changed (non-sequel)	0.02	0.04	0.04	0.04	0.04	0.04
	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
No. of genres changed (sequel)	0.11**	−0.07+	−0.07+	−0.08*	−0.08*	−0.08*
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Major distributor	0.35**	0.35**	0.35**	0.35**	0.35**	0.35**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Constant	55.84**	55.17**	55.16**	54.95**	54.92**	54.44**
	(7.35)	(7.28)	(7.28)	(7.29)	(7.29)	(7.36)
Adjusted R^2	0.896	0.898	0.898	0.898	0.898	0.898
No. of observations	2827	2827	2827	2827	2827	2821

Notes:
1. + p < .10; * p < .05; ** p < .01.
2. Standard errors are in parentheses.
3. One-tailed test for independent variables and two-tailed test for control variables.
4. MPAA Rating, Year Trend and Seasonality, and Genre GoMs, were included in all models as controls, but not reported here.

positive interaction effect provides strong support of H3. Based on model 5, we compared the total impact of genre spanning by three different levels (maximum, mean, and minimum) of a predecessor film's performance. The comparison shows that the total impact of genre spanning under maximum predecessor's success is significantly more positive than the total impact of genre spanning under mean predecessor's success ($p = 0.04$) which in turn is significantly more positive than that under minimum predecessor's success ($p = 0.04$). In fact, any increase in the reputation of a sequel name will significantly reduce the illegitimacy discount of genre spanning.

We again computed the reduction in the illegitimacy discount of genre spanning by comparing non-sequels to sequels with reputation (sequels sharing predecessor naming schemes) at the minimum, mean, and maximum levels of predecessor's success. We found that the illegitimacy discount of genre spanning for non-sequels, as genre spanning increases from 0 to 0.5, is $533,973.27. The illegitimacy discount of genre spanning for reputable sequels, again as genre spanning increases from 0 to 0.5, at the minimum level of predecessor's success is $218,470.47. In contrast, at the mean and maximum levels of predecessor's success, the same increase in genre spanning actually increases the opening week box office revenue by $9,838.59 and $62,431.51, respectively. Overall, our results suggest that the illegitimacy discount associated with genre spanning can be overcome if sequels are named after more successful predecessor films. We conducted a series of supplementary analyses to check the robustness of these findings.

Supplementary Analyses

Endogeneity

Previous studies have suggested that *opening screens* is an endogenous variable in estimating film box office (Sorenson & Waguespack, 2006). Consistent with these studies, we obtained the total number of *available screens* and used it to instrument the number of opening screens for each film. While the number of opening screens of each film likely depends on the total screens available, we do not expect any direct impact of total available screens on the box office revenue of a particular film (Sorenson & Waguespack, 2006).

In addition, we introduced a new set of instruments for opening screens. It is a longstanding rule of thumb in theatrical exhibition that a week-to-week decline of 40 percent or more in box office revenue from a constant pool of theaters indicates that consumer interest is evaporating (Marich, 2005). This suggests that films which experienced a decline of 40 percent or higher in weekly box office revenue are likely to be dropped from theaters. If they are actually dropped from theaters, the screens that they previously occupied become available for opening films or for expansion among existing films. Thus, the number of screens occupied by those films is expected to be positively correlated with the number of opening screens for new films, but is unlikely to be correlated with the market performance of opening films. Based on this argument, we created two instruments for each week in which we observed opening films, and used them as instruments for the number of opening screens for those films. The first instrument is *potential screen drop*, which was computed as the weighted sum of the screens for films which experienced a decline of 40 percent or higher in per-screen box office revenue. Weights were based on the box office revenue of each film in that week. The second instrument is *actual screen drop*, which captures the potential screen drops that were actually dropped from theaters. Statistically, potential screen drop is always larger than or equal to actual screen drop.

The endogeneity problem also applies to the variable *production budget*. Following Sorenson and Waguespack (2006), we included *run time* of a film as an instrumental variable for production

budget. We expect run time to increase the production costs and hence the budget, but with no independent effect on film box office revenue. Then we re-estimated model 5 of Table 2 with the instruments for opening screens and production budget using a two-stage least squares (2SLS) approach (Greene, 2003). The results reported in model 6 again confirmed H3.

Alternative mechanisms

We examined several other potential signals to see whether they can also help gain audience attention and recognition. These other indicators were *star power, director power, major distributor, preexisting creations*, and *the average user rating*, originally included as controls in our main models. We took model 5 of Table 2 as a basis and tested whether the addition of an interaction term between genre spanning and each of these indicators would reduce the significance of the interaction between genre spanning and weighted preexisting name. Results are reported in Table 3. We found that the interaction between genre spanning and weighted reputation remains significantly positive after including each of the additional interaction terms, which suggests that sequel name serves as a more direct and powerful identity marker than these other signals. We speculate that the differential salience between reputable sequel names and these other signals may be due to the particularly strong links built among audiences between a certain name and the specific theme of a film series it signals (Keller, 1993; Keller, Heckler, & Houston, 1998).

Genre dynamics

We recognize that our genre spanning measure based on the 16 distinct genres, although consistent with previous studies, does not account for the popularity of certain combinations of genres over time. Given that genres are socially constructed and "partially constituted by the audiences that support them" (DiMaggio, 1987, p. 441), they are inherently fluid and subject to alteration over time. Indeed, there is a burgeoning line of research studying the emergence and construction of categories such as genre (e.g., Jones et al., 2012; Kennedy, 2008; Khaire & Wadhwani, 2010; Lounsbury & Rao, 2004). To account for the dynamic nature of genre combinations, we created a new genre spanning variable which captures the changing saliency of certain combinations of genres. We then re-estimated models 1–5 of Table 2 with the new genre spanning measure. We found that our main results for our hypotheses remain unchanged. The calculation of this new variable is detailed in the Appendix and estimation results are reported in Table A1.

Selection bias

Our sample of films fall into four non-overlapping categories: sequels with similar names to predecessor films (our reputation variable), sequels with names that differ from predecessor films, non-sequel films that have names using terms from previously successful films (our familiarity variable), and other non-sequel films. We recognize that films are not randomly assigned to each category, which leads to a potential selection bias—the choice of producing a sequel and the choice of different naming strategies could be affected by unobserved factors that potentially influence our outcome variable (Heckman, 1979). To correct these potential selection biases, we used a two-step approach proposed by Lee (1983). Lee's approach extends Heckman (1979) in that it allows for multi-category selection and models the selection decision with multinomial logit estimation techniques. We computed a bias correction term for each category in the first stage, and included these bias correction terms as additional regressors in our main regressions.

Table 3. Interaction Effects between Genre Spanning and Alternative Signals.

Variable	Model 7 OLS	Model 8 OLS	Model 9 OLS	Model 10 OLS	Model 11 OLS
Genre spanning	−0.43**	−0.45**	−0.43**	−0.44**	−0.45**
	(0.16)	(0.16)	(0.17)	(0.16)	(0.16)
Genre spanning × Weighted reputation	0.03*	0.03*	0.03*	0.03*	0.03*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Genre spanning × Weighted different sequel name	0.04	0.04	0.04	0.05	0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Genre spanning × Star power (logged)	0.01				
	(0.01)				
Genre spanning × Director power (logged)		0.004			
		(0.01)			
Genre spanning × Major distributor			0.11		
			(0.15)		
Genre spanning × Based on preexisting creations				0.13	
				(0.13)	
Genre spanning × Average user rating					0.02
					(0.03)
Adjusted R^2	0.898	0.898	0.898	0.898	0.898
No. of observations	2827	2827	2827	2827	2827

Notes:
1. $+ p < .10$; $* p < .05$; $** p < .01$.
2. Standard errors are in parentheses.
3. One-tailed test for independent variables and two-tailed test for control variables.
4. All control variables used in Table 2 were included in corresponding models, but not reported here.

To obtain correct standard errors, we used a bootstrap method with 3,000 replications of the sample with the size equal to our observed sample (i.e., $N = 2,827$). We reported the results corrected for self-selection in Table 4. We found that our main results for hypothesized variables remain unchanged.

Temporality of the naming effects

We also investigated the temporality of the naming effects. Past research has suggested that sustained attention is limited in duration and the probability to detect a certain stimuli decreases over time (Ocasio, 2011; Swets & Kristofferson, 1970). To test whether the naming effects on audience attention decay over time, we constructed a variable *temporality* as the logged number of years between a focal film's release and its predecessor's release. For each variable that involves naming dummies, we created an interaction term between the variable and *temporality*. We included those additional interaction terms in re-estimating models 2–5 of Table 2. We found no support for temporality of the naming effects.

Additional controls

Finally, we tested for additional controls such as the *order of a sequel in a film series* and *common director* and *common star* dummies that indicate whether the focal film and similarly named

Table 4. Correction for Self-Selection.

Variable	Model 12 Lee & Bootstrap	Model 13 Lee & Bootstrap	Model 14 Lee & Bootstrap	Model 15 Lee & Bootstrap	Model 16 Lee & Bootstrap
Genre spanning	−0.45** (0.16)	−0.43** (0.16)	−0.43** (0.16)	−0.41** (0.17)	−0.45** (0.17)
Genre spanning × Familiarity			0.03 (0.33)		
Genre spanning × Reputation				0.33+ (0.23)	
Genre Spanning × Different sequel name				0.56 (1.16)	
Genre spanning × Weighted reputation					0.03* (0.02)
Genre spanning × Weighted different sequel name					0.03 (0.08)
Correction for self-selection (uniquely named non-sequel)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Correction for self-selection (familiarity)	−0.00 (0.00)	−0.00 (0.01)	−0.00 (0.01)	−0.00 (0.01)	−0.00 (0.00)
Correction for self-selection (reputation)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Correction for self-selection (different sequel name)	0.01 (0.02)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)
Adjusted R^2	0.897	0.898	0.898	0.898	0.898
No. of observations	2827	2827	2827	2827	2827

Notes:
1. + $p < .10$; * $p < .05$; ** $p < .01$.
2. Standard errors are in parentheses.
3. One-tailed test for independent variables and two-tailed test for control variables.
4. All control variables used in Table 2 were included in corresponding models, but not reported here.
5. Standard errors were obtained by bootstrapping with 3,000 replications of the sample ($N = 2,827$).

preexisting films (in the case of sequels, whether a sequel and its predecessors) share the same director or star. We found none of these additional controls significant.

Discussion

Our study of the US film industry suggests that genre and name are two critical stimuli that shape audience attention to movies. Without considering naming effects, past studies have demonstrated

that genre spanning led to significant market penalties because of audience inattention and confusion. Our analysis of films released in the US market between 1982 and 2007 again corroborated this proposition. For those that strategically deploy name to gain audience attention, however, the illegitimacy discount of genre spanning was significantly reduced. Our results suggest that different naming strategies, a form of cultural entrepreneurship (Lounsbury & Glynn, 2001), have varied potency in gaining the attention of audiences and offsetting the illegitimacy discount. We find that simply sharing some common name elements with successful past film titles does not necessarily help gain audience attention. To benefit from the naming effect on attention under conditions of genre spanning, it is far more powerful to be credibly connected to previous films with known reputations. Our findings show that sequel naming, especially sequels that are part of a highly successful series, is potent enough to mitigate against the illegitimacy discount and increase audiences' attention to genre-spanning films.

Our study makes several theoretical and empirical contributions. First, it significantly extends the literature on categorization. Recently, scholars studying categories have shifted their attention to relaxing the assumption of categories as taken-for-granted social facts and contextualizing the conditions under which the categorical imperative obtains (see Hannan, 2010; Negro, Koçak, & Hsu, 2010, for reviews). However, scholarly attention has been mainly focused on the changing nature of categories and their evolving boundaries. For example, Ruef and Patterson (2009) found that organizations spanning multiple categories were less problematic when categorization systems themselves were emergent or in flux. Similarly, Rosa et al. (1999) argued that, when categories were emergent and unstable, atypical products were more acceptable. Hsu (2006) suggested that, in markets where economies of scale were present for generalist organizations and where complementarities between different taste positions were high, category spanning was unlikely to have significant negative effects on audience appeal. In a study of the de-institutionalization of categorical boundaries in French gastronomy, Rao et al. (2005) found that category spanning was less penalized where boundaries eroded due to increasing borrowing across boundaries triggered by high-status actors. Taking a relational approach to categories, Wry and Lounsbury (2013) suggested that venture capitalists were less likely to discount carbon nanotube start-ups that spanned vertically and horizontally related patent categories.

While these studies help contextualize the illegitimacy discount argument, they fall short in explaining whether and how particular organizations and products that span highly institutionalized categories can still gain attention from audiences. We address this puzzle by suggesting that categories are only one of several cultural stimuli that shape the attention of audiences. Our findings highlight how cultural resources such as naming can be strategically deployed to mitigate the potential of audience inattention due to category spanning. One natural extension of our study is to consider how other forms of cultural entrepreneurship involving the use of vocabularies, logics, codes, and discourse (Jones et al., 2012; Loewenstein et al., 2012; Phillips et al., 2004; Thornton et al., 2012; Weber, Heinz, & DeSoucey, 2008) may be drawn upon to help category-spanning innovations to be perceived as legitimate or of higher value.

Future research should also examine the possibility that the salience of different sources of audience attention may wax and wane during different time periods. Little is known about why certain stimuli gain potency in audiences' minds while others are subject to dissipation, rejection, or replacement. Future studies should identify the organizational and environmental factors that determine the relevance of various stimuli for audiences' perception and evaluation of an organization or product. Special attention should be paid to those factors that render certain stimuli vulnerable to erosion or rejection over time. The fact that certain stimuli lose potency as evaluative lenses may have significant economic and social implications.

Also, while our focus in this paper has been on category spanning, it would be useful for future research to explore the relationship between category spanning and the dynamics of category boundaries. Very few studies have provided the level of granularity needed to assess how the boundaries of categories shift over time (Jones et al., 2012; Rao et al., 2003, 2005). For example, in their detailed exemplary study of the "modern architecture" category, Jones et al. (2012) document how category boundaries can expand as different actors, drawing on varied institutional logics, enter a category and generate conflicting exemplars. It would be interesting to study whether such expansion (or contraction) of category boundaries constrains or enables category spanning, and leads to higher or lower penalties related to spanning.

More broadly, our study points to an important yet overlooked approach in addressing the dilemma of sameness versus difference faced by organizations (Lounsbury & Glynn, 2001). Organizations embrace novelty and distinctiveness (Lumpkin & Dess, 1996; Rindova, Barry, & Ketchen, 2009; Rosenkopf & McGrath, 2011) and, at the same time, face pressure to conform to institutionalized conventions so as to acquire legitimacy (Meyer & Rowan, 1977; Zuckerman, 1999). Conventional wisdom suggests that organizations deal with the tradeoff between sameness and difference by adopting a moderate level of novelty and pursuing strategic balance—where they are "as different as legitimately possible" (Deephouse, 1999, p. 147). While intuitively appealing, there are important limitations to studies that advocate for strategic balance because they tend to assume away the potential difficulty and incoherence involved in staking out balanced positions. Another approach to solve the dilemma is through "the dual projects of classification and enumeration" (Pedersen & Dobbin, 1997, p. 432). According to this view, conformity and legitimacy occur at a more macro-level of social classification where organizations make clear membership claims, while novelty and uniqueness arise within the broader classifications (Glynn, 2008; Pedersen & Dobbin, 2006). Thus, the dilemma may be alternatively addressed hierarchically at different and nested levels.

Our study goes beyond these two solutions and suggests that perceptions of sameness and difference can be triggered by different kinds of stimuli. Organizations do not necessarily have to compromise and find a middle ground on a specific strategy or structure, as suggested by a balancing strategy. Such a strategy may be either difficult to maintain or turn out to be ineffective due to incoherence. Our findings also depart from the hierarchical solution by suggesting that sameness and difference do not have to be limited to category and intra-category variability. Rather, distinct kinds of attention-generating stimuli, such as genre and name in our particular setting, can jointly contribute to how sameness and difference are perceived. In fact, our study further suggests that the degrees of sameness and difference are not necessarily best understood as a continuum, but as separate dimensions that might be evaluated with different criteria. That is, instead of assuming that similarity and difference should be balanced, we believe that more research attention should be paid to specifying the different criteria by which audiences assess similarity and distinctiveness. Of course, such criteria may also vary within and across different audiences. Our study thus points to the multiplicity of identity, where multiple aspects of identity interactively shape the perception and evaluation of an organization or product by different audiences (Hsu & Hannan, 2005).

Methodologically, we have created a new way of operationalizing genre spanning which accounts for the historical evolution of the genre system. Over time, certain genre combinations may become more and more popular among audiences, and become more accepted. Previous studies on genre spanning have largely failed to take this into account. Our supplementary analysis showed that, even after accounting for the changing popularity of certain genre combinations,

certain names enhance audience attention as reflected in box office revenue. In addition, we have carefully included a comprehensive list of controls and addressed a series of empirical issues such as endogeneity, selection bias, and temporality of the naming effects, which significantly increased the validity of our findings.

Still, our study has several limitations worth mentioning. First, we focused on a specific audience: the general public. Given that an organization or product is subject to evaluations of different types of audiences, this focus may appear limiting. Yet, in our context this narrow focus is necessary because, as we argued, attention triggered by factors related to genre and name is more relevant to this specific audience than to others (e.g., professional critics). Nonetheless, future research may simultaneously examine the evaluation of an organization or product by multiple audiences.

Historically, different segments of audiences may have preferred evaluations through different dimensions. For each audience segment, separating the relevant criteria for evaluation from the irrelevant is a social act. Audiences in each segment may have been specifically socialized to ignore certain dimensions as part of the process of adopting the distinctive evaluative dimensions of their segment (Zerubavel, 1997). Unraveling this socialization process will help us understand why different audience segments adopt different evaluative lenses. In addition, divergence in preferences may have resulted in political contests among different segments of audiences. Studying such political contests may generate insights regarding the emergence and construction of prevailing evaluative schemes. It may also reveal the role political disputes play in shaping the development trajectory of the current evaluation system. Overall, the conceptualization of the evaluation system as a political arena and as a result of socialization will challenge the prevailing view of such systems as being stable and taken for granted.

In addition, although the economic and social significance of films is drawing more attention (Jones & Thornton, 2005), the film industry is only one empirical setting. Some of the findings in this paper may be driven by its unique characteristics. For example, the naming strategies identified in this paper may not be universally applicable, although we observe similar strategies in industries as diverse as video games, automobiles, hotel chains, and cell phones. Moreover, we believe that the findings of this study will be more likely to hold in industries with established genre systems such as entertainment industries like video games and music recording. Future studies should attempt to generalize our findings to other industries. In addition, the temporary and single-project nature of film production and the tendency of audiences to evaluate the characteristics of films rather than that of producers and distributors require us to focus our analysis at the level of film projects. Future research should explore a wider variety of kinds of organizations and products.

Notes

1. We thank editor Candace Jones for this example and for highlighting the importance of copycat naming as an attention-enhancing strategy.
2. Hsu et al. (2009) used one additional website for gathering genre information: Showbizdata.com (SBD). We excluded this source because it only contains data for films released after 1990. We also tried substituting *Variety* for SBD as the third source of genre information. The results remained largely stable no matter whether we used two or three sources in calculating the genre spanning variable.
3. We accounted for the fact that certain genre combinations become popular over time and thus perceived as one broad genre among audiences by creating a new genre spanning variable. We discussed this new variable in the section of supplementary analyses. The calculation of this new variable and the estimation results based on it are reported in the Appendix.
4. We also tried using the past film with the earliest release date. Results remain largely unchanged.
5. For each film, the actors in the cast section of IMDb are listed in credits order, with leading actors and actresses (typically stars) at the top of the order.
6. Alternatively, we could control for the average opening week critics' rating. Unfortunately, this variable is available only for films released after 2000, when RT started publishing critics' reviews with dates posted. Moreover, most films do not receive any critics' reviews in the opening week. Since including this variable in our model would reduce the sample size significantly, we decided not to use this variable as a control. However, we found that among those films with the opening week critics' rating, the correlation between this variable and the average user rating is 0.68, indicating that the average user rating is a good proxy for film quality perceived by both the general public and professional critics. We used the average user rating rather than the opening week user rating because, for most films in our data, there are no user ratings offered in the opening week. By including the average user rating, our intention is to control for the unobserved film quality rather than the word-of-mouth effect. The fact that little information is available on the opening week user rating suggests that the influence of word-of-mouth is minimal in the opening week of a film.
7. We also considered including the evaluation (average user rating) of the predecessor film as a control. But it is highly correlated with the predecessor's box office revenue (0.91). Thus, we dropped it from our reported models.
8. This is equivalent to a change in number of genres from 1 to 2 when both sources (IMDb and RT) agree with the listing of genres.

References

Albert, S. (1998). Movie stars and the distribution of financially successful films in the motion picture industry. *Journal of Cultural Economics*, *22*, 249–270.

Austin, B. A. (1989). *Immediate seating: A look at movie audiences*. Belmont, CA: Wadsworth Pub. Co.

Baker, W. E., & Faulkner, R. R. (1991). Role as resource in the Hollywood film industry. *American Journal of Sociology*, *97*, 279–309.

Baron, J. N. (2004). Employing identities in organizational ecology. *Industrial and Corporate Change*, *13*, 3–32.

Benjamin, B. A., & Podolny, J. M. (1999). Status, quality, and social order in the California wine industry. *Administrative Science Quarterly*, *44*, 563–589.

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

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the American economic review. *American Economic Review*, *81*, 1041–1067.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17*, 475–482.

Cole, S., Cole, J. R., & Simon, G. A. (1981). Change and consensus in peer review. *Science*, *214*, 881–886.

Crane, D. (1965). Scientists at major and minor universities: A study of productivity and recognition. *Americal Sociological Review*, *30*, 699–741.

De Vany, A., & WallsW. D. (1996). Bose-Einstein dynamics and adaptive contracting in the motion picture industry. *Economic Journal*, *106*(439), 1493–1514.

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*, 147–166.

Deephouse, D. L. (2000). Media reputation as a strategic resource: An integration of mass communication and resource-based theories. *Journal of Management*, *26*, 1091–1112.

DiMaggio, P. J. (1987). Classification in art. *American Sociological Review*, *52*, 440–455.

DiMaggio, P. J. (1997). Culture and cognition. *Annual Review of Sociology*, *23*, 263–287.

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

Douglas, M. (1986). *How institutions think*. Syracuse, New York: Syracuse University Press.

Faulkner, R. R., & Anderson, A. B. (1987). Short-term projects and emergent careers: Evidence from Hollywood. *American Journal of Sociology*, *92*, 879–909.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*, 2nd edition. New York: McGraw Hill.

Fombrun, C. (1996). *Reputation: Realizing value from the corporate image*. Boston: Harvard Business School Press.

Fombrun, C, & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of Management Journal*, *33*, 233–258.

Gioia, D. A., & Thomas, J. B. (1996). Identity, image, and issue interpretation: Sensemaking during strategic change in academia. *Administrative Science Quarterly*, *41*, 370–403.

Glynn, M. A. (2008). Beyond constraint: How institutions enable identities. In R. Greenwood, C. Oliver, K. Sahlin & R. Suddaby (Eds.), *The Sage handbook of organizational institutionalism* (pp. 413–430). London: SAGE Publications.

Glynn, M. A., & Abzug, R. (1998). Isomorphism and competitive differentiation in the organizational name game. *Advances in Strategic Management*, *15*, 105–128.

Glynn, M. A., & Abzug, R. (2002). Institutionalizing identity: Symbolic isomorphism and organizational names. *Academy of Management Journal*, *45*, 267–280.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of 'blind' auditions on female musicians. *American Economic Review*, *90*, 715–741.

Greene, W. H. (2003). *Econometric analysis*, 5th edition. Upper Saddle River, NJ: Prentice Hall.

Greve, H. R. (2008). A behavioral theory of firm growth: Sequential attention to size and performance goals. *Academy of Management Journal*, *51*, 476–494.

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

Hansen, M. T., & Haas, M. R. (2001). Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, *46*, 1–28.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–161.

Hoffman, A. J., & Ocasio, W. (2001). Not all events are attended equally: Toward a middle-range theory of industry attention to external events. *Organization Science*, *12*, 414–434.

Hsu, G. (2006). Jacks of all trades and masters of none: Audiences' reactions to spanning genres in feature film production. *Administrative Science Quarterly*, *51*, 420–450.

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

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

Ingram, P. (1996). Organizational form as a solution to the problem of credible commitment: The evolution of naming strategies among US hotel chains, 1896–1980. *Strategic Management Journal*, *17*, 85–98.

Jensen, M. (2010). Legitimizing illegitimacy: How creating market identity legitimizes illegitimate products. *Research in the Sociology of Organizations*, *31*, 39–80.

Jones, C. (2001). Co-evolution of entrepreneurial careers, institutional rules and competitive dynamics in American film, 1895–1920. *Organization Studies*, *22*, 911–944.

Jones, C., Livne-Tarandach, R., & Balachandra, L. (2010). Rhetoric that wins clients: Entrepreneurial firms' use of institutional logics when competing for resources. In W. Sine & R. David (Eds.), *Institutions and Entrepreneurship: Research in the Sociology of Work*, 21, 183–218. Amsterdam: Elsevier.

Jones, C., Maoret, M., Massa, F. G., & Svejenova, S. (2012). Rebels with a cause: Formation, contestation, and expansion of the de novo category "modern architecture," 1870–1975. *Organization Science*, *23*, 1523–1545.

Jones, C., & Thornton, P. (2005). Introduction to transformations in cultural industries. *Research in the Sociology of Organizations*, *23*, ix–xix.

Karthikeyan, S., & Wezel, F. (2010). Identity repositioning: The case of Liberal Democrats and audience attention in British politics, 1950–2005. *Research in the Sociology of Organizations*, *31*, 295–320.

Keller, K. L. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *Journal of Marketing*, *57*(1), 1–22.

Keller, K. L., Heckler, S. E., & Houston, M. J. (1998). The effects of brand names suggestiveness on advertising recall. *Journal of Marketing*, *62*(1), 48–57.

Kennedy, M. T. (2005). Behind the one-way mirror: Refraction in the construction of product market categories. *Poetics*, *33*, 201–226.

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

Kennedy, M. T., Lo, Y-C., & Lounsbury, M. (2010). Category currency: The changing value of conformity as a function of ongoing meaning construction. *Research in the Sociology of Organizations*, *31*, 369–397.

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

Kovács, B., & Hannan, M. T. (2010). The consequences of category spanning depend on contrast. *Research in the Sociology of Organizations*, *31*, 175–201.

Kraatz, M. S., & Zajac, E. (1996). Exploring the limits of the new institutionalism: The causes and consequences of illegitimate organizational change. *American Sociological Review*, *61*, 812–836.

Lawrence, T. B. (1998). Examining resources in an occupational community: Reputation in Canadian forensic accounting. *Human Relations*, *51*, 1103–1131.

Lee, L-F. (1983). Generalized econometric models with selectivity bias. *Econometrics*, *51*, 507–512.

Lee, P. M. (2001). What's in a name.com? The effects of ".com" name changes on stock prices and trading activity. *Strategic Management Journal*, *22*, 793–804.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*(3), 74–89.

Loewenstein, J., Ocasio, W., & Jones, C. (2012). Vocabularies and vocabulary structure: A new approach linking categories, practices, and institutions. *Academy of Management Annals*, 1–46.

Lounsbury, M., & Crumley, E. T. (2007). New practice creation: An institutional perspective on innovation. *Organization Studies*, *28*, 993–1012.

Lounsbury, M., & Glynn, M. A. (2001). Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic Management Journal*, *22*, 545–564.

Lounsbury, M., & Rao, H. (2004). Sources of durability and change in market classifications: A study of the reconstitution of product categories in the American mutual fund industry, 1944–1985. *Social Forces*, *82*, 969–999.

Lumpkin, G. T., & Dess, G. G. (1996). Clarifying the entrepreneurial orientation construct and linking it to performance. *Academy of Management Review*, *21*, 135–172.

Marich, R. (2005). *Marketing to moviegoers: A handbook of strategies used by major studios and independents*. Burlington, MA: Elsevier Focal Press.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*, 340–363.

Moul, C. C. (2007). Measuring word of mouth's impact on theatrical movie admissions. *Journal of Economics & Management Strategy*, *16*, 859–892.

MPAA ratings. http://mpaa.org/FlmRat_Ratings.asp.

Navis, C., & Glynn, M. A. (2011). Legitimate distinctiveness and the entrepreneurial identity: Influence on investor judgments of new venture plausibility. *Academy of Management Review*, *36*, 479–499.

Negro, G., Koçak, Ö., & Hsu, G. (2010). Research on categories in the sociology of organizations. *Research in the Sociology of Organizations*, *31*, 3–35.

Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal*, *4*, 154–167.

Ocasio, W. (2011). Attention to attention. *Organization Science*, *22*, 1286–1296.

Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review*, *16*, 145–179.

Patterson, J. (2008). It's in the name. *The Guardian*, 5 April (http://www.guardian.co.uk/film/2008/apr/05/3).

Pedersen, J. S., & Dobbin, F. (1997). The social invention of collective actors: On the rise of the organization. *American Behavioural Scientist*, *40*, 431–443.

Pedersen, J. S., & Dobbin, F. (2006). In search of identity and legitimation: Bridging organizational culture and neoinstitutionalism. *American Behavioral Scientist*, *49*, 897–907.

Perrow, C. (1961). Organizational prestige: Some functions and dysfunctions. *American Journal of Sociology*, *66*, 335–341.

Peterson, R. (1997). *Creating country music: Fabricating authenticity*. Chicago: University of Chicago Press.

Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: A resource dependence perspective*. New York: Harper & Row.

Phillips, D. J., & Kim, Y-K. (2009). Why pseudonyms? Deception as identity preservation among jazz record companies, 1920–1929. *Organization Science*, *20*, 481–499.

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

Phillips, N., Lawrence, T. B., & Hardy, C. (2004). Discourse and institutions. *Academy of Management Review*, *29*, 635–652.

Podolny, J. M., & Phillips, D. (1996). The dynamics of organizational status. *Industrial and Corporate Change*, *5*, 453–471.

Pollock, T., & Rindova, V. (2003). Media legitimation effects in the market for initial public offerings. *Academy of Management Journal*, *46*, 631–642.

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

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*, 203–227.

Powell, W. W. (1991). Expanding the scope of institutional analysis. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press.

Power, M., Scheytt, T., Soin, K., & Sahlin, K. (2009). Reputational risk as a logic of organizing in late modernity. *Organization Studies*, *30*, 301–324.

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*, 795–843.

Rao, H., Monin, P., & Durand, R. (2005). Border crossing: Bricolage and the erosion of categorical boundaries in French gastronomy. *American Sociological Review*, *70*, 968–991.

Rindova, V., Barry, D., & Ketchen, D. J. (2009). Introduction to special topic forum: Entrepreneurial as emancipation. *Academy of Management Review*, *34*, 477–491.

Rosa, J. A., Porac, J., Runser-Spanjol, J., & Saxon, M. S. (1999). Sociocognitive dynamics in a product market. *Journal of Marketing*, *63*, 64–77.

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

Rosenkopf, L., & McGrath, P. (2011). Advancing the conceptualization and operationalization of novelty in organizational research. *Organization Science*, *22*, 1297–1311.

Ruef, M., & Patterson, K. (2009). Credit and classification: The impact of industry boundaries in nineteenth century America. *Administrative Science Quarterly*, *54*, 486–520.

Scott, R. (2001). Institutions and organizations. Thousand Oaks, CA: SAGE Publications.

Simon, H. A. (1947). *Administrative behavior: A study of decision-making processes in administrative organizations*. Chicago: Macmillan.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688–688.

Sorenson, O., & Waguespack, D. M. (2006). Social structure and exchange: Self-confirming dynamics in Hollywood. *Administrative Science Quarterly*, *51*, 560–589.

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

Swets, J. A., & Kristofferson, A. B. (1970). Attention. *Annual Review of Psychology*, *21*, 339–366.

Taylor, V., & Whittier, N. (1992). Collective identity in social movement communities: Lesbian feminist mobilization. In A. Morris & C. Mueller (Eds.), *Frontiers in social movement theory*. New Haven, CT: Yale University Press.

Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure, and process*. Oxford: Oxford University Press.

Vogel, H. L. (2007). *Entertainment industry economics: A guide for financial analysis*, 7th edition. Cambridge: Cambridge University Press.

Weber, K., & Dacin, M. T. (2011). The cultural construction of organizational life: Introduction to the special issue. *Organization Science*, *22*, 287–298.

Weber, K., & Glynn, M. A. (2006). Making sense with institutions: Context, thought and action in Karl Weick's theory. *Organization Studies*, *27*, 1639–1660.

Weber, K., Heinz, K., & DeSoucey, M. (2008). Forage for thought: Mobilizing codes for grass-fed meat and dairy products. *Administrative Science Quarterly*, *53*, 529–567.

Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: SAGE Publications.

Wry, T., & Lounsbury, M. (2013). Contextualizing the categorical imperative: Category linkages, technology focus, and resource acquisition in nanotechnology entrepreneurship. *Journal of Business Venturing*, *28*, 117–133.

Zerubavel, E. (1997). *Social mindscapes: An invitation to cognitive sociology*. Cambridge, MA: Harvard University Press.

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

Zuckerman, E. W. (2000). Focusing the corporate product: Securities analysts and de-diversification. *Administrative Science Quarterly*, *45*, 591–619.

Zuckerman, E. W., & Kim, T-Y. (2003). The critical trade-off: Identity assignment and box-office success in the feature film industry. *Industrial and Corporate Change*, *12*, 27–67.

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*, 1018–1074.

Author biographies

Eric Yanfei Zhao is a doctoral student in strategy and organization theory at the University of Alberta School of Business. He is also a research fellow at the National Institute for Nanotechnology (NINT) and the Canadian Center for Corporate Social Responsibility (CCCSR). His research interests lie at the intersection of strategy and organization theory, with a particular focus on social and technological entrepreneurship. His work has won various best paper awards from the Academy of Management and the Academy of International Business.

Masakazu Ishihara is an assistant professor of marketing at the Leonard N. Stern School of Business, New York University. His research focuses on quantitative marketing and empirical industrial organization, with particular interest in the dynamic effects of marketing strategies, forward-looking decision making by consumers and firms, and marketing in the entertainment industry. His work has appeared in *Quantitative Marketing* and *Economics* and *Management Science*. He received his PhD in marketing from the Rotman School of Management at the University of Toronto.

Michael Lounsbury is a Professor, Thornton A. Graham Chair, and Associate Dean of Research at the University of Alberta School of Business. He is also a Principal Investigator at the Canadian National Institute for Nanotechnology. His research focuses on institutional emergence and change, entrepreneurship, and the cultural dynamics of organizations and practice. He serves on a number of editorial boards and is the Series Editor of *Research in the Sociology of Organizations*, Associate Editor of *Academy of Management Annals*, as well as Co-Editor of *Organization Studies*. He received his PhD in sociology and organizational behavior in 1999 from Northwestern University.

Appendix

The Calculation of the New Genre Spanning Variable

Let G be the set of all distinct genres. Let j index the focal film and G_j^s be the set of genres for film j listed on classification source s in S, and let N_S be the number of classification sources. The fundamental idea of the new genre spanning variable is to allow for the possibility that audiences might view the genres in G_j^s differently from how the genres are listed in the source. For example, suppose $G_j^s = \{$Action, Adventure$\}$. In constructing GS_j, we assumed that audiences view Action and Adventure as separate genres. However, audiences might view them as one genre Action/Adventure. To allow for this possibility, we define the new genre spanning variable as a weighted sum of GS_j, where GS_j is computed under the assumptions that certain genre combinations are viewed as one genre by audiences, and weights are given by the probability that how frequently we observe such genre combinations in the historical set of films.

Formally, let H_j^s be the set of all possible combinations of distinct genres in G_j^s and let $H_j = \cup_{s \in S} H_j^s$. Let $P(H_j)$ be the power set of H_j, i.e., the set of all subsets of H_j including the empty set (ϕ) and H_j itself. Let h be an element of $P(H_j)$ and represent an event that audiences view each of genre combinations in h as one genre. For example, suppose that there are two classification sources and that $G_j^1 = \{Action, Adventure\}$ and $G_j^2 = \{Action, Horror\}$. Then, $H_j = H_j^1 \cup H_j^2 = \{Action/Adventure\} \cup \{Action/Horror\} = \{Action/Adventure, Action/Horror\}$, and $P(H_j) = \{\phi, \{Action/Adventure\}, \{Action/Horror\}, \{Action/Adventure, Action/Horror\}\}$. We interpret $h = \phi$ (empty set) as representing the event that audiences view neither *Action/Adventure* nor *Action/Horror* as one genre, $h = \{Action/Adventure\}$ as representing the event that audiences view *Action/Adventure* as one genre, but view *Action/Horror* as two separate genres. The same applies to $h = \{Action/Horror\}$. For $h = \{Action/Adventure, Action/Horror\}$, audiences view both *Action/Adventure* and *Action/Horror* as one genre. Note that in this example we do not consider *Adventure/Horror* because they never appeared together on the same classification source. We then define the weighted genre spanning variable, WGS_j, as

$$WGS_j = \Sigma_{h \in P(Hj)}\, w_j(h) * GS_j(h),$$

where $w_j(h)$ is the weight attached to an event h, and $GS_j(h)$ is the genre spanning measure that we defined on page 17, but computed under the event that audiences view each of genre combinations in h as one genre.

To compute $w_j(h)$, we define C to be the set of all possible combinations of all distinct genres. For each c in C, let $g(c)$ be the set of distinct genres in c. Let F be the set of all films released in the past five years prior to film j, and let N_F be the number of such films. We define the probability that we observe c in the historical set of films as

$$p(c) = \Sigma_{i \in F}\, \Sigma_{s \in S}\, I(g(c) \subset G_i^s)\, /\, (N_F * N_S),$$

where $I(\cdot)$ is an indicator function. Then, for h in $P(H_j)$, define

$$w_j(h) = \Pi_{c \in Hj}\, p(c)^{I(c \in h)}(1-p(c))^{1-I(c \in h)}.$$

In sum, WGS_j captures the notion that popular/familiar combinations of genres are viewed as spanning fewer genres than unpopular/unfamiliar combinations if the number of genres in the combinations is identical. It also accounts for the dynamic nature of the popular and familiar combinations of genres. In operationalizing this variable, we focused only on pairs of genres as possible genre combinations. This simplification is made because most important genre combinations are pairs rather than triplets or higher-order combinations. Yet, the procedure above can be extended to higher-order combinations of genres.

Table A1. Effects of Weighted Genre Spanning and Naming on the Opening Week Box Office Revenue.

Variable	Model 17 OLS	Model 18 OLS	Model 19 OLS	Model 20 OLS	Model 21 OLS
Weighted genre spanning	−0.46** (0.17)	−0.45** (0.17)	−0.45** (0.17)	−0.43** (0.17)	−0.46** (0.17)
Weighted genre spanning × Familiarity			0.06 (0.31)		
Weighted genre spanning × Reputation				0.29+ (0.23)	
Weighted genre spanning × Different sequel name				0.71 (0.68)	
Weighted genre spanning × Weighted reputation					0.03* (0.02)
Weighted genre spanning × Weighted different sequel name					0.04 (0.05)
Adjusted R^2	0.896	0.898	0.898	0.898	0.898
No. of observations	2827	2827	2827	2827	2827

Notes:
1. + p < .10; * p < .05; ** p < .01.
2. Standard errors are in parentheses.
3. One-tailed test for independent variables and two-tailed test for control variables.
4. All control variables used in Table 2 were included in corresponding models, but not reported here.