ACADEMY OF Management Journal

# CATEGORY SPANNING, EVALUATION, AND PERFORMANCE: REVISED THEORY AND TEST ON THE CORPORATE LAW MARKET



| | |
|---|---|
| Journal: | *Academy of Management Journal* |
| Manuscript ID: | AMJ-2013-0651.R3 |
| Manuscript Type: | Revision |
| Keywords: | Organization and management theory (General) < Organization and Management Theory < Topic Areas, Reputation < Organization and Management Theory < Topic Areas, Panel/Pooled < Research Design < Research Methods |




Abstract: Studies suggest that category-spanning organizations receive lower evaluation and perform worse than organizations focused on a single category. We propose that (1) these effects are contingent on clients' theory of value and that as clients expect more sophisticated services, they tend to value category spanners more positively and (2) the evaluation of producers mediates the relationship between category spanning and performance. We test our hypotheses using original data on corporate legal services in three markets (London, New York City, and Paris) over the decade 2000–2010. We find that (1) category spanners receive a better evaluation, and more so when their categorical combination is more inclusive and (2) evaluation mediates significantly the relationship between category spanning and performance. This study enriches our understanding of how audiences apprehend a whole market category system and why organizations span categories.

# Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market

**Lionel Paolella**
University of Cambridge
l.paolella@jbs.cam.ac.uk

**Rodolphe Durand**
HEC Paris
durand@hec.fr

**Acknowledgement**
We would like to thank Tim Pollock and the three anonymous reviewers for their excellent guidance. We would also like to thank Damon J. Phillips, Elizabeth G. Pontikes, Ezra W. Zuckerman, and David B. Wilkins for valuable feedback. The HEC Foundation and the Society & Organizations Research Center provided financial support for this research.

# CATEGORY SPANNING, EVALUATION, AND PERFORMANCE: REVISED THEORY AND TEST ON THE CORPORATE LAW MARKET

## ABSTRACT

Studies suggest that category-spanning organizations receive lower evaluation and perform worse than organizations focused on a single category. We propose that (1) these effects are contingent on clients' theory of value and that as clients expect more sophisticated services, they tend to value category spanners more positively and (2) the evaluation of producers mediates the relationship between category spanning and performance. We test our hypotheses using original data on corporate legal services in three markets (London, New York City, and Paris) over the decade 2000–2010. We find that (1) category spanners receive a better evaluation, and more so when their categorical combination is more inclusive and (2) evaluation mediates significantly the relationship between category spanning and performance. This study enriches our understanding of how audiences apprehend a whole market category system and why organizations span categories.

**Key words**: category, evaluation, performance, mediation, law firms.

In this paper, we question the dominant wisdom that category spanning is detrimental to an organization. *Category spanning* refers to organizations' involvement in several activities that comprise distinct cognitive sets; we use detrimental to describe activities leading to either damaged evaluation or reduced performance. Compared with their "purer" rivals, category spanners have been found to be less acknowledged and rewarded. Accordingly, they are penalized by being granted lower evaluation and experiencing poorer performance (Hsu, 2006; Negro, Hannan, & Rao, 2010; Zuckerman, 1999). To account for variations in category spanners' penalties, past research has associated different cognitive attributes with distinct audiences and contrasted mechanisms of sanctions. For instance, investors are category-makers, while clients are category-takers (Pontikes, 2012), and producers who span multiple categories suffer from various perceptual disadvantages when accounting for typicality and underlying quality differences (Kovács & Johnson, 2014; Leung & Sharkey, 2014). These studies, however, have left untouched the principle that clients prefer focused producers.

Counter to the above arguments, organizations continue to combine categories and to redefine their activity portfolio: everyday evidence includes Google, Amazon, and Apple but also GE and Xerox. Furthermore, if specialists were the fittest everywhere, at any time, a state of equilibrium would materialize between audience expectations and corresponding producers, but category systems continue changing and producers continue to combine categories (Granqvist, Grodal, & Woolley, 2013; Wry, Lounsbury, & Jennings, 2014). Finally, past research has studied the effects of category spanning on evaluation and performance independently, despite the fact that further associations may exist between these factors.

We address these issues and propose that clients have no general and unconditional preference for organizations focused on a single category. What matters is the theory of value that clients stand for and perform (Lamont, 2012; Zuckerman & Rao, 2004). By *theory of value*, we refer to audiences' identification of issues and solutions, and their ascription to solution providers of a value order, which, in competitive markets translates into a willingness to pay and higher prices (Podolny, 1993; Uzzi & Lancaster, 2004). Depending on the issue at stake—entertaining oneself or acquiring a rival firm—audiences' theories of value will differ, and hence audiences will differ in how they prioritize the valuable offerings (Durand & Paolella, 2013; Shiller, 1990).

Past research suggests that clients prefer typical products and focused producers (Hsu, 2006; Hsu, Koçak, & Hannan, 2009). This is undoubtedly true but whereas specialist producers equate with simpler and more focused outputs, category spanners equate with more complex clients' demands. Hence, we argue that when issues are complex, i.e., require more sophisticated sets of services, are non-recurrent, and/or involve high financial stakes, clients place greater value on category spanners, as they are considered more capable of handling the case in point, such as in our example of the acquisition of a rival. In these situations, clients mete out the goal they pursue before the typicality of each producer relative to a category prototype. Clients' theory of value in these cases passes from a prototypical view to a goal-based view (Barsalou, 1991; 2005; Durand & Paolella, 2013), leading clients to value positively those suppliers that span categories and thereby signal their sensitivity to elaborated demands. These instances of positive valuation should be truer as the categories combined are more inclusive—that is, as the categories occur together more frequently at multiple levels of combination. Inclusiveness acts as a marker

of the saliency and acceptability of a category combination (Murphy, 2004; Kennedy, Lo, & Lounsbury, 2010) and thus should moderate the relationship between category spanning and evaluation.

Importantly, when clients find category spanners to be more attractive, by seeing them as more capable of handling their complex cases, the resulting better evaluation should manifest both in higher revenues for producers and in a partial account of the direct relationship between category spanning and performance. That the effect of category spanning on performance could be negatively or positively mediated by audiences' evaluation as a function of their theory of value has been ignored in prior research, which focused on the direct negative effects of category spanning on performance (Hsu et al., 2009; Kovács & Johnson, 2014).

The setting for our study is the corporate legal services market. Whereas looking for a good movie or a great *Barolo* wine requires identification and selection based on straightforward preferences at the product level (Hsu, 2006; Negro & Leung, 2013), corporate legal transactions are case-specific and involve multiple category expertise—that is, numerous practice areas of law at the organizational level (Chatain, 2011; Kor & Leblebici, 2005). For instance, a client who proceeds with acquisitions, or invests in foreign countries, needs advice simultaneously in Corporate-M&A, Tax, and Employment laws. Corporate legal services also involve significant economic consequences, as their costs can amount to hundreds of thousands or millions of dollars. Clients' theory of value therefore comprises the criteria of sophistication, singularity, and expensiveness. We test our predictions on a unique dataset of clients' evaluation of eight practice categories (Competition-Antitrust, Litigation, Intellectual Property, Real Estate, Tax, Corporate-

M&A, Bankruptcy, and Employment) over a decade (2000–2010) in three different locations (London, New York City, and Paris). The results support our conjecture that in this setting where output requirements are complex, category spanning is positively associated with evaluation, and that spanning inclusiveness reinforces this association. Moreover, we find a positive association between evaluation and performance, and that evaluation mediates 37% of the effect of category spanning on performance.

This paper studies concurrently categorical spanning, evaluation, and performance and evidences the mediation of evaluation on the category spanning-performance relationship. It expands the research on organizations and category systems by suggesting conditions at which audiences' theory of value changes to favor a goal-based perspective over the classical prototypical perspective. Although we do not simultaneously study simple and complex clients' requirements and the corresponding theories of value, this paper helps identify the validity domain of past research (Hannan, Polos, & Carroll, 2007; Hannan, 2010; Hsu et al., 2009; Leung & Sharkey, 2014; Negro & Leung, 2013). This work broadens the scope of category research at the international, firm-to-firm level and demonstrate also the importance of the degree of inclusiveness of category spans within the market's whole category system. Finally, this study speaks to the diversification literature and answers the call to research "beyond the "golden cage" of categorical membership norms within which organizations are held" (Durand & Kremp, 2015: 38).

## THEORY BACKGROUND

The functioning of markets relies on common references for a reciprocal understanding between actors. Categories fill that need by acting as sociocognitive classification apparatus shared among market actors (Hsu & Hannan, 2005; Rosa, Porac, Runser-

Spanjol, & Saxon, 1999). As collection points for information, categories streamline market interactions by grouping together similar organizations and excluding dissimilar ones based on their similarity with prototypical features. A prototype, i.e., the ideal member, constitutes the cognitive reference point for defining category membership based on organizations' possession of certain attributes (Durand & Paolella, 2013; Rosch & Mervis, 1975). As such, category memberships foster understanding between actors and guide behaviors in recognizable ways by developing the expectations of and for participants.

In this sense, categories enable the identification of firms, and those firms that fail to fit into a single recognizable category confuse clients, and thereby risk being ignored or penalized. For example, in the case of asset markets, Zuckerman (1999) argues that category-spanning organizations experience greater coverage mismatch (i.e., the difference between the categories claimed by a firm and the categories in which third parties assign that firm). Hence, categories are mainly understood as a vector of prohibitions or prescribed obligations: due to the expectations tied to prototypes, they can be seen as disciplining frameworks that restrict organizations' leeway by limiting what they can and must do (Hannan et al., 2007). Market participants who are members of multiple categories face disadvantages, whereas firms that fit neatly into a category receive better evaluation and gain higher audience consensus (Hsu, 2006). According to this perspective, the notion of category is therefore grounded in features of similarity (i.e., their closeness to prototypes) and exists with relative independence from clients' goals and valuation. A logical correspondence thus exists between category spanning—understood as a misfit—lowered evaluation, and, in turn, firm performance.

However, first, we question the generalizability of this theory and argue that what matters more to audiences is not whether producers trespass category boundaries but their own ability to identify and comprehend the category combinations offered by producers. For instance, Wry and colleagues (2014) find that, depending on the category header—the dominant category used for identification—category spanning produces distinct effects: science-based startups that add technology patents fare better than technology-based startups that add science patents because the former associate cutting-edge advancements with a will to reach the market, whereas the latter gives venture capitalists the impression that it aims to move upward and away from the market. Hence, a first insight suggests that preceding any evaluation (either pro or con category spanning) is how audiences define their requirements and determine which producers represent the best match for their situations.

Second, as Durkheim (2008: 145) has argued, "when a classification is reduced to two classes, these are almost necessarily conceived as antitheses", which accentuates the cognitive confusion brought about by category spanning, as in the *classical* vs. *nouvelle* categories of the haute cuisine context (Rao, Monin, & Durand, 2005) or in the more recent wine studies on *Barolo* vs. *Barbaresco* (Negro et al., 2010) and the science vs. technology-based startups in the carbone nanotube industry (Wry et al., 2014). Hence, both theorizing and empirical investigations need to be less restrictive, and instead portray markets with a broader system of categories not limited to oppositional cases[1]. When considering multiple categories, audiences may find some category combinations valuable and others meaningless. For the producers offering valuable combinations, a

[1] Hsu (2006: 425) explains that the decreasing relationship between category spanning and organizational appeal holds under the evident condition that "when categories are incompatible or oppositional, producers who attempts to span positions encounter substantial difficulty in appealing to and retaining consumers".

question is in order: Can they receive even better evaluation than mono-category producers? This may be true, according to some evidence from prior studies. For instance, Rosa and colleagues (2005) find that motorcycle manufacturers whose various products represented multiple categories (e.g., cruisers and tourers) experienced greater appeal than makers of motorcycles that represented only one category. Similarly, in their historical case study about a household product manufacturer, Rindova and colleagues (2011) describe precisely how an organization can draw on multiple market categories to build its identity and gain the appreciation of its clients.

Third, most studies about category spanning are at the product level (e.g., movies, recipes, auctions, wines), and the results are aggregated at the organizational level. This inference makes perfect sense from a prototypical view of categories, a producer being judged as the aggregate of its full array of offerings. Yet, when the theory of value makes clients judge first the producer and second its products, the fact that some products do not appear to fit preexisting categories could be seen as secondary to whether the producer possesses a coherent market identity, congruent with the clients' complex requirements (Alexy & George, 2013; Granqvist et al., 2013; Rindova et al., 2011). These reflections prompt the question: What fundamental relationships link category spanning, evaluation, and performance at the organizational level?

## HYPOTHESES

As Lamont states, evaluation, as a set of processes that contributes to making judgments and assessing entities, requires categorization (Lamont, 2012: 206). As such, judgments and assessments are not necessarily pre-set to favoring prototypical producers. Instead, they vary in terms of the goals and issues at stake, i.e., audiences' theory of value. By

*theory of value*, we refer to how audiences identify issues and solutions, ascribe value, and rank solution providers (Lamont, 2012; Shiller, 1990; Zuckerman & Rao, 2004). To use Lamont's terms, an audience's valuation—in effect, the application of its theory of value in connection with a category system—must necessarily precede its evaluation of an organization as being suitable for its needs. Past studies have evoked theory of value as a condition for their results to hold, but have not fully embraced its consequences on evaluation. For instance, in his notable paper on the categorical imperative, Zuckerman (1999: 1431) writes that the "industry-based category structure analyzed in this article is contingent on the prevalence of a particular theory of value […]. The possibility of alternative theories of value and classification schemes persists, as indicated by the welcome received by the conglomerate firm in the 1960s". Here and in other of his works (Phillips, Turco, & Zuckerman, 2013; Zuckerman, Kim, Ukanwa, & Rittmann, 2003: 1030-31), Zuckerman draws attention to the importance of understanding which theory of value prevails within a market, the risk being to take the functioning and theory of value prevalent in asset markets (an extreme case) as a rule for all markets.

A sound departure point is therefore an audience's theory of value, and whether and to what extent producers match its requirements and are capable of addressing its cases. Hence, an audience attends to producers' signaled capacities to solve issues and fix situations, whether or not they are categorically pure. Therefore, relative to the more common cases of consumption previously studied, as the issues and situations become more complex, i.e., involve more sophisticated requirements, singular cases, and higher stakes, audiences are likely to identify and evaluate differently narrow- and broad-range producers.

We posit as the core of our theorizing that the preference for specialists or category spanners is therefore a consequence of the theory of value held by audiences in relation to the situations and cases they face. As a result, categories can appear not compatible for an exogenous taxonomist—say, a scholar attributing categories to a set of producers—but desirable and needed for clients in terms of the goal they pursue. This distinction has been underscored in the categorization literature, for instance, in Durand and Paolella (2013), where the authors separate the prototypical view from the goal-based view on categories. The former assumes an accepted prototype that audiences use as a reference point to measure distance; and as the distance between a producer's identity and the corresponding prototypes grows, the evaluation diminishes, as does, one supposes, performance. The latter, goal-based view, starts with what audiences need and with the category or category combination that fit the purpose they demand (Ratneshwar, Barsalou, Pechmann, & Moore, 2001). This approach constitutes an audience's theory of value that is not exclusive from considering spanners as valuable suppliers.

Precisely, the prototypical view is congruent with standard purchase decisions, where simple and focused outputs are offered by producers–and for which category spanning is detrimental. When considering settings that involve more complex outcome requirements, the goal-based view predominates over the prototypical view on categories and audiences will pay heed to producers as a whole. As audiences deal with complex cases, organizations that span broad categories receive better evaluation, as they are likely to attend better to audiences' requirements and the specifics of their cases. In these situations, producers offering various activities concurrently transfer neither vagueness of identity nor doubts about competence to clients (Smith, 2011). Category-spanning

producers are identified as experts in the activity categories they offer, and are regarded as being more capable than focused players of tackling the variety of specificities within and across categories involved by the transaction—and so receive higher evaluation (Kim & Jensen, 2011). Therefore, category spanning signals a supplier's capacity to handle complex situations and meet their clients' theory of value. In response, clients value category spanners more highly. Thus:

> ***Hypothesis 1***. *When outcome requirements are complex, category spanning will be associated with more positive evaluation.*

Whereas the prototypical view uses a simple rule to determine how valuation leads to evaluation (i.e., distance from a prototype), for the goal-based view, not all category configurations are equal. Audiences do not adopt analytical and isolated perceptions of each category independently of others but perceive categories simultaneously. Thus, combinations of categories have been found to matter as much as, or even more than, the grade of membership in each category (Meyers-Levy & Tybout, 1989)[2]. Some of them are perceived as making more sense than others (Phillips et al., 2013; Leung, 2014; Ruef & Patterson, 2009; Wry et al., 2014), in particular due to their inclusiveness (Kennedy et al., 2010), i.e., the cumulative frequency of conjoint categories at different levels of combination (duos, trios, and so on). Audiences cognitively associate the more interlocked combinations of categories with more recognizable answers to their situations. Inclusive spans constitute thematic relations that are viewed as being more salient and thus more valuable (Lin & Murphy, 2001). We expect that audiences will value more category spans

[2] To make an analogy, people understand an adjective-noun phrase more easily when the adjective is frequently associated with this noun. For example, the expression "loud museum" is more difficult to understand than "quiet museum" because the adjective "loud" does not match with the expectations attached to the noun "museum" (Murphy, 2004: 398).

that group together categories that co-occur more frequently, as these spans "make more sense" and are more accepted (Kim & Jensen, 2011; Lo & Kennedy, 2015).

For example, suppose that a market comprises four different categories, A, B, C, and D. Category spanners may combine as follows: AB, AC, AD, BC, BD, and CD for pairs; ABC, ABD, ACD, BCD for trios; and ABCD for a quartet. Imagine now that, for pairs of categories, only AB and BD are present in this market; for the trios, only ABD, ACD, and BCD exist; and some firms also adopt the quartet. From this example, we can observe that AB and BD are represented a great deal—as pairs, in the trios ABD and BCD, and in the quartet. Clients are prone to endorse more these category combinations that likely cohere better with their goals and expectations than those combinations that appear rarely or randomly, as say, AC, which does not exist as a pair. In our trios' examples, ABD will likely be most inclusive–and most meaningful– as it includes the two pairs AB and BD, then BCD which contains BD only, and finally ACD is the least inclusive since it exists only on its own.

Hence, category spans are not uniformly rewarded, and a category combination's degree of inclusiveness positively moderates the relationship between category spanning and evaluation since producers whose category spanning attains high levels of inclusiveness will be more salient and viewed as more capable than those offering less inclusive categorical combinations. Therefore:

> ***Hypothesis 2***. *The adoption of more inclusive category combination will have a positive influence on the relationship between category spanning and evaluation.*

**Evaluation as mediation**

In competitive markets, as value orders coincide with a willingness to pay and prices (Podolny, 1993), producers who are better evaluated are more likely to benefit from superior revenues (Greenwood, Li, Prakash, & Deephouse, 2005). Indeed, where quality is difficult to observe and judge, evaluation is taken as indirect indicator of capability and commitment (Karpik, 2010; Phillips et al., 2013; Benjamin & Podolny, 1999). For instance, in the corporate law industry, Uzzi and Lancaster (2004) show that producers' evaluation enhances the client's own image—which incites to contract with more expensive law firms, underscoring the positive relationship between a producer's evaluation and its performance.

We expect too that performance is positively associated with better evaluation but consider jointly the relationships between category spanning and evaluation on performance. Past studies use several arguments to explain the direct relationship between category spanning and performance. Most of them are related to evaluation such as audiences' ability to identify and assess producers and competence signaling. For instance, in their analysis of a sample of auctions distributed across 23 different product categories, Hsu and colleagues (2009) observe that sellers engaged in more than one category were less likely to sell items since they were less able to use the correct acronyms and quality descriptors for their products. Other arguments pertain to the influence of resource synergies or duplicates (Stimpert & Duhaime, 1997), of production factors, and of coordination interfaces across activities and geographical areas (Jones & Hill, 1988; Kor & Leblebici, 2005).

However, by not isolating the net effect of evaluation from other causes leading to performance, previous results on the direct effect of category spanning on performance may have been partially- or under- specified. Indeed, the bulk of the explanation of the relationship linking category spanning and performance consists of evaluation arguments, which are contingent on the clients' demands and theory of value. Therefore, evaluation very likely mediates partially the effect of category spanning on performance. Through spanning, depending on their clients' theory of value, producers expect better or lower evaluation and, indirectly, better or lower economic returns. In cases where outcome requirements are complex, spanning producers expect positive evaluation and with higher evaluation, higher performance. In these instances, evaluating spanning producers positively redounds economic value upon them. Other causes associating directly category spanning with performance still hold (i.e., resource characteristics, production factors, and intra-organizational coordination), but we expect that a significant portion of the total effect passes through evaluation. Combined with prior arguments on category spanning and evaluation and on positive association between evaluation and performance, it follows:

> ***Hypothesis 3***. *A producer's evaluation mediates partially the relationship between category spanning and performance.*

## EMPIRICAL CONTEXT AND DATA

Our empirical context is the corporate law market, a service industry where producers deal with complex cases and clients face difficulties in assessing producers' quality ex ante. We conducted a series of 17 preparatory interviews with general counsels, corporate lawyers, guides' editors and legal experts in three economic hubs that we use in our

quantitative analysis: London, New York City, and Paris. They explained that the "judicialization" of the economic world led to the emergence of an international corporate legal services market based on the American model (Dezalay & Garth, 2004). Unlike in the past, clients today have less need for a lawyer who acts exclusively for their litigation cases, preferring instead a partner who can navigate them through their whole legal life and advise on major corporate events, such as mergers or acquisitions, or the creation of foreign subsidiaries (Wilkins, 2009). Thus, law firms across countries are mushrooming into large legal "department stores", offering a panoply of services delivered by many lawyers (Galanter & Palay, 1994; Harper, 2013). Structured by practice disciplines (often regarded as "silos"), corporate law firms have been offering combinations of their various competencies rather than offering them each independently in an effort to fit their clients' theory of value. As expressed by two interviewees:

> Our firm has historically understood our clients' needs and how we can best help them address those needs. Our clients want sophisticated and responsive legal service [and] are looking for a full package of services. We need to offer a set of diversified practices in different regional areas. (Interview with a UK law firm partner, London office)

> If we want to develop our more profitable practice [areas], which is corporate/M&A or litigation, we need skills in tax, intellectual property, real estate, employment, occasionally in environmental law and so on. To close the deal, you generally need expertise in diverse areas. So, having partners and teams in these areas helps us to enhance our core practice to attract clients and make deals. For us, being a multi-practice firm is necessary to close the deal. (Interview with a US law firm partner, Paris office)

Clients confirm that corporate law firms "take so much the stress out of even the complex deals, anticipating our needs before they arise" (*The Chambers and Partners UK*, 2010: 1352, about the law firm *Ashurst*). Therefore, more corporate law firms tend to

cater to their clients with a broad array of services in which multiple categories of expertise are represented to manage deals worth millions of dollars.

In this setting, clients also use the potential corporate law partners' social standing as an observable signal of their unobservable quality in an attempt to reduce uncertainty surrounding the contractual decision (Uzzi & Lancaster, 2004). Since legal services are expensive and their direct value difficult to assess, clients' attention tends to shift from the product to the potential partner as a whole (Podolny, 1993). As a result, law firms aim to enhance their evaluation (i.e., rankings) in the market as a means to become more "desirable". When successful, such effort "expands the number of clients and strengthens the intensity of client-brand relations" (Karpik, 2010: 164), which, in turn, should generate positive economic returns. Thus, the corporate law market constitutes a relevant setting for testing our predictions.

We collected data on corporate law firms over a period of one decade (2000–2010), from the three most recognized guides in the profession: *The Chambers and Partners*, *The Legal 500*, and *PLC Which lawyer*. We coded eight law practice categories (Competition-Antitrust, Litigation, Intellectual Property, Real Estate, Tax, Corporate-M&A, Bankruptcy, and Employment) in three locations: New York City, because it is the largest legal services market around the world (in terms of volume and values of transactions), and London and Paris the two main European hubs for corporate law firms. According to our interviewees, these eight practice categories covered the main scope of business law firms. The three legal guides rank law firms for each practice and each location by reflecting market opinions of various audience members (*International In-house Counsel Journal*, 2010: 21; Coates, DeStefano, Nanda, & Wilkins, 2011). More

precisely, the guides' research is based on the combined analysis of law firms' submissions of deals or reported contentious issues, commentaries from the lawyers, and opinions and feedbacks from clients interviewed in each jurisdiction. For example, *The Chambers and Partners* conducted more than 7,000 in-depth interviews with clients for the U.S guide in 2010. The three guides use the same methodology and similar practice denominations:

> The guides are similar, they adopt the same process for their survey: submissions, interviews with clients and lawyers. And we send the same submissions for all guides…with the same kind of interviews with the lawyers. I think you can definitely compare them. (Interview with a Marketing Communication Advisor, US law firm, New York City office)

Our dataset comprises all law firms that have been ranked in at least one practice category by at least one of the three guides over the period, and our level of analysis is the law firm's branch office—the firm-location dyad. For example, in our dataset, *Linklaters*-London, *Linklaters*-New York, and *Linklaters*-Paris represent three distinct entities. For each guide, a given entity is covered in maximum 8 practices, giving a maximum of 24 ratings for a firm if present in the three countries.

In addition, for all law firms ranked by the guides, we collected complementary data (e.g., gross revenue, total number of partners and lawyers, number of offices, location of offices, etc.), using three professional journals that report such data: *The Lawyer* (for London), *American Lawyer* (for New York City), and *Juristes et Associés* (for Paris).

## MEASURES AND METHOD

***Evaluation.*** The three guides adopt an ordered scale from 1 to n bands, with 1 representing the highest rank. Both the total number of bands used and the number of firms covered vary from practice to practice over time. The number of bands is therefore

not meaningful in itself—for example, the ranking of a firm in band 2 depends on whether the guide uses four or seven (or another number) of bands. Contrary to a system that uses a constant number of stars (Rao et al., 2005), we cannot take only the figure of the band in which the firm is ranked. For each guide *g*, we thus propose the following calculation to rate a firm-location *x* in a given practice *p*:

$$Rating_{g,p}\ x = 1 - \frac{band\ of\ \ firm\ x_{g,p} - 1}{total\ bands\ in\ ranking_{g,p}}$$

As a result, *rating* ranges in a decreasing value order from 1 (the highest-rated firms rank in the first band and hence gain a *rating* of 1) to value close to zero for firms at the bottom of the guides' classification. We first calculated the average rating each firm gained in each practice area over the three guides. Thus, in Paris, a firm ranked in "Tax" in two guides (with the *ratings* 1 and 0.8) and in "Real Estate" in three guides (with the *ratings* 0.5, 0.2, and 0.2) received an average rating of 0.9 for its "Tax" practice, and 0.3 for in "Real Estate". Next, we computed the evaluation over all the practice areas in which the firm was ranked. Following the example above, the final value of evaluation for this firm-location is 0.6: (0.9 + 0.3) / 2. Therefore for a firm-location *x*:

$$Evaluation\ x_t = \sum_{p=1}^{8}\left(\frac{1}{p} \times \sum_{g=1}^{3}\frac{rating_{g,p}}{g}\right),$$

where *p* is the number of practice areas in which the firm-location *x* is covered by at least one guide, and *rating* as explained above the rating obtained in the guide *g* for each practice area that the firm *x* covered at time *t*.

***Performance.*** The variable *performance* is based on the Revenue Per Lawyer (RPL) at the law firms' branch office, a commonly used measure of performance in law firm and professional service firm studies (Greenwood et al., 2005; Malos & Campion, 2000). To

calculate this variable, we multiplied RPL (i.e., worldwide revenues divided by worldwide number of lawyers) by the branch office's number of lawyers. *Performance* estimates the portion of total revenue generated by a branch office as a function of its number of lawyers. In our models, it is expressed in hundreds of millions dollars (to avoid distortions between locations, we used financial figures in international dollars adjusted for purchasing power parity using World Bank estimates).

***Category spanning.*** First, in the same fashion as used in previous studies (e.g., Negro et al., 2010), we calculated for each firm the function $\mu(p, x, t)$, which indicates the Grade of Membership (GoM) of the firm-location $x$ in the practice area $p$ (in at least one guide) at time $t$. An organization's practice portfolio is the set of GoMs: [$\mu$(*competition, x, t*), $\mu$(*litigation, x, t*), $\mu$(*intellectual property, x, t*), $\mu$(*real estate, x, t*), $\mu$(*tax, x, t*), $\mu$(*corporate-m&a, x, t*), $\mu$(*bankruptcy, x, t*), $\mu$(*employment, x, t*)]. For example, for a specialist law firm in "Real Estate", its GoM vector can be represented by [0, 0, 0, 1, 0, 0, 0, 0]. For a full-service law firm covered in the eight practice areas, its GoM vector is [0.125, 0.125, 0.125, 0.125, 0.125, 0.125, 0.125, 0.125]. We then computed category spanning as 1 minus the Simpson's index of its GoM vector:

$$Category\ spanning\ [\mu(p, x, t)] = 1 - \sum_{p \in P} \mu^2(p, x, t)\ ,$$

that is, 1 minus the sum of the squared GoMs of each firm-location $x$ at time $t$, with $p$ is the set of practice areas in which the firm operates. The minimum value for *category spanning* is 0 (for specialists engaged in only one practice area), and the maximum 0.875 (for generalist law firms covering the eight practice areas). Figure 1 shows the evolution of the average levels of the category spanning variable over the observed years. Over the period, we see that firms' average category spanning across practice areas has grown

slightly. These values are higher than values noted in previous studies that used product-level data in simpler business-to-consumer transactions (e.g., Negro et al., 2010: 1407; in which the average category spanning is between 0 and 0.15 over the period of observation). As an alternative measure, we first calculated a firm's category spanning for each guide in which the firm is present and then averaged the values to obtain the average *category spanning* across the three guides. This measure correlated highly with the measure we adopted and led to the same results.

[INSERT FIGURE 1 ABOUT HERE]

***Inclusive category combination.*** What matters for clients is the co-occurrence of practices associations—that is, whether practices A and B frequently occur or not at various degrees of combination. To determine a firm's *inclusive category combination,* we evaluated the degree of insertion of category combinations in the entire population of producers for each location each year. More precisely, for each firm-location in our dataset, we used a set of eight dummy variables to flag their coverage in the eight categories in at least one guide (e.g., 1 = presence and 0 = absence). We concatenated this ordered set of categories to obtain a vector (an octet) reflecting the category combination of each law firm. For example, if a firm is affiliated with the first, the second, and the eighth category, its organizational vector is 11000001. In this example, the octet includes three pairs: 10000001, 11000000, and 010000001. Our variable *inclusive category combination* corresponds to each year and each location, for a given octet, to the logged sum of the frequencies for all the existing lower-degree n-uplets, i.e., septuplets, sextets, quintets, quartets, trios, and pairs that are included in the octet. Therefore, the vector 11111111 is the most inclusive configuration and contains the sum of all the other existing combinations in a given year in a given location, whereas the least inclusive is a

pair that is present only once (e.g., 10000010). Specialists do not combine categories and therefore get a value of zero for this variable.

Instead of treating only the frequency of exactly identical combinations as being indicative of level of acceptance of a given combination (a measure of frequency-based legitimacy as in many past studies), and in accordance with clients' theory of value, our measure allows us to capture the vertical insertion of every pair, trio, etc., in the whole category system. Hence, one quartet, q1, could be more represented in the population than another, q2, but comprise less frequently observed pairs and trios than q2—that is, q1 is less inclusive and coherent than q2 with clients' theory of value. From our data, Figure 2 shows examples of a pair (Tax/Corporate-M&A) included in a trio and in two quartets. We can observe that the trio Litigation/Tax/Corporate-M&A is actually more inclusive than one of the quartets at the beginning of the period (Competition/Tax/Corporate-M&A/Employment), and that the other quartet (Competition/Litigation/Tax/Corporate-M&A) remains consistently the most inclusive of all these combinations over the period.

[INSERT FIGURE 2 ABOUT HERE]

***Control variables***. We controlled for other factors that might influence our dependent variables. Most research invokes a cognitive confusion mechanism to explain the association between category spanning and lower evaluation, as "membership in multiple (non-nested) categories likely confuses the audience and makes a producer appear to fit poorly to any of the schemata that an agent applies to the categories" (Hannan et al., 2007: 108). Multi-category memberships that straddle identifiable categories therefore would lead to audience confusion, and to clients misattributing both properties and expectations. To capture the clarity of a focal firm in the eyes of its clients,

we included the variable *categorical consensus* among the three guides to measure their agreement as to firm's areas of practice coverage. We calculated the average similarity between each pair of guides that cover a firm and, following previous studies (Hsu, 2006), used the Jaccard similarity index to capture similarity between the coverage of firm's practice areas in each pair of guides. The Jaccard coefficient takes the following form: $J(A,B) = \frac{A \cap B}{A \cup B}$ where $A \cap B$ indicates the cardinality of the set of practice areas in which the firm is covered in both guides A and B, and $A \cup B$ the cardinality of the set of practice areas in which the firm is covered in guides A and/or B. For example, in 2001, the firm-location dyad *Sullivan and Cromwell* in Paris is covered in the areas "Corporate/M&A" and "Tax" in both *PLC Which Lawyer* and the *Legal 500* but only in "Corporate/M&A" in the *Chambers and Partners*. Here, the values of each pairwise comparison are 1 (for the pair *PLC Which Lawyer/ Legal 500*), 1/2 (for the pair *PLC Which Lawyer/ Chambers and Partners*), and 1/2 (for the pair *Legal 500/ Chambers and Partners*). The average value of the three pairwise comparisons is (1 + 1/2 + 1/2) / 3 = 0.67. The range of the variable *categorical consensus* is between 0 and 1, with some firms showing no consensus at all, whereas others reach partial or full consensus on their coverage.

Previous research also shows that clarity regarding the meaning of a category increases the appeal of all its members (Kovács & Hannan, 2010). In contrast, multi-category firms blur the saliency of the categories to which they are affiliated with, leading to audiences reacting negatively to their lack of clarity. We therefore controlled for the average *categorical contrast* of categories that a firm spans. Based on previous studies (e.g., Kovács & Hannan, 2010: 184-85), we measured the contrast of a category as the sum

of the grades of membership of the category members divided by the total number of members belonging to that category. As an example, if the category "Litigation" has three members with grades of membership equal to 0.2, 1, and 0.5, then the categorical contrast of the category equals (0.2 + 1 + 0.5) / 3 = 0.57. We also included the average *tenure in guides* of the firm's coverage across the three guides (Pontikes, 2012) to measure the duration since the firm-location was first registered by any of the three guides. We first calculated the cumulative number of years for the firm's successive presence in each guide and then computed the firm's average number of years over the three guides.

At the firm level, independent of a law firm's category spanning, its visibility on the national market may increase its evaluation (Karpik, 2010: 163), so we controlled for the local *size* of the firm, using the log of the total number of partners it employed nationwide. We also controlled for the *age* of the firm-location—that is, the number of years since it opened an office in London, New York City, or Paris. Over the past three decades, in a remarkable trend, law firms have expanded globally in the corporate legal services markets (Dezalay & Garth, 2004). We therefore included the control variable *internationalization degree* as the logged number of countries in which the firm had a branch office.

Regarding their independence and objectivity of research, all guides state that law firms cannot pay for their inclusion in the rankings—for instance, *The Chambers and Partners'* editorial introduction claims: "inclusion in sections of the guide is based solely on the research team's findings. No-one can buy their way in". However, all these guides contain a directory section consisting of professional cards and profiles and in case the purchase of a full-page profile influences a guide's ranking decisions, we controlled for

the potential effect of *advertisement in guides* by counting each firm's average number of page profiles purchased across the three guides. As some practice areas of law are more prestigious or profitable than others, and the size of each department within firms may also vary, firm's presence in a particular category may affect clients' evaluation and firm performance. We therefore included dummy variables for *coverage in a focal category* to control for specific effect of each practice area. Finally, to account for the potential effect of firm's *nationality* on the dependent variables, we added dummy variables to signal the country of the firm's headquarter. To capture guide-specific effects and the possibility of measurement errors in reflecting clients' opinions that might influence ratings of firms, we included dummy variables for the presence in each guide. We finally captured *time effects* and *location effects* by including a set of dummy variables in our models.

**Model choice**

We opted for random-effects generalized least squares (GLS) estimations for several reasons. First, our theoretical argument explains the differences between firms in evaluation and performance, conditional on their category spanning and inclusive category combination; thus, we used analyses that estimate between-firm differences over years. Second, by decomposing the variance of our regressors, we observed very low variance within firms; thus, it was preferable not to use fixed-effects models, as the coefficients of regressors with little within-variation are imprecisely estimated (Cameron & Trivedi, 2010: 244). Third, random-effects regressions allow time-invariant factors as explanatory variables, which is the case in our setting (e.g., the variable location or the category dummies). We confirmed that random-effects regressions were appropriate by using the Breusch-Pagan Lagrange multiplier test to confirm that variances across entities

differed from zero ($p < .001$). However, the error terms of regressions with time-series data may not be independent, as errors are often serially correlated and not identically distributed. Both the Durbin-Watson and Wooldridge tests indicated the presence of autocorrelation in our data, and the likelihood ratio test for heteroskedasticity was significant. To ensure valid statistical inferences, despite these two violations of the regression model's assumptions, we used clustered-robust standard errors, which controlled for errors both not distributed identically across firms (i.e., heteroskedasticity) and correlated within firms across time periods (i.e., serial correlation) (Petersen, 2009: 465). We therefore computed clustered-robust standard errors at the firm-worldwide level—that is, one level above our unit of analysis (Pepper, 2002)—to correct for the lack of independence of evaluation obtained by the same firm.

We rejected potential issues related to multicollinearity among the explanatory variables by using a variance inflation factor test and regression collinearity diagnostic procedures described by Belsley, Kuh, and Welsch (2004) (respectively ranges between [2.48; 4.08] for the VIF of our independent variables and between [20.30; 24.13] for the condition number across all models, both below the acceptable thresholds). To mitigate the reverse causality issue, all independent variables and control variables were lagged by one year: we collected data from the 2000 editions of the three guides to construct lagged independent variables for the first year of observing our dependent variables (2001).

## RESULTS

Descriptive statistics and pairwise correlations for the variables used are presented in Table 1. Our dataset contains 377 firm-location dyads and 2180 observations over 10 years. Because the variable *categorical consensus* is highly correlated with the dependent

variable *evaluation* (0.62), which might raise concerns about multicollinearity, we ran separate estimations without this control variable: the results obtained are similar to those reported below in Tables 2-4.

[INSERT TABLES 1 & 2 ABOUT HERE]

Model 1 in Table 2 contains controls only. The *categorical consensus* variable is positive and highly significant across models, echoing the findings of previous studies (Hsu, 2006) that find that evaluation is higher when clients have greater consensus regarding firm's practice areas. The non-significance of the variable *categorical contrast* suggests that the average grade of membership of categories does not affect clients' evaluation, most likely since the practice areas of law are discrete and do not overlap. The average *tenure in guides* of the firm's coverage across the three guides positively affects evaluation: that is, the longer the guides have tracked an organization, the higher that organization's evaluation. As expected, clients' evaluation is enhanced by the control variable *size,* which relates to law firm's visibility in the national market. The variable *age* is also positive but does not reach statistical significance for any models. The estimated coefficient for *internationalization degree* is positive and significant, which indicates that clients favor law firms that invest in providing a broader scope of international operations. The *advertisement in guides* variable is non-significant on evaluation, which seems to indicate that a firm's purchase of full-profile advertisement pages in the three guides does not increase its evaluation.

Model 2 includes the independent variable *category spanning* and *inclusive category combination,* and Model 3 adds their joint effect. Regarding controls, coefficient estimates and significance levels are robust compared with those of Model 1, except for internationalization degree that loses significance in Model 3. Hypothesis 1 proposes a

positive association between category spanning and evaluation, and the results support this hypothesis. Firm's *category spanning* has a significant and positive coefficient ($\beta = 0.07$, $p < .001$ in Model 3), which suggests that law firms that span a broader range of categories are more likely to obtain more positive assessments. To illustrate, per Model 3, the predicted average gain of spanning one more category on *evaluation* is +.015, the predicted value of *evaluation* for a firm spanning three categories is .37 while that of a firm spanning seven categories is .44.

The direct effect of *inclusive category combination* is positive and marginally significant in Model 2 and significant in Model 3, suggesting that the inclusiveness of a firm's category combination favors higher evaluation. In Model 3, the interaction effect with *category spanning* is positive and significant, supporting Hypothesis 2. To probe this moderation effect on the full range of observations, we computed and represented graphically the marginal effect of category spanning on evaluation at various levels of the variable *inclusive category combination*. Figure 3 shows that the marginal effect of category spanning on evaluation increases in line with its inclusiveness. This positive moderation corroborates the existence of a theory of value that drives clients to identify unambiguously and evaluate positively those legal service providers with a broad span of activities, the more so as their category combination is inclusive.

[INSERT FIGURE 3 ABOUT HERE]

Model 4 estimates the same effects as well as the direct effect of *evaluation* on performance. While clients positively value organizations spanning multiple categories, the marginally significant ($p < .10$) estimated coefficient of the variable *category spanning* suggests that category spanning has less evident positive economic benefit, consistent with previous research (Hitt, Biermant, Shimizu, & Kochhar, 2001; Kor &

Leblebici, 2005). *Inclusive category combination* fails to impact performance directly but the interaction effect is significant, indicating that category spanners who use inclusive category combinations fare better and generate more revenues. The estimated coefficients of the variables *categorical consensus* and *tenure in guides* are both positive and significant, suggesting that organizations with a clearer category portfolio and longer reported presence experience higher revenues— while *categorical contrast* does not significantly affect firm performance. *Internationalization degree* variable is positive and significant ($\beta = 0.13$, $p < .05$) indicating that firms with multiple offices worldwide correlate with better RPL in the main business hubs of London, NYC, and Paris. *Age* and *advertisement in guides* are found to marginally affect positively performance.

Most notably, we found in Model 4 that *evaluation* has a positive impact on performance ($\beta = 0.94$, $p < .001$; with an improvement in model fit revealed by $\Delta R^2$ and $\Delta\chi^2$ statistics compared to the model without the variable *evaluation*) providing a hint in support of the positive association between evaluation and better performance. In the context of professional service firms where quality is difficult to observe directly, clients rely on indirect indicators (Greenwood et al., 2005), which point to evaluation-related benefits (Benjamin & Podolny, 1999; Uzzi & Lancaster, 2004).

Hypothesis 3 suggests that evaluation should mediate the relationship between category spanning and firm performance, such that category spanning predicts the level of positive evaluation of firms, and evaluation, in turn, affects the performance of firms. Several methods exist for testing the indirect effects in mediation analysis (Edwards & Lambert, 2007; Hayes, 2013; Shaver, 2005). As we are more concerned about Type I errors—that is, claiming an indirect effect exists when it does not—than Type II errors

(i.e., the opposite), we ran both the conservative Sobel test (1982) with the bootstrapping mediation approach (Hayes, 2013: 116; Zhao, Lynch, & Chen, 2010). The Sobel test model estimates the standard error for the product of coefficients in each path of the hypothesized mediation. The equation is $z_{value} = ab/\sqrt{b^2 s_a^2 + a^2 s_b^2}$ where $a$ and $s_a$ are coefficients and standards errors (from Model 3) for the impact of independent variable on mediator, and $b$ and $s_b$ are coefficients and standards errors (from Model 4) for the impact of mediator on the dependent variable. We also used these coefficients in the bootstrapping method, which tends to have higher power than the Sobel test, as no assumption is made about the shape of the sampling distribution of $a \times b$ (Hayes, 2013). The results of the Sobel test and bootstrapping mediation are consistent. The product of coefficients—i.e., the indirect effect— was statistically significant in the expected direction ($a \times b = 0.069$; $SE = 0.021$; $p < .01$). Since zero is also not included in the 95% bias-corrected bootstrap confidence interval ([.1983, .3613], estimated with 10,000 replications), we can conclude that the indirect effect of category spanning on performance is significantly different from zero. Further, computing the ratio of indirect to total effect, we found that the proportion of the total effect of category spanning on performance mediated by evaluation is equal to 37%. This finding is crucial, as it shows category spanning's association with performance through evaluation.

**Robustness checks**

As predicted, our results show that category spanners enjoy higher evaluation. However, this result could be due to our sample's inclusion of both focused firms and category spanners; for our theory to hold, our hypotheses should remain supported when analyzing category spanners only. We therefore ran a separate model that included only category-

spanning firms—those firms that cover at least two practice areas. The results of importance confirm support to Hypotheses 1, 2, and 3 (see Table 3, Models 5 and 6). The estimated coefficient of the direct effect of category spanning on performance appears non-significant, suggesting that the positive and weakly significant corresponding coefficient in Model 4 is mainly driven by the specialist vs. spanners difference in performance, specialists obtaining a lower performance than spanners on average (on average .6 vs. 1.4). We further wondered whether the effects we observed were linear or curvilinear. In unreported models, we tested for curvilinear effects without success: we found categorical spanning had a curvilinear effect neither on evaluation nor on performance. Instead of worldwide revenue, we used worldwide revenue per partner weighted by the relative size of the local branch office (i.e., number of lawyers at the branch office divided by the worldwide number of lawyers) as another measure of performance. This alternative measure accounts for the partners' contribution to revenue generation. We found consistent results.

The two regressions (on evaluation and on performance) may also be related because the errors terms associated with the two dependent variables may be correlated. To provide joint estimates of this system of equations dealing with the correlation issue between the two error terms, we followed the recommended approach by Shaver (2005: 338) and used the random-effects two-stage least squares model (Balestra & Varadharajan-Krishnakumar, 1987). That is, we treated the predicted values of evaluation from the first regression as an endogenous variable and entered those values in the second regression on performance. Table 3 (Models 7 and 8) shows the results, which, for our variables of interest, are again similar to those of the main models as shown in Tables 2.

[INSERT TABLE 3 ABOUT HERE]

**Alternative mechanisms**

Beyond their signaled and objective capacity to handle complex requirements, what other factors could explain why category-spanning organizations receive superior evaluation? One alternative explanation could be that corporate clients may choose to work with only one legal services provider to minimize the transaction costs of searching, bargaining, and policing the exchange (e.g., Chatain, 2011). But there is no evidence for this alternative explanation. First, the results of a 2009 survey of 200 companies from 60 countries (*International In-house Counsel Journal*, 2010: 9) indicate that 56% of the responding clients reported using up to 10 multi-practice law firms. Second, other scholarly work on corporate law firms tends to reject transaction cost-saving and risk-reduction interpretations (e.g., Garicano & Hubbard, 2009). Third, if correct, this mechanism would have been aligned with a strong direct improved performance for category spanners, which we did not find (see Models 4, 6 and 8) in accordance with previous studies (Hitt et al., 2001; Kor & Leblebici, 2005).

As economies of scope do not materialize (since there is no positive effect on performance for spanners), economies of scale may exist. Although the corporate law industry gives no evidence of law firms benefiting from any economies of scale (Harper, 2013; Heinz et al., 2001; Pearlstein, 2004), corporate clients may favor category-spanning law firms for advantages related to size. We tested this explanation by including the interaction term of the two variables *category spanning* and *size* in the additional Models 9 and 10 in Table 4. The results of interest for testing H1-H3 did not change, and we found a positive estimated coefficient on evaluation (Model 9: $\beta = 0.01$, $p < .01$; Figure 4), indicating that law firms that are both large and multi-practice absorb

uncertainties about legal cases, making such firms more appealing to clients with complex requirements. In these additional models, the main effects keep providing further evidence in favor of our hypotheses.

Finally, category spanning could be a strategy adopted by top corporate law firms, and spanning could conflate a true organizational capacity to answer clients' requirements and a legitimacy effect. To disentangle this potential conflation, we calculated the sum of the absolute differences between each individual firm's vector of the presence or absence of a practice (8-uplet vector of 0/1 values) and the average of the 10 most highly evaluated firms' practice portfolios in each location (8-uplet vector of $\bar{x}$, x being the presence (1) or absence (0) of the practice): the greater the distance, the less a focal firm's category spanning resembled that of the top 10 firms. The results of interest for testing H1-H3 did not change, and we found a negative and significant effect on both evaluation and performance (Models 11 and 12). Thus, quite logically, clients tend to refer to the highest-ranked firms' portfolio to assess a focal firm's portfolio and more distant firms from the top 10 fare less well in terms of revenue per lawyer.

[INSERT TABLE 4 & FIGURE 4 ABOUT HERE]

## DISCUSSION

In questioning the dominant wisdom that category spanning is detrimental to an organization, this paper departs from past studies in two main ways. First, by accounting for audiences' theory of value, it proposes to stretch the "categorical imperative" (Durand & Paolella, 2013; Zuckerman, 1999), and opens the "golden cage of categorical membership norms within which organizations are held" (Durand & Kremp, 2015: 38). The core of past research hinges on prototypical distance, which affixes organizations to prototypes, and fixes audiences and their expectations. We relax this core assumption and

posit that audiences have diverse theories of value depending on their requirements, some anchored in crisp prototypes (e.g., for well-established consumption products), whereas others evade typecasting (for more complex requirements). We argue that when issues are complex, i.e., sophisticated, singular, and high stake, clients' theory of value passes from type- to goal-based. Hence, clients assess a producer as a whole entity as potentially addressing their complex requirements before whether its products taken in isolation meet an ideal type.

Second, most past research has ignored the concurrent effects of category spanning on evaluation and performance and the possibility of a mediated effect between category spanning and performance through evaluation. While category spanning can transfer or not into better performance–depending on which theory of value materializes in given contexts –evaluation will mediate the relationship between category spanning and performance such that the impact on performance will depend on whether evaluation of spanning is positive or negative. We therefore connect sparse literatures and prior findings in a more coherent set of relationships and open new avenues for research.

Exploring the corporate law market, we found (at the branch-office level) that category spanners obtain better evaluation than more focused firms and that evaluation mediates partially the direct influence of category spanning on performance. Before potential clients consider firms' products, they value firms for the coherence of their categorical identity, which affects audience evaluation and indirectly performance, 37% of revenues being explained by the mediation. We found complementary support for this theory in the moderating effect of *inclusive category combination* on the category spanning–evaluation relationship: not all categories and not all configurations of categories

benefit producers' evaluation identically–more frequently interlocked combinations at various degrees (duos, trios, n-uplets) being more influential in reinforcing the relationship between spanning and evaluation. These findings enrich our understanding of how audiences apprehend markets' category systems and category spanners.

**Implications for category research**

Some prior works find evidence that category spanning is acceptable in particular settings. For instance, when categories are in flux or when spanning helps mitigate value conflicts, the spanners can outnumber purer players, be more accepted and even rewarded (Ruef & Patterson, 2009, Alexy & George, 2013; Vergne, 2012). Our general approach helps recast prior studies in a broader context. Following our work, audiences' theory of value is type- or goal-based depending on what audiences need and seek. For more standard demands, according to prototype-based view, a product's features contain enough information and act as stimuli to which an audience responds by comparing current features to an ideal type (Mervis & Rosch, 1981). For more complex requests, when for instance the audience pursues goals such as "fixing the legal issues of my last acquisition", those organizations that meet these sophisticated, singular, and high stake requests will be deemed more attractive (Barsalou, 1991; Durand & Paolella, 2013). Hence, in those situations, categorization does not operate on a similarity basis, but rather according to criteria that coincide with case-by-case expectations. Therefore, a corollary of our findings would be that, as a theory of value involves less sophisticated, less recurrent, and less expensive conditions the prototypical view will tend to prevail over the goal-based view of categorization at the product level. As the clients demands become more complex, audiences consider producers as more or less coherent entities able to respond to their

goals, leading them to value more highly category spanners, and to a partial mediation by evaluation of the direct association between category spanning and performance.

We found also evidence that audiences do not mobilize analytical distance assessments independently of a category system. Category combination's inclusiveness matters to understand both evaluation and performance. As the theory of value involves more complex requirements, the interlocking of category co-occurrences at multiple degrees of potential combinations (2, 3, 4 and more degrees depending on the number of categories present in the market) positively moderates the association between category spanning and evaluation. Note that we found that spanners with inclusive combinations obtain better revenue per lawyer as well. These findings bring evidence that beyond distance from product types, taken in isolation or even aggregated over at the producer level, sedimentation of co-occurring category associations within a market is related to an audience's evaluation of a producer (and to its performance). Not just horizontal distances from types matter but also the vertical layering of frequently joint categories that are more salient and make more sense for audiences. Broader and more inclusive category spanning is viewed positively, as indicating "the skill or ability to produce a wide variety of situated conceptualizations that support goal achievement in specific contexts" (Barsalou, 2005: 626). It is likely that even for more standard requirements, where the type-based view predominates, inclusiveness will have a positive moderation on the category spanning-evaluation relationship, i.e., will reduce the penalty borne by category spanners.

Hence, this paper helps identify and reconcile the different validity domains of past research. It enriches our understanding of the valuation conditions that precede any evaluation in organizational and market contexts (Lamont, 2012; Zuckerman, 1999).

**Implications for diversification research**

Producers are assessed based on the audiences' theory of value and on the signals they project, in terms of both their range of expertise and coherence among the spanned categories. Thus, independent of the economic or strategic rationale related to diversification benefits, in markets as the one we studied, category spanners are likely to gain better evaluation, which transfers to better performance and counterbalances the category-spanning's potential penalties associated with this positioning. As such, this paper contributes to the study of diversification by suggesting that the relatedness between activities can take different forms: not only strategic (rent-accruing through economies of scope) and economic (reducing unit costs by economies of scale), as seen in past research, but also sociocognitive in the cases of category spanning, as in the context we have studied. As stated by Durand and Paolella (2013: 1112): "It is not the fact of spanning categories *per se* (i.e., increasing the total cognitive distance relative to established prototypes) that might matter to audiences, but their capacity to make coherent sense of the categorical combinations they observe". Therefore, even in absence of economies of scale and scope, producers continue to span categories to meet audiences' theory of value and then increase their revenues, the more so when they obtain higher evaluation and when their category combination is inclusive. Producers should therefore take audiences' theory of value into account when defining their category-spanning strategies, both in absolute terms (by accepting more or fewer categories of activities) and in relative terms (as based on configurations' inclusiveness).

The limitations of our paper suggest promising opportunities for future research. First, some data limitations stopped us in our investigation. For example, as we lacked

the details of contracts and deals, we needed to consider clients in a homogenous way, although it may well be important to tease out the significance of clients' heterogeneity on their appreciation of category spanners (i.e., in terms of their variations in theories of value). Also, more fine-grained measures of category spanning could be used, particularly weighted measures of spanning that would account for the respective value of each practice in a firm's portfolio. Unfortunately, we lacked the data with which to calculate this variable as we lacked the data on branch-office profits and we had to consider that RPL is equally distributed over a firm's branch offices. With our setting however, we could not test concurrently the two theories of value and leave it to future research. Probing further audiences' theories of value in different settings that include both standard and complex requirements will stimulate theory building on category order and valuation at different levels of analysis (i.e., products/services, firms, and industry), and help determine the strength of the mediation effect through evaluation under different theory-of-value conditions.

Second, although our paper is longitudinal, it does not isolate the emergence of new categories in firms' portfolios. By shifting interest from prototypes to audiences' theory of value, it prompts for more studies on changes in audiences' theory of value (Zuckerman & Rao, 2004: 236). The underlying reasons for an audience to elaborate a more or less complex theory of value and for a market being categorized in one way or another depend chiefly on cultural, social, and institutional factors, none of which we explore. Those dimensions and their associations with history and traditions matter significantly more than is currently acknowledged in explaining both audiences' and producers' categorizations and conformity behaviors (Durand & Kremp, 2015; Kennedy

& Fiss, 2013; Smith, 2011). Here, new avenues for research exist to better connect the existing sociological and ecological work on categories with other institutional, social, and cultural approaches that aim to describe and understand not just the discipline involved in markets' sociocognitive infrastructures but also the networks of meaning that emerge, propagate, and reproduce.

More generally, explicitly recognizing the sociocultural foundations of theories of value and category systems opens opportunities to better understand the dynamic and long-term evolution of markets. Together, these future works will complement both extant studies and this study. They will move us further toward a more refined understanding of category spanning, and more broadly toward a better comprehension of the causes and consequences of organizational deviance and conformity relative to audiences' expectations.

## REFERENCES

Alexy, O., & George, G. 2013. Category divergence, straddling, and currency: Open innovation and the legitimation of illegitimate categories. ***Journal of Management Studies***, 50: 173–203.

Balestra, P., & Varadharajan-Krishnakumar, J. 1987. Full information estimations of a system of simultaneous equations with error component structure. ***Econometric Theory***, 3: 223–246.

Barsalou, L. W. 1991. Deriving categories to achieve goals. In G. H. Bower (Ed.), ***The psychology of learning and motivation: Advances in research and theory***, Vol. 27: 1–64. San Diego, CA: Academic Press.

Barsalou, L. W. 2005. Situated conceptualization. In H. Cohen, & C. Lefebvre (Eds.), ***Handbook of categorization in cognitive science***, 619–650. Oxford:UK: Elsevier.

Belsley, D. A., Kuh, E., & Welsch, R. E. 2004. ***Regression diagnostics: Identifying influential data and sources of collinearity***. Hoboken, NJ: John Wiley & Sons.

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

Cameron, A. C., & Trivedi, P. K. 2010. ***Microeconometrics using stata***. Lakeway Drive, College Station, TX: Stata Press

***The Chambers and Partners UK***. 2010. London: UK: Chambers and Partners Publishing.

Chatain, O. 2011. Value creation, competition, and performance in buyer□supplier relationships. ***Strategic Management Journal***, 32: 76–102.

Coates, J. C., DeStefano, M. M., Nanda, A., & Wilkins, D. B. 2011. Hiring teams, firms, and lawyers: Evidence of the evolving relationships in the corporate legal market. ***Law and Social Inquiry-Journal of the American Bar Foundation***, 36: 999–1031.

Dezalay, Y., & Garth, B. G. 2004. The confrontation between the Big Five and Big Law: Turf battles and ethical debates as contests for professional credibility. ***Law & Social Inquiry***, 29: 615–638.

Durand, R., & Kremp, P.-A. Classical Deviation: Organizational and Individual Status as Antecedents of Conformity. ***Academy of Management Journal***: Forthcoming.

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

Durkheim, E. 2008. ***The elementary forms of the religious life*** (J. W. Swain, Trans.). Mineola, NY: Dover Publications.

Edwards, J. R., & Lambert, L. S. 2007. Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. ***Psychological Methods***, 12: 1–22.

Galanter, M., & Palay, T. 1994. ***Tournament of lawyers: The transformation of the big law firm***. Chicago, IL: University of Chicago Press.

Garicano, L., & Hubbard, T. N. 2009. Specialization, firms, and markets: The division of labor within and between law firms. ***Journal of Law, Economics, and Organization***, 25: 339–371.

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

Greenwood, R., Li, S. X., Prakash, R., & Deephouse, D. L. 2005. Reputation, diversification, and organizational explanations of performance in professional service firms. ***Organization Science***, 16: 661–673.

Hannan, M. T. 2010. Partiality of Memberships in Categories and Audiences. ***Annual Review of Sociology***, 36: 159-181.

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

Harper, S. J. 2013. ***The Lawyer bubble. A Profession in crisis***. New York, NY: Basic Books.

Hayes, A. F. 2013. ***Introduction to mediation, moderation, and conditional process analysis: A regression-based approach***. New York, NY: Guilford Press.

Heinz, J. P., Nelson, R. L., & Laumann, E. O. 2001. The scale of justice: Observations on the transformation of urban law practice. ***Annual Review of Sociology***: 337–362.

Hitt, M. A., Biermant, L., Shimizu, K., & Kochhar, R. 2001. Direct and moderating effects of human capital on strategy and performance in professional service firms: A resource-based perspective. ***Academy of Management Journal***, 44: 13–28.

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., Koçak, O., & Hannan, M. T. 2009. Multiple category memberships in markets: An integrative theory and two empirical tests. ***American Sociological Review***, 74: 150–169.

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

***International In-house Counsel Journal***. 2010. Agenda For Action 2010. Annual Survey Report.

Jones, G. R., & Hill, C. W. 1988. Transaction cost analysis of strategy□structure choice. ***Strategic Management Journal***, 9: 159–172.

Karpik, L. 2010. ***Valuing the unique: The economics of singularities*** (N. Scott, Trans.). Princeton, NJ: Princeton Univ Press.

Kennedy, M. T., & Fiss, P. C. 2013. An ontological turn in categories research: From standards of legitimacy to evidence of actuality. ***Journal of Management Studies***, 50: 1138–1154.

Kennedy, M. T., Lo, J., & Lounsbury, M. 2010. Category currency: The changing value of conformity as a function of ongoing meaning construction. In G. Hsu, O. Kocak, & G. Negro (Eds.), ***Categories in markets: Origins and evolution. Research in the Sociology of Organizations***, Vol. 31: 369–397. Bingley, UK: Emerald.

Kim, B. K., & Jensen, M. 2011. How product order affects market identity repertoire ordering in the US opera market. ***Administrative Science Quarterly***, 56: 238–256.

Kor, Y. Y., & Leblebici, H. 2005. How do interdependencies among human□capital deployment, development, and diversification strategies affect firms' financial performance? ***Strategic Management Journal***, 26: 967–985.

Kovács, B., & Hannan, M. T. 2010. The consequences of category spanning depend on contrast. In G. Hsu, O. Kocak, & G. Negro (Eds.), ***Categories in markets: Origins and evolution. Research in the Sociology of Organizations***, Vol. 31: 175–201. Bingley, UK: Emerald.

Kovács, B., & Johnson, R. 2014. Contrasting alternative explanations for the consequences of category spanning: A study of restaurant reviews and menus in San Francisco. ***Strategic Organization***, 12: 7–37.

Lamont, M. 2012. Toward a comparative sociology of valuation and evaluation. ***Annual Review of Sociology***, 38: 201–221.

Leung, M. D. 2014. Dilettante or Renaissance Person? How the Order of Job Experiences Affects Hiring in an External Labor Market. ***American Sociological Review***, 79: 136-158.

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

Lin, E. L., & Murphy, G. L. 2001. Thematic relations in adults' concepts. ***Journal of Experimental Psychology General***, 130: 3–28.

Lo, J. Y.-C., & Kennedy, M. T. 2015. Approval in Nanotechnology Patents: Micro and Macro Factors That Affect Reactions to Category Blending. ***Organization Science***, 26: 119-139.

Malos, S. B., & Campion, M. A. 2000. Human resource strategy and career mobility in professional service firms: A test of an options-based model. ***Academy of Management Journal***, 43: 749-760.

Mervis, C. B., & Rosch, E. 1981. Categorization of natural objects. ***Annual Review of Psychology***, Vol. 32: 89–115.

Meyers-Levy, J., & Tybout, A. M. 1989. Schema congruity as a basis for product evaluation. ***Journal of Consumer Research***, 16: 39–54.

Murphy, G. L. 2004. ***The big book of concepts***. Cambridge, MA: The MIT Press.

Negro, G., Hannan, M. T., & Rao, H. 2010. Categorical contrast and audience appeal: Niche width and critical success in winemaking. ***Industrial and Corporate Change***, 19: 1397–1425.

Negro, G., & Leung, M. D. 2013. "Actual" and perceptual effects of category spanning. ***Organization Science***, 24: 684–696.

Pearlstein, S. 2004. Scaling the myth of merger's efficiencies. ***The Washington Post***, April 21.

Pepper, J. V. 2002. Robust inferences from random clustered samples: An application using data from the panel study of income dynamics. ***Economics Letters***, 75: 341–345.

Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. ***Review of Financial Studies***, 22: 435–480.

Phillips, D. J., Turco, C., & Zuckerman, E. W. 2013. Betrayal as market barrier: identity-based limits to diversification among high-status corporate law firms. ***American Journal of Sociology***, 118: 1–32.

Podolny, J. M. 1993. A status-based model of market competition. ***American Journal of Sociology***, 98: 829–872.

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

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.

Ratneshwar, S., Barsalou, L. W., Pechmann, C., & Moore, M. 2001. Goal-derived categories: The role of personal and situational goals in category representations. ***Journal of Consumer Psychology***, 10: 147–157.

Rindova, V. P., Dalpiaz, E., & Ravasi, D. 2011. A cultural quest: A study of organizational use of new cultural resources in strategy formation. ***Organization Science***, 22: 413–431.

Rosa, J. A., Judson, K. M., & Porac, J. F. 2005. On the sociocognitive dynamics between categories and product models in mature markets. ***Journal of Business Research***, 58: 62–69.

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

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

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

Shaver, J. M. 2005. Testing for mediating variables in management research: Concerns, implications, and alternative strategies. ***Journal of Management***, 31: 330–353.

Shiller, R. J. 1990. Speculative prices and popular models. ***The Journal of Economic Perspectives***, 4: 55-65.

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

Sobel, M. E. 1982. Asymptotic confidence intervals for indirect effects in structural equation models. ***Sociological Methodology***, 13: 290–312.

Stimpert, J. L., & Duhaime, I. M. 1997. Seeing the big picture: The influence of industry, diversification, and business strategy on performance. ***Academy of Management Journal***, 40: 560–583.

Uzzi, B., & Lancaster, R. 2004. Embeddedness and price formation in the corporate law market. ***American Sociological Review***, 69: 319–344.

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

Wilkins, D. B. 2009. Team of rivals. Toward a new model of the corporate attorney-client relationship. ***Fordham Law Review***, 78: 2067–2136.

Wry, T., Lounsbury, M., & Jennings, P. D. 2014. Hybrid vigor: Securing venture capital by spanning categories in nanotechnology. ***Academy of Management Journal***, 57: 1309–1333.

Zhao, X., Lynch, J. G., & Chen, Q. 2010. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. ***Journal of Consumer Research***, 37: 197–206.

Zuckerman, E. W. 1999. The Categorical imperative: Securities znalysts and the illegitimacy discount. ***American Journal of Sociology***, 104: 1398–1438.

Zuckerman, E. W., Kim, T.-Y., Ukanwa, K., & Rittmann, J. v. 2003. Robust identities or nonentities? Typecasting in the feature-film labor market. ***American Journal of Sociology***, 108: 1018–1074.

Zuckerman, E. W., & Rao, H. 2004. Shrewd, crude or simply deluded? Comovement and the internet stock phenomenon. ***Industrial and Corporate Change***, 13: 171-212.

**TABLE 1**
**Descriptive Statistics and Pairwise Correlations**

| | Variables | Mean | SD | Min. | Max. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | Performance | 1.25 | 1.51 | 0.01 | 10.04 | | | | | | | | | | |
| 2. | Evaluation | 0.40 | 0.19 | 0.07 | 1 | 0.47 | | | | | | | | | |
| 3. | Category spanning | 0.44 | 0.32 | 0 | 0.88 | 0.41 | 0.31 | | | | | | | | |
| 4. | Inclusive category combination | 3.39 | 1.34 | 0 | 5.63 | 0.39 | 0.33 | 0.49 | | | | | | | |
| 5. | Categorical consensus | 0.42 | 0.38 | 0 | 1 | 0.44 | 0.62 | 0.39 | 0.43 | | | | | | |
| 6. | Categorical contrast | 0.46 | 0.08 | 0.25 | 0.80 | -0.27 | 0.03 | -0.37 | -0.20 | -0.11 | | | | | |
| 7. | Tenure in guides | 4.17 | 2.51 | 1 | 10 | 0.26 | 0.28 | 0.38 | 0.35 | 0.34 | -0.21 | | | | |
| 8. | Size | 3.50 | 0.96 | 0 | 6.50 | 0.48 | 0.11 | 0.31 | 0.33 | 0.17 | -0.23 | 0.12 | | | |
| 9. | Age | 49.23 | 56.46 | 0 | 308 | 0.30 | 0.10 | 0.19 | 0.15 | 0.14 | -0.17 | 0.15 | 0.31 | | |
| 10. | Internationalization degree | 1.27 | 1.07 | 0 | 3.66 | 0.31 | 0.22 | 0.40 | 0.32 | 0.30 | -0.23 | 0.12 | 0.08 | -0.01 | |
| 11. | Advertisement in guides | 0.67 | 0.55 | 0 | 3.25 | 0.35 | 0.24 | 0.32 | 0.21 | 0.26 | -0.25 | 0.01 | 0.26 | 0.20 | 0.25 |

**TABLE 2**
**Random-Effects GLS Estimations: Effects on Evaluation and Performance**

| Variables | Evaluation | | | Performance |
|---|---|---|---|---|
| | Model 1: Control | Model 2 | Model 3: H1-H2 | Model 4: H3 |
| Evaluation | | | | 0.94*** |
| | | | | (0.169) |
| Category spanning | | 0.06*** | 0.07*** | 0.11+ |
| | | (0.017) | (0.019) | (0.051) |
| Inclusive category combination | | 0.01+ | 0.01* | -0.01 |
| | | (0.003) | (0.003) | (0.010) |
| Spanning ×Inclusive category combination | | | 0.01** | 0.10*** |
| | | | (0.005) | (0.023) |
| Categorical consensus | 0.11*** | 0.10*** | 0.10*** | 0.16** |
| | (0.014) | (0.013) | (0.013) | (0.056) |
| Categorical contrast | 0.01 | 0.03 | 0.07 | 0.27 |
| | (0.069) | (0.072) | (0.077) | (0.209) |
| Tenure in guides | 0.02*** | 0.02*** | 0.02*** | 0.07*** |
| | (0.003) | (0.003) | (0.003) | (0.013) |
| Size | 0.02** | 0.01** | 0.01* | 0.05 |
| | (0.006) | (0.005) | (0.005) | (0.032) |
| Age | 0.001 | 0.001 | 0.001 | 0.002+ |
| | (0.000) | (0.000) | (0.000) | (0.001) |
| Internationalization degree | 0.02* | 0.01+ | 0.01 | 0.13* |
| | (0.007) | (0.007) | (0.007) | (0.055) |
| Advertisement in guides | 0.01 | 0.01 | 0.01 | 0.09+ |
| | (0.007) | (0.007) | (0.007) | (0.048) |
| Constant | 0.17** | 0.16* | 0.15* | -1.71*** |
| | (0.050) | (0.051) | (0.052) | (0.275) |
| Category dummies | Yes | Yes | Yes | Yes |
| Nationality dummies | Yes | Yes | Yes | Yes |
| Guide dummies | Yes | Yes | Yes | Yes |
| Year dummies | Yes | Yes | Yes | Yes |
| Location area dummies | Yes | Yes | Yes | Yes |
| Observations | 2180 | 2180 | 2180 | 2180 |
| Firm-location | 377 | 377 | 377 | 377 |
| Firm-worldwide | 288 | 288 | 288 | 288 |
| $R^2$ | 0.53 | 0.54 | 0.55 | 0.61 |
| | | | | (Δ: .04) |
| $\chi^2$ | 606.44 | 641.74 | 646.89 | 739.31 |
| | | | | (Δ: 84.66) |

*Note:* Clustered-robust standard errors are shown in parentheses. In model 4, Δ refers to goodness of fit measures of Model 4 without evaluation as an independent variable.

+ $p < .10$
* $p < .05$
** $p < .01$
*** $p < .001$

**TABLE 3**
**Robustness Checks**

| Variables | Spanners only | | G2SLS Estimations | |
|---|---|---|---|---|
| | **Model 5: Evaluation** | **Model 6: Performance** | **Model 7: Evaluation** | **Model 8: Performance** |
| Evaluation | | 1.69*** | | 1.49* |
| | | (0.254) | | (0.574) |
| Category spanning | 0.05* | -0.13 | 0.07*** | 0.07 |
| | (0.024) | (0.094) | (0.012) | (0.083) |
| Inclusive category combination | 0.004 | 0.01 | 0.01+ | 0.01 |
| | (0.003) | (0.013) | (0.002) | (0.012) |
| Spanning × Inclusive category combination | 0.01** | 0.05* | 0.01*** | 0.10*** |
| | (0.004) | (0.024) | (0.003) | (0.020) |
| Categorical consensus | 0.11*** | 0.25** | 0.08*** | 0.17* |
| | (0.015) | (0.078) | (0.013) | (0.071) |
| Categorical contrast | 0.14 | 0.30 | 0.05 | 0.24 |
| | (0.093) | (0.331) | (0.041) | (0.216) |
| Tenure in guides | 0.02*** | 0.08*** | 0.02*** | 0.06*** |
| | (0.003) | (0.019) | (0.002) | (0.006) |
| Size | 0.01 | 0.13* | 0.01* | 0.12*** |
| | (0.006) | (0.056) | (0.005) | (0.024) |
| Age | 0.001 | 0.001 | 0.001 | 0.002*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Internationalization degree | 0.02* | 0.10+ | 0.01 | 0.10* |
| | (0.007) | (0.062) | (0.007) | (0.034) |
| Advertisement in guides | 0.01 | 0.10+ | 0.01 | 0.11*** |
| | (0.006) | (0.060) | (0.005) | (0.026) |
| Constant | 0.08* | -2.66*** | 0.19*** | -0.59 |
| | (0.063) | (0.371) | (0.051) | (0.210) |
| Category dummies | Yes | Yes | Yes | Yes |
| Nationality dummies | Yes | Yes | Yes | Yes |
| Guide dummies | Yes | Yes | Yes | Yes |
| Year dummies | Yes | Yes | Yes | Yes |
| Location area dummies | Yes | Yes | Yes | Yes |
| Observations | 1592 | 1592 | 2180 | 2180 |
| Number of firm-location | 291 | 291 | 377 | 377 |
| Number of firm-worldwide | 217 | 217 | 288 | 288 |
| $R^2$ | 0.59 | 0.67 | 0.55 | 0.59 |
| $\chi^2$ | 532.52 | 471.71 | 663 | 983.05 |

*Note:* Clustered-robust standard errors are shown in parentheses.
+ $p < .10$
* $p < .05$
** $p < .01$
*** $p < .001$

**TABLE 4**
**Supplementary Analyses for Alternative Mechanisms**

| Variables | Scale effect | | Top 10 distance effect | |
|---|---|---|---|---|
| | Model 9: Evaluation | Model 10: Performance | Model 11: Evaluation | Model 12: Performance |
| Category spanning ×Size | 0.01* | 0.15*** | | |
| | (0.005) | (0.028) | | |
| Distance top 10 firms | | | -0.06** | -0.51*** |
| | | | (0.020) | (0.088) |
| Evaluation | | 0.94*** | | 0.88*** |
| | | (0.170) | | (0.165) |
| Category spanning | 0.07*** | 0.04 | 0.07*** | 0.08 |
| | (0.019) | (0.057) | (0.018) | (0.053) |
| Inclusive category combination | 0.01+ | -0.02 | 0.01* | -0.01 |
| | (0.003) | (0.011) | (0.003) | (0.010) |
| Spanning × Inclusive category combination | 0.01** | 0.08*** | 0.01** | 0.09*** |
| | (0.005) | (0.022) | (0.004) | (0.021) |
| Categorical consensus | 0.10*** | 0.15** | 0.10*** | 0.17** |
| | (0.013) | (0.055) | (0.013) | (0.053) |
| Categorical contrast | 0.07 | 0.37 | 0.06 | 0.28 |
| | (0.076) | (0.208) | (0.076) | (0.208) |
| Tenure in guides | 0.02*** | 0.06*** | 0.02*** | 0.06*** |
| | (0.003) | (0.013) | (0.003) | (0.012) |
| Size | 0.02** | 0.13*** | 0.01* | 0.04 |
| | (0.006) | (0.039) | (0.005) | (0.032) |
| Age | 0.001 | 0.002+ | 0.001 | 0.002+ |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Internationalization degree | 0.01 | 0.12* | 0.01 | 0.12* |
| | (0.007) | (0.054) | (0.007) | (0.054) |
| Advertisement in guides | 0.01 | 0.09+ | 0.01 | 0.11* |
| | (0.007) | (0.047) | (0.007) | (0.048) |
| Constant | 0.14** | -1.84*** | 0.19*** | -1.35*** |
| | (0.051) | (0.258) | (0.050) | (0.250) |
| Category dummies | Yes | Yes | Yes | Yes |
| Nationality dummies | Yes | Yes | Yes | Yes |
| Guide dummies | Yes | Yes | Yes | Yes |
| Year dummies | Yes | Yes | Yes | Yes |
| Location area dummies | Yes | Yes | Yes | Yes |
| Observations | 2180 | 2180 | 2180 | 2180 |
| Number of firm-location | 377 | 377 | 377 | 377 |
| Number of firm-worldwide | 288 | 288 | 288 | 288 |
| $R^2$ | 0.56 | 0.65 | 0.57 | 0.63 |
| $\chi^2$ | 703.37 | 894.42 | 646.02 | 808.80 |

*Note:* Clustered-robust standard errors are shown in parentheses.
+ $p < .10$
* $p < .05$
** $p < .01$
*** $p < .001$

**FIGURE 1**

**Level of Average Category Spanning over Observation Period**

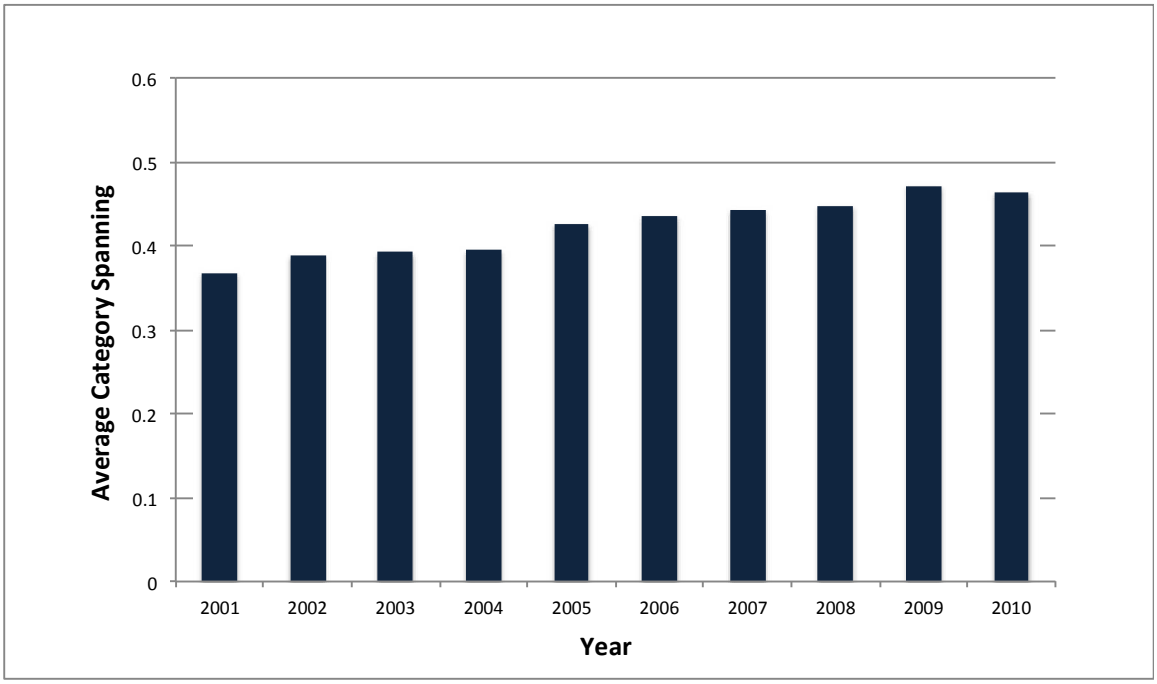



**FIGURE 2**

**Examples of Inclusive-Category Combinations**

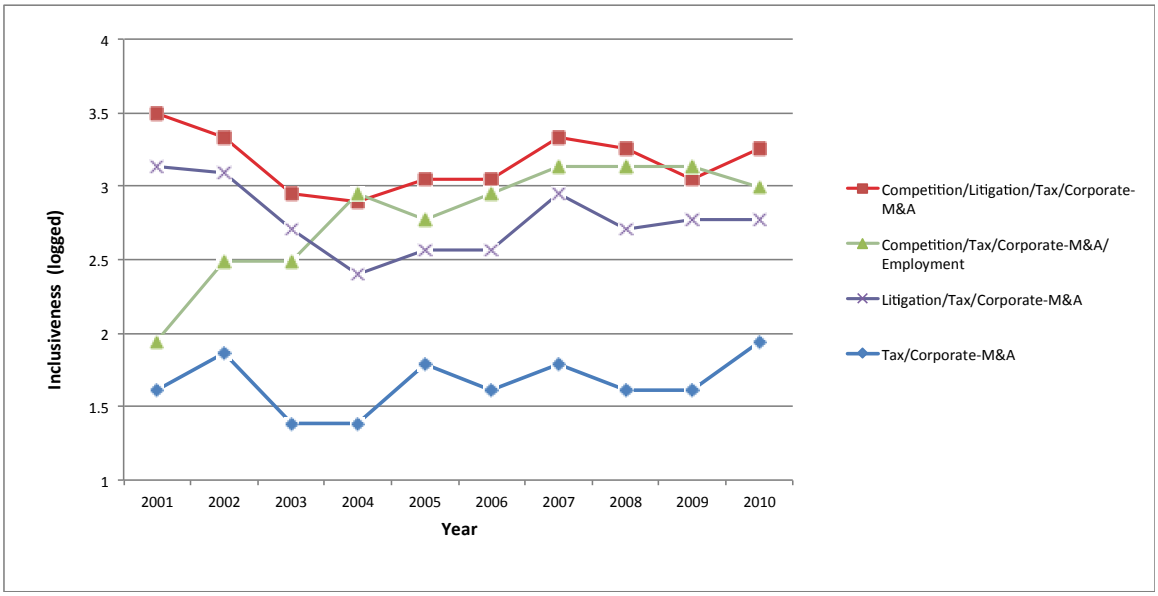


**FIGURE 3**
**Marginal Effect of Category Spanning on Evaluation Conditional on Inclusive-Category Combination**

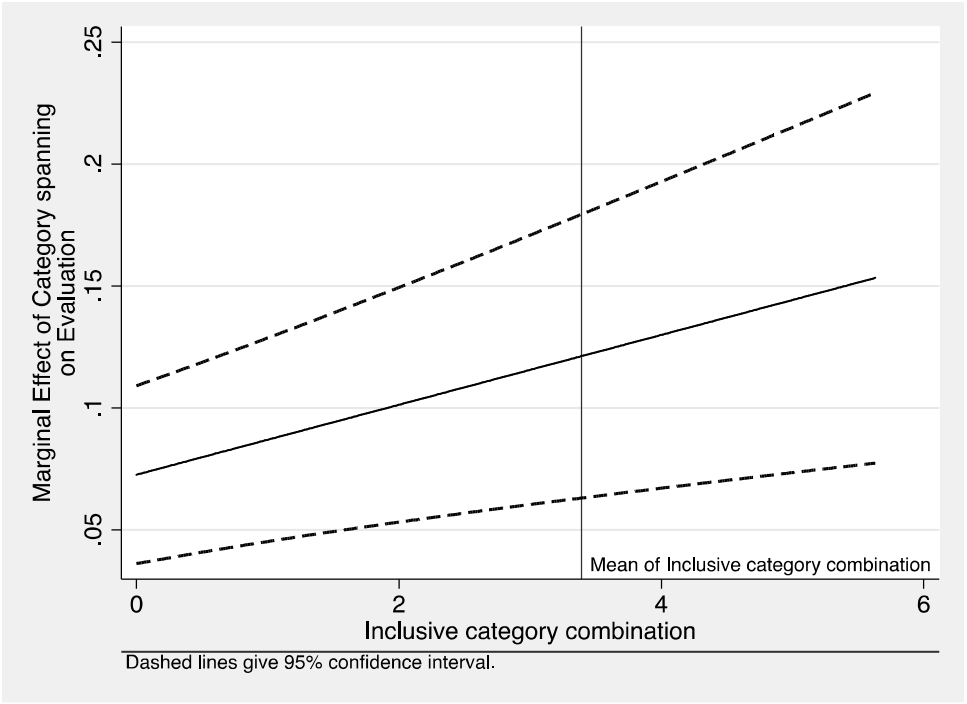
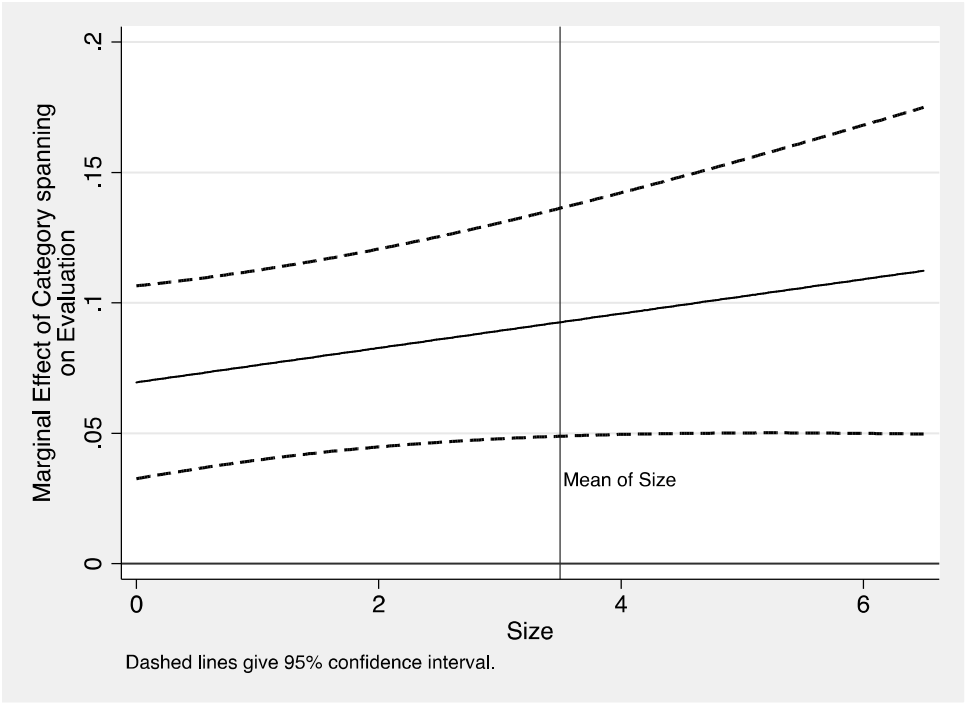



**FIGURE 4**
**Marginal Effect of Category Spanning on Evaluation Conditional on Size**



**Lionel Paolella** (l.paolella@jbs.cam.ac.uk) is an Assistant Professor of Strategy and Organization at the Cambridge Judge Business School, University of Cambridge, and affiliated with the Center on the Legal Profession at Harvard Law School. He received his Ph.D. from HEC Paris in 2014. His main line of work explores the categorization processes in markets.

**Rodolphe Durand** (durand@hec.fr) is Professor of Strategy at HEC Paris, from where he received his Ph.D. His primary research interests concern the social and institutional sources of competitive advantage and organizational performance.