Explaining the Variability of Audiences' Valuations : An Approach Based on Market Categories and Natural Language Processing

Paul Gouvard

INSTITUT POLYTECHNIQUE DE PARIS

HEC PARIS

NNT : 2019IPP000

Thèse de doctorat

Explaining the Variability of Audiences' Valuations: An Approach Based on Market Categories and Natural Language Processing

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à HEC Paris

École doctorale de l'Institut Polytechnique de Paris – (ED IP Paris) n°626
Spécialité de doctorat: Sciences de Gestion

Thèse présentée et soutenue à Jouy-en-Josas, le 01/07/2020 par

GOUVARD Paul

Composition du Jury :

Mary-Ann Glynn Professeur titulaire, Caroll School of Management, Boston College	Présidente du jury
Sameer Srivastava Professeur associé, Haas Business School, UC Berkeley	Rapporteur
Rose Luo Professeur associé, INSEAD	Examinateur
Rodolphe Durand Professeur titulaire, HEC Paris	Directeur de thèse

REMERCIEMENTS

Je tiens à remercier tous ceux qui, de près ou de loin, m'ont aidé, accompagné et soutenu dans l'achèvement de cette thèse au cours des cinq années qui viennent de s'écouler. Je souhaiterais notamment remercier :

Les membres du jury, pour avoir pris le temps de lire ce travail imparfait,

Marieke Huysentruyt, pour m'avoir impliqué au sein d'un projet de recherche au cours duquel j'ai beaucoup appris et pour la qualité de nos échanges,

Sameer Srivastava et Amir Goldberg, pour m'avoir accueilli au sein du Computational Culture Lab, pour avoir partagé leurs ressources avec moi, pour leurs précieux conseils et leur aide lors de mon séjour à Berkeley,

Arnaud et Laure, pour tous les moments passés ensembles et tous ceux à venir,

Tous les doctorants d'HEC, pour leur bonne humeur et leur camaraderie,

Toute l'équipe du doctorat d'HEC, pour toute l'aide apportée au cours des cinq dernières années,

Le département stratégie d'HEC et tous ses membres, pour leur soutien, leur bienveillance et leurs encouragements,

SNO, pour tous nos déjeuners riches en découvertes,

Le GREGHEC, pour avoir financé une partie de mon échange à Berkeley, et Corinne Tessier, pour m'avoir aidé dans mes démarches et attendu mon retour à HEC immédiatement après ma descente d'avion,

L'ensemble de la communauté d'HEC Paris, pour cinq excellentes années en son sein,

L'université Berkeley, pour m'avoir reçu l'espace d'un semestre riche en apprentissage en 2018,

Tous les membres du Computational Culture Lab, pour leur écoute et leur bienveillance,

Je tiens plus spécialement à exprimer toute ma gratitude et ma reconnaissance à :

Rodolphe, pour avoir accepté de me rencontrer quand je l'ai contacté 'out-of-the-blue' en 2015, pour la richesse de nos entretiens, son constant soutien, sa bonne humeur et son enthousiasme, son humanisme et sa passion pour la recherche,

Mes amis, qui ont toujours été là pour m'encourager quand je perdais la foi et pour me distraire quand j'avais besoin d'ailleurs,

Ma famille, sans laquelle je ne serais pas l'éternel chercheur que je suis,

Sylvain, pour tout,

TABLE OF CONTENT

REMERCIEMENTS 2
LIST OF FIGURES 4
LIST OF TABLES 5
INTRODUCTION 6
1. The prototype-based model of valuation and the stability of audiences' valuations 9
1.1. Origins in cognitive psychology 9
1.2. The prototype-based model of valuation in organization studies 10
2. Exploring the variability of audiences' valuation 11
2.1. The variability of audiences' valuations and its importance for research on categories 11
2.1.1. The variability of audiences' valuations as an empirical fact 12
2.1.2. The heterogeneity of audiences within the prototype-based model of valuation 15
2.1.3. The multiplicity of audiences' modes of valuation 17
2.2. Main research question and three research gaps 20
2.2.1. Main research question 20
2.2.2. Three research gaps 20
3. Introduction to the three essays 23
3.1. Essay 1 - The (relative) effects of typicality on volatility: A study using word embeddings 23
3.2. Essay 2 - Organizational appeal and market valuation: A natural language processing study of IPO first-day returns 24
3.3. Essay 3 - Valuing organizations: An integrated theory 25
4. Methodology and setting 26
4.1. NLP methods as a tool to inform category research 26
4.2. Publicly listed firms in the U.S. and their relevance to study audiences' valuation using NLP 28
5. Overall structure of the dissertation 29
CHAPTER 1 37
CHAPTER 2 75
CHAPTER 3 118
CONCLUSION 158
1. Contributions 159
1.1. Contributions to the study of market categories and audiences' valuations 159
1.1.1. Contribution to the literature on prototype-based categories 160
1.1.2. Contributions to the literature on goal-based and exemplar-based categories 161
1.1.3. Contributions to the literature on optimal distinctiveness 162
1.2. Contributions to computational approaches to organizations 164
1.3. Implications for practice 165
2. Limitations and future research 167
2.1. Current work limitations 167
2.2. Avenues for future research 168
Final words 171

LIST OF FIGURES

INTRODUCTION - Figure 1. Monthly reviews for three games released on the platform Steam in the second half of 2016 14
INTRODUCTION - Figure 2. Articulation of the three essays 26
INTRODUCTION - Figure 3. Structure of the dissertation 31
CHAPTER 1 - Figure 1. Main and moderated effects of typicality on the standard deviation in monthly returns 69
CHAPTER 1 - Figure 2. Main and moderated effects of typicality on the standard deviation in daily returns 69
CHAPTER 2 - Figure 1. Effects of typicality and attractiveness on first-day returns (model 4) 110
CHAPTER 2 - Figure 2. Effects of typicality and attractiveness on first-day returns as a function of investor sentiment (model 5 and 6) 110
CHAPTER 3 - Figure 1. Determinants of audiences' valuations given their interests 152
CHAPTER 3 - Figure 2a. Illustration of the case where audiences all behave as prototype-based evaluators (Proposition 2a) 153
CHAPTER 3 - Figure 2b. Illustration of the case where audiences all behave as exemplar-based evaluators (Proposition 2b) 154
CHAPTER 3 - Figure 2c. Illustration of a case with both prototype-based and exemplar-based evaluators (Proposition 2c) 155
CHAPTER 3 - Figure 3a. Illustration of the case where audiences all behave as goal-based evaluators (Proposition 3a) 156
CHAPTER 3 - Figure 3b. Illustration of the case with both prototype-based and goal-based evaluators (Proposition 3b) 157

LIST OF TABLES

CHAPTER 1 - Table 1. Most similar words to selected words as evaluated using the word embedding model trained on business overview sections 70
CHAPTER 1 - Table 2. 15 most similar words to industry prototypes for the 10 most represented industries in 2017 71
CHAPTER 1 - Table 3. Five firms most similar and least similar to the industry prototype for the business services industry (SIC code 73) in 2017 72
CHAPTER 1 - Table 4. Descriptive statistics and correlations 73
CHAPTER 1 - Table 5. OLS regressions of monthly volatility (models 1 to 5) and daily volatility (models 6 to 10) on typicality... 74
CHAPTER 2 - Table 1. Three selected IPOs and their five most similar peers (in descending order) 111
CHAPTER 2 - Table 2. Established firms most similar to the selected industry prototypes for 2000, 2007, and 2014 (in descending order).. 112
CHAPTER 2 - Table 3. The 5 firms with the highest level of first-day returns in 2000, 2007, and 2014 and the top 5 most attractive firms in 2001, 2008 and 2015.................. 113
CHAPTER 2 - Table 4. Descriptive statistics and correlations 114
CHAPTER 2 - Table 5. OLS regressions of first-day returns on typicality and attractiveness 115
CHAPTER 2 - Table 6. OLS regressions of first-day returns on relative typicality and attractiveness 116
CHAPTER 2 - Table A1. OLS regressions of first-day returns on typicality and attractiveness when using 100 dimensions (model A7) or 200 dimensions (model B7) for document vectors.. 117
CHAPTER 3 - Table 1. Main features of all three models of valuation......................... 151

INTRODUCTION

In January 2019, the office rental company WeWork, the ‘Uber of shared offices’, was worth $47 billion. At the end of September 2019, after a failed attempt at going public, WeWork’s value fell to about $11 billion. While the revelation of wrong doings by WeWork’s co-founder played an important role in this debacle, it is not the only culprit. The socio-cognitive dynamics at play greatly contributed to this drastic revision in WeWork’s value. First, investors started to doubt that ‘tech’ companies were still ‘hot’: in May 2019, Uber itself did its IPO and experienced a negative return of 6.7% on its first day of trading. Second, investors noticed that already existing real estate companies, such as BXP, were much more similar to WeWork than ‘tech’ companies such as Facebook or Google. These companies were generally valued with smaller multiples than WeWork. Third, the use by WeWork’s management of new and poorly understood measures of performance in the IPO prospectus led investors to question the appropriateness of previous valuations. Fourth, the contrast between the well-established nature of WeWork’s activities and its founder’s attempts at presenting it as a disruptive company created unease among investors. Overall, the realization that WeWork did not qualify as a ‘tech’ company and growing scepticism about both this market category and WeWork’s categorization of its activities all contributed to its downfall.

This example shows that audiences’ perceptions of what an organization is or is not in part determine how they value it. In fact, models of valuation as developed in organizational research generally present the identification of what an organization is as a necessary first step for valuation (Hannan et al., 2019; Zuckerman, 2017). In other words, they hold that audiences’ valuations of organizations depend on how they categorize them. Categorization involves ascribing entities to categories, where categories are the symbolic and material attributes of products, organizations, and industries that are both shared among actors and that distinguish these entities from others (Durand & Thornton, 2018). Categories bring stability to markets by ensuring that audiences can converge on comparable assessments of organizations and their products (Zuckerman, 1999).

The impact of categories on audiences' valuation is a key area of research in organization studies given the importance of valuation for the survival of any kind of organizations. Indeed, audiences only support organizations if they find them *worthy* of their support. This seems relatively self-evident in market contexts: companies have to convince various stakeholders -such as shareholders, employees or NGOs- of the legitimacy of their products and activities if they want to avoid contestation or leverage resources from them (Bitektine, 2011; Vergne, 2012). However, the same is true of all kinds of organizations such as political parties, churches, opera houses or individual projects (Durand & Kremp, 2016; Jones & Massa, 2013; Karthikeyan, Jonsson, & Wezel, 2015; Kim & Jensen, 2011; Leung & Sharkey, 2014). Understanding the socio-cognitive processes through which audiences (clients, suppliers, critics, analysts, investors, activists, etc.) value organizations is thus key to understanding why some organizations thrive and persist while others wither and disappear (Cattani, Porac, & Thomas, 2017; Vergne & Wry, 2014), which explains why over 100 articles on market categories were published since 2011 in top management journals (Durand & Thornton, 2018).

Studies of organizations focus on one type of categories determining audiences' valuations: *prototype-based categories*. Examples of prototype-based categories are movie or literary genres (Hsu, 2006; Kovács & Hannan, 2010), types of cuisines (Rao, Monin, & Durand, 2005), industry categories (Kennedy, 2008; Navis & Glynn, 2010; Porac, Thomas, & Baden-Fuller, 1989; Ruef & Patterson, 2009; Vergne, 2012; Zuckerman, 1999, 2004), or product categories (Barlow, Verhaal, & Angus, 2019; Rosa, Porac, Runser-Spanjol, & Saxon, 1999; Zhao, Ishihara, Jennings, & Lounsbury, 2018). These categories are widely known among audiences and gather items which share family resemblance. As their name indicates, prototype-based categories are defined by a prototype, which is an abstract representation of the most typical member of a category, and membership within the category is a function of one's similarity to the prototype. Prototype-based categories are a key component of the *prototype-based model of valuation*. This model holds that prototype-based categories are relatively fixed, shared among audiences and that audiences would systematically value typical firms more positively (Hannan et al., 2019). Thus, prototype-based categories stabilize valuation in markets (Hannan et al., 2019; Schneiberg & Berk, 2010; Zuckerman, 1999).

As the prototype-based model of valuation emphasizes audiences' reliance on pre-existing, well-established and relatively fixed categories, it is ill-equipped to account for the variability of audiences' valuations. Yet, the value of many entities can vary substantially, even in contexts where there exist well-established categories. For example, the stock price of publicly listed firms is generally volatile, and this volatility in part reflects underlying fluctuations in investors' (mis)perceptions (Brandt, Brav, Graham, & Kumar, 2010; Foucault, Sraer, & Thesmar, 2011; Stambaugh, Yu, & Yuan, 2015). Different audiences can also have different preferences so that the value of typical products may vary from one audience to another. For example, while some movie-goers love films fitting into existing genres, others have a preference for films blending different genres (Goldberg, Hannan, & Kovacs, 2016). Finally, audiences do not always define objects using prototype-based categories. Sometimes audiences may create *ad hoc* categories (Durand & Paolella, 2013). For example, one may explicitly seek a restaurant to take someone on a first date rather than a restaurant specialized in a type of cuisine. Other times, audiences define what an entity is in terms of other, already known entities (Zhao et al., 2018). For instance, one may categorize a game as a 'Rogue-like', after the videogame Rogue. These different modes of categorization in turn leads to valuations which do not necessarily favour typical entities.

Since audiences' perceptions can be fuzzy and shifting, since different audiences can have different preferences for typicality and since audiences may use multiple modes of valuation, the relationship between categorization and valuation can be more complex than pictured by the prototype-based model of valuation. Organizations face multiple and potentially heterogeneous audiences and they have to understand and embrace this complexity if they want to appear worthy in the eyes of some of these audiences. This dissertation seeks to shade some light on the different mechanisms that drive the variability of audiences' valuations.

1. The prototype-based model of valuation and the stability of audiences' valuations

1.1. Origins in cognitive psychology

Prototype-based categories were first studied in socio-cognitive psychology research on category learning which adapted Wittgenstein's idea that knowing what a word means does not involve learning a precise definition of its meaning but rather learning how to use it from overlapping similarities between different contexts in which the word is used (Wittgenstein, 1953). Wittgenstein proposed that the different entities to which a noun can refer share *family resemblance* – i.e. relationships of similarity along multiple and variable dimensions – rather than a definite set of features.

Following this insight, socio-cognitive psychologists proposed that the categories to which common nouns refer, such as 'table', 'chair' or 'birds', are defined by a prototype, which is an abstract representation of the most typical members of a category (Reed, 1972), rather than by a set of well-defined features or rules of membership, and that membership into a category is a function of an entity's similarity to the prototype, i.e. its typicality (Mervis & Rosch, 1981; Rosch, 1975; Rosch & Mervis, 1975). For example, a table knife is typical of the category knife and clearly belongs in this category as it is very similar to our abstract representation of a knife, while the Swiss Army knife, which is also a bottle opener, a scissor and many other tools, is an atypical knife and does not clearly belongs to this category as it is very dissimilar from the prototypical knife. Humans learn prototypes by abstracting a representation of the most typical member of the category from observed exemplars of the category. Categorizing and entity as a member of a prototype-based category allows to set one's expectations relative to the entity (Cantor, Mischel, & Schwartz, 1982). For example, defining an event as a party allows to formulate generally accurate expectations regarding who will come and how to behave.

1.2. The prototype-based model of valuation in organization studies

Prototype-based categories were leveraged in organization studies to explain audiences' valuations, becoming part of a *prototype-based model of valuation*. The prototype-based model of valuation holds that audiences make sense of newly observed entities in terms of pre-existing categories shared among audiences (Hannan et al., 2019). Under this view, when determining what an entity is, audiences are primarily concerned with how well it can be defined in terms of already known categories and value typical entities more positively for two distinct reasons.

First, atypical entities are harder to categorize using existing categories, resulting in disfluency in the processing of their features, which leads audiences to penalize them (Hannan et al., 2019; Hsu, Koçak, & Hannan, 2009). Second, provided that pre-existing categories have a positive valence in the eyes of audiences – which is the standard assumption in most settings in organization studies –, audiences view their members more positively, in proportion of their typicality (Hannan et al., 2019; Hsu, 2006). In other words, typical entities have a higher intrinsic appeal in the eyes of audiences (Hsu et al., 2009). The positive relationship between typicality and valuation and its mediation by appeal in the eyes of audiences have been tested and supported in many different settings: typical restaurants

(Kovács & Hannan, 2010), typical movies (Hsu, 2006), typical books (Kovács & Hannan, 2015), typical winemakers (Negro & Leung, 2013), even typical persons (Leung & Sharkey, 2014) are all seen more positively, leading to various advantages such as critical acclaim or an ability to gather more resources.

Based on this account, prototype-based valuation operates as a disciplining mechanism and bring stability to audiences' valuations. Prototype-based valuation ensures that deviants are systematically weeded out and typical organizations systematically valued more positively and thus more likely to thrive. Moreover, prototype-based categories tend to change slowly over time, further stabilizing audiences' valuations.

2. Exploring the variability of audiences' valuation

While the predictions of the prototype-based model of valuation are supported in numerous contexts, recent results suggest that audiences' valuations can vary substantially, both over time and from audience to audience. Given the prototype-based model's emphasis on audiences' reliance on relatively stable and fixed categories to structure their valuation, this leads to wonder whether the prototype-based model of valuation is well equipped to explain the variability of audiences' valuations. In this section, I discuss the variability of audiences' valuations in more depth and introduce alternative models of valuation. I then present the main research question addressed in this dissertation.

2.1. The variability of audiences' valuations and its importance for research on categories

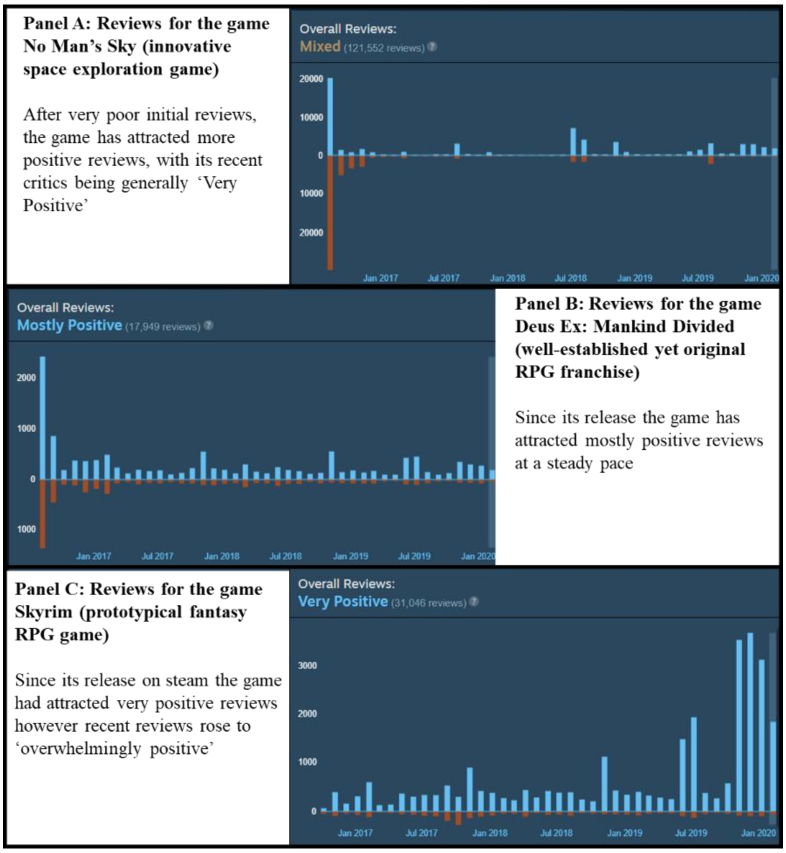
The variability of audiences' valuations can appear as a potential limitation for the prototype-based model of valuation for different reasons. First, while the prototype-based model of valuation does not exclude the possibility that audiences' valuations may vary, it does emphasize audiences' reliance on relatively fixed and stable categories to structure their valuation. Thus, observations of widespread variations in audiences' valuations begs further examination of the adequacy of the model. Notably, variations in values may result from changes in the meaning of categories, leading to question their stability (Lo, Fiss, Rhee, & Kennedy, 2019). Different audiences can also have different preferences for typicality, introducing variations in the valuation of typical objects from one audience to another (Goldberg et al., 2016; Pontikes, 2012). Finally, audiences do not always use prototype-based categories to structure their valuations, so that typical entities are not valued more positively in all contexts (Durand & Paolella, 2013; Paolella & Durand, 2016; Paolella & Sharkey, 2017).

2.1.1. The variability of audiences' valuations as an empirical fact

Audiences' valuations are subject to constant change and adjustments in a variety of settings. In financial markets, stock prices – which reflect investors' consensus estimate of the worth of a firm at a given point in time – can be highly volatile (Brandt et al., 2010; Foucault et al., 2011; Stambaugh et al., 2015) although valuation in financial markets is heavily structured by established industry categories (Zuckerman, 1999, 2004). Notably, the stock price of atypical publicly listed firms is more volatile after the first quarterly earnings announcement of the year, suggesting that investors have difficulty interpreting information about atypical firms (Zuckerman, 2004). More traditional products also experience variability in their prices. For example, there is a greater variability in the prices set by wine producers when critics do not use clear evaluative schemas in their reviews (Hsu, Roberts, & Swaminathan, 2012). Critical consensus can change over time as well, as one can easily observe nowadays on online platforms gathering reviews of movies, books or video games. For example, the video game platform Steam allows user to give either a positive or a negative review to games and indicates both the overall critical consensus across all reviews ever published on the platform and the recent critical consensus across recent reviews. The platform also shows the proportion of positive and negative reviews published each month about any game, and it is easy to identify different trajectories for different games.

Figure 1 presents histograms of positive and negative reviews for three different games added to the platform Steam in the second half of 2016. Panel A shows reviews for the game No Man's Sky, released on the platform in August 2016. The game initially received predominantly negative reviews and the overall consensus across all reviews is still 'Mixed' today. However, recent reviews contain a much higher proportion of positive reviews and the recent consensus showed on the platform is 'Very Positive', suggesting that the game is being re-evaluated much more positively, several years after its initial release. Panel B shows reviews for the game Deus Ex: Mankind Divided, also released in August 2016. Here we see a generally stable pattern, with a relatively constant dominance of positive reviews, leading to a 'Mostly Positive' overall consensus as well as a 'Mostly Positive' recent consensus. Finally, Panel C shows reviews for the game Skyrim Special Edition, released on the platform in October 2016. This game is a re-edition of a well-known game initially released in 2011. Interestingly, while the overall consensus for the game is 'Very Positive', recent reviews experienced a spike of positive reviews and are 'Overwhelmingly Positive'. These different trajectories for different games, as well as monthly variations in the quantity and quality of reviews show that audiences' valuations can follow multiple patterns and are not necessarily stable over time.

Figure 1. Monthly reviews for three games released on the platform Steam in the second half of 2016



One important factor introducing variability in audiences' valuations is the evolving meaning of categories. Being typical of a category ensures superior valuation only to the extent that audiences value membership into the category, i.e. to the extent that the category has a high currency or viability (Kennedy & Fiss, 2013; Kennedy, Lo, & Lounsbury, 2010; Lo et al., 2019). If a category loses some of its intrinsic appeal or if it becomes less coherent, it loses some of its currency (Kennedy et al., 2010). Similarly, categories that lack contrast relative to other categories or, on the contrary, are too sharply delineated, are less viable in the eyes of audiences and thus fall in disfavour (Lo et al., 2019). New categories can also be created *ex nihilo* by coalitions of interested actors, increasing the value of previously unrelated objects (Durand & Khaire, 2017). For example, the overlapping interests of art historians and auction houses led to the recognition of previously unrelated pieces of art as members of a modern Indian art category, in turn enhancing their values (Khaire &

Wadhwani, 2010). The meaning of existing categories can also be modified by organizations in order to increase their value in the eyes of targeted audiences (Delmestri & Greenwood, 2016; Weber, Heinze, & DeSoucey, 2008). For instance, the re-evaluation of Grappa as a high status category of alcohol is the result of purposeful actions from a small group of producers who laboured to convince high-brow audiences of the value of their product (Delmestri & Greenwood, 2016). Similarly, grass-fed meat producers were able to convince young urbans of the value of their products by emphasizing their authenticity and quality (Weber et al., 2008).

Taken together, these arguments suggest that audiences' valuations can and do vary, sometimes substantially, both from one audience to another and over time. Thus, audiences' valuations may not be as stable as they would be if audiences always relied on pre-existing, relatively stable prototypes as a basis for their valuation. Researchers have acknowledged this phenomenon and tried to explain it, either within the prototype-based model of valuation or by producing alternative models of valuation.

2.1.2. The heterogeneity of audiences within the prototype-based model of valuation

Within the prototype-based model of valuation, recent research suggests that some audiences prefer atypical entities to typical ones. In the context of cultural consumption, studies using data from Netflix and Yelp show that some audiences have an interest in movies which span multiple genres or in restaurants which mix different types of cuisine (Goldberg et al., 2016). More precisely, consumers of cultural products can be decomposed in four different types of audiences. 'Mono-purists' have a preference for a small number of categories and prefer typical entities within these categories. 'Poly-purists' like numerous different categories but prefer typical entities within each of them. 'Mono-mixers' appreciate a small number of categories but prefer entities mixing features from this small pool of categories. Finally, ‘Poly-mixers’ have an inclination for a large number of categories and prefer entities mixing features from these different categories.

Other findings suggest that venture capitalists prefer to invest in firms which are associated with ambiguous labels rather than in firms associated with clear labels – unlike consumers who prefer the products of typical firms (Pontikes, 2012). Atypical hedge funds also seem to avoid punishment from investors following periods of poor performance and attract more investments following periods of good performance (Smith, 2011). Thus, although none of these studies contest that audiences use prototypes as a basis for their valuation, all of them suggest that audiences can be heterogeneous with respect to their preferences for typicality, resulting in substantial variations in the valuation of typical entities from audience to audience.

This finding is important because it suggests that audiences are much less conservative than previously thought. Some audiences purposively look for atypical entities and expect organizations to go against categorical expectations. Thus, producers of typical products risk alienating ‘mixers’ – and conversely, innovative or avant-garde producers are likely to appeal only to them. Organizations should thus consider carefully who their intended audiences are when they choose to emphasize the typicality or atypicality of their offerings. Furthermore, organizations have to balance the need to appeal to their audiences with other benefits which may stem from atypicality or ambiguity, such as greater flexibility and reduced scrutiny (Pontikes & Barnett, 2015; Pontikes & Kim, 2017).

Going beyond the prototype-based model of valuation, two lines of research suggest that audiences do not always use pre-existing categories to value organizations and their products. One line of research proposes that audiences sometimes derive idiosyncratic, *goal-based categories* (Durand & Paolella, 2013; Granqvist & Ritvala, 2016; Paolella & Durand, 2016), while another line of research proposes that they sometimes rely on *salient exemplars* rather than prototypes to value entities (Barlow et al., 2019; Zhao et al., 2018).

2.1.3. The multiplicity of audiences' modes of valuation

The goal-based model of valuation. The goal-based model of valuation presents audiences as deriving idiosyncratic, goal-based categories to value newly observed entities. The goal-based model of valuation is based on socio-cognitive research which proposes that humans do not always rely on pre-existing, well-established categories to categorize entities, but sometimes derive new categories which will help them achieve their particular goals (Barsalou, 1985, 1991; Durand & Boulongne, 2017). For example, someone seeking to lose weight might seek to identify products belonging to the goal-based category 'food to eat on a diet'. Goal-based categories are defined based on an *ideal*, which is an abstract representation of the best tool to achieve one's goal (Barsalou, 1985). As an illustration, 'zero-calorie food' might be the ideal of the category 'food to eat on a diet'. Goal-based categories and the ideals defining them are actively created by audiences by combining features which might help them achieve their goals through a process of conceptual combination (Barsalou, 1991). Goal-based categories may or may not overlap with pre-existing prototype-based categories, depending on the type of solutions that one seeks. Under the goal-based view, when determining what an entity is, audiences are primarily concerned with determining whether it can be a tool to achieve their current goals.

Goal-based categories provide the basis for the *goal-based model of valuation*, according to which audiences value more positively entities which are similar to the ideal that they use to screen audiences (Zuckerman, 2017). It follows that if the ideal candidate combines features from multiple prototype-based categories, and is thus atypical of each of them, atypical entities can sometimes be valued more positively (Paolella & Durand, 2016; Paolella & Sharkey, 2017). For example, law firms which span multiple categories of legal services offer to their client a much needed flexibility and are thus valued more positively (Paolella & Durand, 2016). Importantly, audiences may use *both* prototype-based and goal-based valuation, resulting in a situation where both typical entities and those aligned with ideals are valued more positively. In the law firm context, this leads to a U-shape relationship between typicality and valuation as both very typical and very atypical firms are valued more positively (Paolella & Sharkey, 2017). While goal-based valuation can guarantee a relative stability of market exchanges as long as there exists ways of limiting the diversity of the goal-based categories with which audiences can come up with (Glaser, Krikorian Atkinson, & Fiss, 2019), it can also leads to less stability in audiences' valuations as different audiences have different goals and audiences' goals shift over time.

The exemplar-based model of valuation. The exemplar-based model of valuation builds on exemplar-based categorization as developed in socio-cognitive research on categories. Exemplar-based models of categorization suggests that humans categorize newly observed entities into pre-existing categories not only based on their similarity to prototypes but also based on their similarity with specific members of these categories – i.e. exemplars (Cohen & Basu, 1987; Homa, Sterling, & Trepel, 1981; Nosofsky & Zaki, 2002). In organization studies, the exemplar-based model of valuation was first introduced in a context void of pre-existing categories and aimed at explaining how 'proto-categories' appear around salient exemplars (Zhao et al., 2018). The exemplar-based model of valuation proposes that salient exemplars can be used as a yardstick by audiences to value newly observed entities. For example, in the early days of the video-game industry, games which were similar to recent successes generally sold more copies and received better evaluations from critics (Zhao et al., 2018). Under this view, when determining what an entity is, audiences are primarily concerned with determining whether it has the features of others, already known entities.

Several factors contribute to render an exemplar salient. First, success and critical acclaims can bring an exemplar to the forefront and lead to extant discussion in the public discourse of the exemplar and its most specific features (Barlow et al., 2019; Zhao et al., 2018). For example, apps that uses words similar to those of successful apps on the Google Play platform receive more reviews from the platform's users (Barlow et al., 2019). Second, an exemplar can be salient simply due to its stronger than the average association with features characteristic of a category – i.e. by exhibiting conventionality (Durand & Kremp, 2016). As an illustration, musical directors of middle status orchestra tend to program canonical pieces more often as it allows them to shine more brightly among their peers (Durand & Kremp, 2016). Third, an exemplar can become salient due to its ability to represent extant theorizing within a field or to foster new theorizing (Nigam & Ocasio, 2010). Fourth, an exemplar can be consecrated as a highly salient member of its category through the active involvement and dedication of a small group of devotees (Jones & Massa, 2013).

Audiences may rely on both prototype-based valuation and exemplar-based valuation, leading to potentially incongruent valuations (Barlow et al., 2019). Results in multiple settings tend to concur with the observation that audiences can behave either as prototype-based, goal-based or exemplar-based evaluators. This introduces a new source of variability in audiences' valuations: their shifts from one model of valuation to another.

Understanding the model(s) of valuation used by audiences is paramount for organizations to thrive. Offering prototypical offerings when audiences actually use goal-based categories mixing features from multiple prototype-based categories can lead to poorer valuations (Paolella & Durand, 2016) and ultimately to one's own demise. Organizations cannot simply assume that audiences use prototype-based categories; a good understanding of their goals and a working knowledge of the salient exemplars that structure their valuations can be as important as a good fit with established categories.

2.2. Main research question and three research gaps

2.2.1. Main research question

Recent developments in the literature on audiences' valuations of organizations thus generally challenge the idea that audiences primarily rely on relatively stable and fixed prototype-based categories to structure their valuation and systematically value typical entities more positively. On the contrary they emphasize the variability of audiences' valuations. The meanings tied to category change and evolve, and new categories emerge, leading to shifts in audiences' valuation (Delmestri & Greenwood, 2016; Durand & Khaire, 2017; Lo et al., 2019). Some audiences do not systematically prefer typical entities and are more inclined toward atypical ones (Goldberg et al., 2016; Pontikes, 2012; Smith, 2011). Audiences that favour typical entities in one context may behave as goal-based evaluators in another (Durand & Boulongne, 2017; Paolella & Sharkey, 2017). All these findings point toward an inherent variability of audiences' valuations which needs to be accounted for. In fact, current shifts in research on categories and valuation may point to a broader re-orientation of the fundamental question addressed by this research. While studies which originated the field sought to answer the question 'Why are audiences' valuations so stable?' by highlighting the stabilizing roles of prototype-based categories in market exchanges (Hsu et al., 2009; Zuckerman, 1999), recent research seems to ask 'Why are audiences' valuation so variable?'. This is precisely the overarching question that this dissertation seeks to address:

Why are audiences' valuations of organizations so variable?

2.2.2. Three research gaps

This dissertation focuses on three research gaps related to this overarching research question. The first gap relates to the direct impact of typicality on the variability of audiences' valuations. Since the results introduced above suggest that audiences' valuations may vary substantially, even in the presence of market categories, it is important to re-examine this impact which, surprisingly, has been the object of only a few studies. In a 2004 paper, Zuckerman proposed that typical publicly listed firms experience less volatility in the days following the first quarterly earnings announcement of the year as it is easier for investors to converge on a common interpretation of new pieces of information regarding typical firms (Zuckerman, 2004). Per this account, the introduction of a new piece of information triggers periods of volatility. Hsu and colleagues also showed that atypical wine producers are less able to rely on the clarity of critics' evaluative schemas to price their wines (Hsu et al., 2012). Thus, for a given level of clarity of critics' evaluative schemas, prices set by atypical producers tend to exhibit greater variability around expected levels. Their study focused on the variability of wine prices at a given point in time around expected levels rather than on the volatility of prices over time. It also focused on producers' valuations of their own products based on the clarity of critics' evaluative schemas rather than on audiences' assessments. Since audiences' valuations are often much more variable than it seems even in the presence of market categories, this dearth of studies on the link between typicality and the variability of audiences' valuations is puzzling. The first essay of this thesis thus asks the question:

<u>*Gap 1.*</u> *Does typicality lead to less variability in audiences' valuations?*

The second gap addressed in this dissertation relates to the existence of temporary attractions among audiences toward certain features which influence their valuations alongside stable and well-established prototypes. The dominant perspective within the prototype-based model of valuation is that audiences make sense of entities in terms of their similarity to pre-existing prototypes, i.e. their typicality. Typical entities are more appealing to audiences which in turn value them more positively. As prototypes change slowly, this dimension of appeal tends to be relatively stable, in other words, for a given level of typicality, one tends to be rewarded with the same premium in valuation. However, this approach neglects the impact that temporary trends or hypes, i.e. temporary attractions toward certain features, can have on the appeal of a given entity (Abrahamson & Fairchild, 1999; Lee, 2001). Notably, audiences are attracted toward organizations exhibiting features which are currently expected to lead to higher performance due to recent successes (Zhao et al., 2018). As past successes fade away and new successes occur, audiences' attractions toward certain features come and go, and the attractiveness of organizations -i.e. their similarity to past successes- varies over time, which introduces variability in audiences' valuations. This second, destabilizing dimension of appeal has been generally ignored by the literature. Hence, the second essay of this thesis asks:

<u>Gap 2.</u> *What is the impact of organizations' attractiveness on audiences' valuations and is it congruent with that of typicality?*

The third gap addressed in this dissertation relates to audiences' reliance on multiple models of valuation. Audiences' shift from one model of valuation to another is clearly a potential source of variability in audiences' valuations. While we already possess several results suggesting that audiences do rely on these alternative models of valuation, little theoretical effort has been dedicated to understanding when and why audiences sometimes behave as prototype-based evaluators and sometimes behave as goal-based or exemplar-based evaluators. Integrating the three models of valuation is important to produce a comprehensive account of empirical findings which may otherwise seem contradictory, notably regarding the relationship between typicality and valuation. Hence, the third essay of this thesis asks:

<u>Gap 3.</u> *When and why do audiences behave as prototype-based, goal-based or exemplar-based evaluators? What are the consequences for their valuations?*

This dissertation thus tries to lay the groundwork to the study of variability of audiences' valuations under the lens of market categories. Taking stock of the inherent variability of audiences' valuations even in the presence of pre-existing categories, this dissertation first seeks to re-examine how typicality relates to the variability of audiences' valuations (Gap 1). It then seeks to study how organizations' attractiveness influence their valuations alongside typicality as, unlike categories, temporary attractions among audiences toward certain features change quickly over time and are thus susceptible to induce variability in audiences' valuations (Gap 2). Finally, since audiences do not rely solely on prototype-based valuation, this dissertation seeks to determine the factors influencing audiences' use of different models of valuation and to explore how audiences' uses of different models induce variability in their valuations (Gap 3).

3. Introduction to the three essays

This dissertation follows a three-essay format and applies innovative Natural Language Processing (henceforth NLP) methods to financial documents (annual reports and IPO prospectuses) to study the valuation of publicly listed firms in the U.S.

3.1. Essay 1 - The (relative) effects of typicality on volatility: A study using word embeddings

The first essay of this dissertation aims at addressing Gap 1: *Does typicality leads to less variability in audiences' valuations?* It does so by studying the volatility of the stock price of publicly listed firms in the U.S.. Since stock prices reflect investors' consensus estimate of the worth of a given firm, stock price volatility is an appealing measure of the overall variability of investors' valuations. Past studies in the categories literature showed that typical firms experience less volatility in the days following the first quarterly earnings announcement of the year (Zuckerman, 2004). Thus, typicality would reduce volatility spurred by the production of new information. This essay adopts a different angle by proposing that the stock price of typical firms is generally less volatile but that this relationship is contingent on the overall ambiguity of the firm's industry category.

While finance scholars have come up with many different explanations and interpretations of stock price volatility, this study focuses on explanations which tie volatility to the actions of uninformed investors (De Long, Shleifer, Summers, & Waldmann, 1990). Uninformed investors generate volatility in the valuations of publicly listed firms by trading on 'noise' rather than information and thus mispricing firms (Brandt et al., 2010; Foucault et al., 2011; Stambaugh et al., 2015). As a baseline hypothesis, I contend that general knowledge about category members encoded in the category's prototype provides relevant information to value typical firms. Thus, typical firms are generally less exposed to uninformed investors. Hence, the stock price of typical firms is generally less volatile. However, in ambiguous industry categories, the information encoded in prototypes is less relevant to value typical firms, resulting in an attenuation of the relationship between typicality and volatility.

3.2. Essay 2 - Organizational appeal and market valuation: A natural language processing study of IPO first-day returns

The second essay of this dissertation addresses Gap 2: *What is the impact of organizations' attractiveness on audiences' valuations and is it congruent with that of typicality?* The literature on typicality tends to present appeal as resulting mainly from a focal organization's similarity to relatively stable and well-established prototypes. Thus, the appeal of an organization would tend to be a relatively fixed trait. In this essay, we point out that audiences often feel temporary attraction toward certain features which they associate with success due to fleeting hypes or trends. Thus, organizational appeal also incorporates a less stable dimension, which we call attractiveness, i.e. an organization's similarity to recent successes. We study the market for IPO, which is especially sensible to temporary hypes or trends as some IPOs regularly experience very high first-day returns (Ibbotson, Sindelar, & Ritter, 1994; Loughran & Ritter, 2002), which render them salient in the eyes of investors who will subsequently be on the look out for similar firms. In line with expectations, we find that typicality does not have a significant effect on first-day returns but does have a marginally significant negative impact on first-day returns when investors' sentiment (Baker & Wurgler, 2006, 2007) is high. By contrast, attractiveness has a positive impact on first-day returns which is enhanced when investors' sentiment is high.

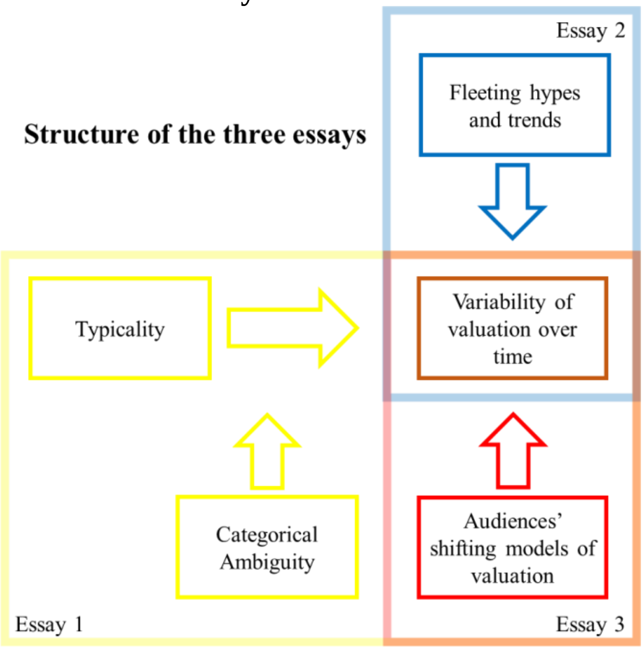
3.3. Essay 3 - Valuing organizations: An integrated theory

The third essay of this dissertation is motivated by the observation that while various attempts have been made at explaining results which conflict with the predictions of the prototype-based model of valuation, we still lack and integrated theory of audiences' valuations able to reconciliate these findings. We thrive to produce such a theory by showing that all three models of valuation -the prototype-based, the goal-based and the exemplar-based models- are three different takes on a single mechanism which posits that audiences value more positively organizations which align with their current center of interest.

Based on this initial observation, we formulate propositions predicting when a focal audience will behave as a prototype-based, goal-based or exemplar-based evaluator. We then consider how audiences' heterogeneity, defined in terms of their sharing the same or different interests, and breadth of interest, defined in terms of the number of features in which they have an interest, interact to determine the shape of the relationship between typicality and valuation as well as the likelihood that a new category will emerge. In so doing, we reconcile multiple findings on audiences' valuation in a single, coherent framework which accounts for the variability of audiences' valuations in numerous settings.

Figure 2 graphically represents all three essays and their articulations.

Figure 2. Articulation of the three essays



4. Methodology and setting

Theories of categorization emphasize distributional approaches to measuring typicality, based on organizations or audiences' uses of category labels and co-occurrences of labels (Hannan et al., 2019; Kovács & Hannan, 2015). As such, they relate naturally to Natural Language Processing methods which represent the meaning of words based on their occurrences in similar contexts (Lenci, 2018). In this dissertation, I embrace this proximity and use distributional approaches to model words' meanings to measure typicality and, more generally, semantic similarities between firms. To do so, I use large corpora of financial documents -annual reports and IPO prospectuses- which are highly suited for this kind of analysis.

4.1. NLP methods as a tool to inform category research

NLP methods, i.e. methods used to study large corpuses of texts written in natural language using machine learning, have gained a lot of attention in organization studies in the past few years. Topic models (Blei, Ng, & Jordan, 2003) are probably the most well-known of these methods among organizational scholars (Corritore, Goldberg, & Srivastava, 2019; Croidieu &

Kim, 2017; DiMaggio, Nag, & Blei, 2013; Haans, 2019; Hannigan et al., 2019). They offer the opportunity to infer from a large corpus of documents the common topics discussed within the corpus, as well as the proportion of each document dedicated to each topic (Blei, 2012; Blei et al., 2003).

However, numerous other NLP methods exist and can be leveraged by organizational scholars to study phenomena of interest. Notably, word embeddings models, which seek to capture the meaning of words by locating them in a semantic space (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013; Pennington, Socher, & Manning, 2014), offer an incredible opportunity to represent the meanings associated to organizations and their products. These models are especially relevant to study categories as they are built upon principles which largely mirrors those supporting extant theorizing about categories. Indeed, just like Wittgenstein held that the meaning of a word is a function of the context in which it is used, word embeddings models are built upon the hypothesis that the meaning of a word is a function of the word with which it co-occurs – i.e. the *distributional hypothesis* (Harris, 1952, 1954; Lenci, 2008, 2018). Word embedding models learn the meaning of a word by iterating over a corpus of documents and trying to predict either a target word given its context words (Continuous Bag-Of-Word model or CBOW) or context words given a target word (Skip-Gram model) (Mikolov et al., 2013). Consequently, word embeddings models produce semantic spaces in which words appearing in similar contexts will tend to be in the same region of the space. This effectively operationalizes the proposition that the meaning of a word is a function of overlapping similarities in the context in which it is used.

Word embeddings can be used to represent the meaning of entire documents by averaging the word vectors of the words composing it or by learning a document vector alongside word vectors during training (Dai, Olah, & Le, 2015; Mikolov et al., 2013). In turn, the position of organizations in the semantic space can be assimilated to that of the documents it produces (or which some other authors produce about it). Thus, word embeddings models offer the opportunity to represent organizations in a semantic space where organizations using similar words in similar documents are located in the same region of the space, i.e. to operationalize the notion of *family resemblance* based on the similarities between different organizations' vectors. Essays 1 and 2 leverage this specificity of word embeddings to study the typicality of publicly listed firms in the U.S. based on the content of their annual reports and their IPO prospectuses.

More generally, NLP models entice scholars to think of meaning both in distributional terms and in spatial terms. Most models assimilate the meaning associated to a word or a document to its position in a semantic space which itself depends on the distributional properties of words (the contexts in which they appear). This approach to meaning resonates with recent attempts in the literature on audiences' valuations to locate organizations in semantic spaces based on the distributional properties of the words used to define them (Corritore et al., 2019; Haans, 2019; Hannan et al., 2019). Essay 3 embraces this intuition by proposing an integrative model of audiences' valuation which represents audiences as having an interest in different points of a semantic space.

4.2. Publicly listed firms in the U.S. and their relevance to study audiences' valuation using NLP

Both Essays 1 and Essay 2 study publicly listed firms in the U.S. using annual reports and IPO prospectuses. Two main considerations drove the choice of this empirical setting to explore Gap 1 and Gap 2. First, one of the most influential paper which spurred interest in categories specifically studied publicly listed firms in the U.S. and established that investors value typical firms at a premium (Zuckerman, 1999). This finding is especially surprising given that financial markets are generally assumed to behave efficiently, valuing investments solely based on information linked to fundamentals (Sharpe, 1964). Essay 1 and 2 prolong category research on publicly listed firms in two ways. Essay 1 shows that typicality affects the volatility of firms' value. Typical firms enjoy less volatility but only when they belong to unambiguous industry categories. As a consequence, the premium that typical firms enjoy may reflect investors' preferences for low volatility firms. In Essay 2, we find that typical IPOs tend to be less underpriced by underwriters when investors' sentiment is high. This suggests that high investors' sentiment leads investors to discount typical IPOs rather than value them at a premium.

The setting of publicly listed firms in the U.S. is also interesting because publicly listed firms are required by law to provide an accurate description of their activities in the financial documents that they publish. Companies are prohibited from omitting material information needed to make a disclosure made in their annual form 10-K (often referred to as their annual report) not misleading. The Sarbanes-Oxley Act (2002) further requires CEOs and CFOs to certify the accuracy of the form 10-K. Similar requirements bind issuers when producing their IPO prospectuses. The semantic content of both these documents can thus be expected to reliably capture a firm's activities and main features. Notably, finance research has shown that these documents can be used to accurately measure product-based similarities and differences between firms and to construct text-based industries (Hoberg & Phillips, 2010, 2016). The setting of publicly listed firms is thus especially appropriate to study firms' typicality using Natural Language Processing.

5. Overall structure of the dissertation

Figure 3 presents the overall structure of the dissertation. Each of the three essays address a different aspect of the overarching question: why are audiences' valuations of organizations so variable? In the first chapter of the dissertation, I find, in line with expectations, that typical firms tend to experience less volatility in their stock prices; however, this relationship is contingent on the ambiguity of their industry category. In other words, the persistence of ambiguous market categories may explain persistent variability in audiences' valuations of typical entities. Chapter 2 of this dissertation presents results in the IPO setting suggesting that attractiveness has an important impact on audiences' valuation alongside that of typicality and that the strength of this impact is influenced by the dominant sentiment among audiences at a given point in time. Finally, in chapter 3 of this dissertation, we present a theoretical model predicting which model of valuation a focal audience is likely to use and how the co-existence of multiple types of evaluators may modify the nature of the relationship between typicality and valuation. Taken together, these essays suggest that persistent variability in audiences' valuations is contingent on 1) the ambiguity of existing market categories, 2) the importance of audiences' temporary attractions toward certain features and 3) the proportion of prototype-based, goal-based or exemplar-based evaluators among audiences. In the conclusion, we highlight the main contributions of this dissertation in more details and discuss areas of future research.

Figure 3. Structure of the dissertation

<table>
<tr><th colspan="3">Introduction</th></tr>
<tr><td colspan="3">Research gap</td></tr>
<tr><td colspan="3">Current literature emphasizes one model of audiences' valuation of organizations on the basis of their similarity to pre-existing and relatively fixed prototypes. This leads to present audiences' valuations as overly stable when evidence indicates that audiences' valuations can sometimes be highly variable from one audience to another and over time</td></tr>
<tr><td colspan="3">Main research question</td></tr>
<tr><td colspan="3">Why are audiences’ valuations of organizations so variable?</td></tr>
<tr><th>Chapter 1</th><th>Chapter 2</th><th>Chapter 3</th></tr>
<tr><td>Research question</td><td>Research question</td><td>Research question</td></tr>
<tr><td>Does typicality lead to less variability in audiences’ valuations?</td><td>What is the impact of organizations’ attractiveness on audiences' valuations and is it congruent with that of typicality?</td><td>When and why do audiences behave as prototype-based, goal-based or exemplar-based evaluators?</td></tr>
<tr><td>Method and data</td><td>Method and data</td><td>Method and data</td></tr>
<tr><td>Doc2Vec model trained on over 140,000 annual reports and IPO prospectuses to represent firms in a shared semantic space

Data on firm fundamentals from Compustat and CRSP</td><td>Word2Vec model trained on over 100,000 annual reports to represent firms in a shared semantic space

Data on firm fundamentals from Compustat and CRSP</td><td>Theoretical paper</td></tr>
<tr><td>Results</td><td>Results</td><td>Results</td></tr>
<tr><td>Typicality leads to lower volatility in firms' stock prices, confirming that typicality has a stabilizing effect on audiences' valuations

However, this effect is a function of the level of ambiguity within a firm's industry category, suggesting that the stabilizing effect of typicality is limited</td><td>Typicality leads to lower first-day returns for IPOs when investor sentiment is high

Attractiveness leads to higher first-day returns. This effect is stronger when investor sentiment is high</td><td>Audiences have different centers of interest and value organizations aligned with them more positively

The breadth of audiences' centers of interests; their alignment with pre-existing prototypes and audiences’ propensity to share the same centers of interests determine their valuations

Exemplar-based valuation favours the emergence of new categories while goal-based valuation hinders it</td></tr>
<tr><td colspan="3">Main results</td></tr>
<tr><td colspan="3">Audiences' valuations can vary substantially both from audience to audience and over time, even in the presence of established categories. Typicality reduces the variability of audiences' valuations only as a function of categorical ambiguity. Temporary attractions toward certain features introduce temporary shifts in audiences' valuations. Audiences' centers of interests need not align with pre-existing prototypes, resulting in variable valuations from one audience to another. Organizations seeking to achieve superior value in the eyes of audiences have to embrace this complexity and develop a good understanding of the distribution of audiences' interests toward know features, idiosyncratic goals, or salient exemplars to be able to secure the support of some audiences.</td></tr>
</table>

REFERENCES

Abrahamson, E., & Fairchild, G. 1999. Management fashion: Lifecycles, triggers, and collective learning processes. ***Administrative Science Quarterly***, 44(4): 708–740.

Baker, M., & Wurgler, J. 2006. Investor Sentiment and the Cross-Section of Stock Returns. ***The Journal of Finance***, 61(4): 1645–1680.

Baker, M., & Wurgler, J. 2007. Investor Sentiment in the Stock Market. ***The Journal of Economic Perspectives***, 21(2): 129–151.

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

Barsalou, L. W. 1985. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. ***Journal of Experimental Psychology. Learning, Memory, and Cognition***, 11(4): 629–654.

Barsalou, L. W. 1991. Deriving Categories to Achieve Goals. In G. H. Bower (Ed.), ***Psychology of Learning and Motivation***, vol. 27: 1–64. Academic Press.

Bitektine, A. 2011. Toward a Theory of Social Judgments of Organizations: The Case of Legitimacy, Reputation, and Status. ***Academy of Management Review***, 36(1): 151–179.

Blei, D. M. 2012. Probabilistic topic models. ***Communications of the ACM***, 55(4): 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent Dirichlet Allocation. ***Journal of Machine Learning Research***, 3(Jan): 993–1022.

Brandt, M. W., Brav, A., Graham, J. R., & Kumar, A. 2010. The Idiosyncratic Volatility Puzzle: Time Trend or Speculative Episodes? ***Review of Financial Studies***, 23(2): 863–899.

Cantor, N., Mischel, W., & Schwartz, J. C. 1982. A prototype analysis of psychological situations. ***Cognitive Psychology***, 14(1): 45–77.

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

Cohen, J. B., & Basu, K. 1987. Alternative Models of Categorization: Toward a Contingent Processing Framework. ***Journal of Consumer Research***, 13(4): 455.

Corritore, M., Goldberg, A., & Srivastava, S. B. 2019. Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance. ***Administrative Science Quarterly***, 0001839219844175.

Croidieu, G., & Kim, P. H. 2017. Labor of Love: Amateurs and Lay-expertise Legitimation in the Early U.S. Radio Field. ***Administrative Science Quarterly***, 0001839216686531.

Dai, A. M., Olah, C., & Le, Q. V. 2015. Document Embedding with Paragraph Vectors. ***NIPS 2014 Deep Learning Workshop***. http://arxiv.org/abs/1507.07998.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. 1990. Noise Trader Risk in Financial Markets. ***Journal of Political Economy***, 98(4): 703–738.

Delmestri, G., & Greenwood, R. 2016. How Cinderella Became a Queen: Theorizing Radical Status Change. ***Administrative Science Quarterly***, 61(4): 507–550.

DiMaggio, P., Nag, M., & Blei, D. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. ***Poetics***, 41(6): 570–606.

Durand, R., & Boulongne, R. 2017. Advancing category research: Theoretical mapping and under-researched areas. ***The Sage Handbook of Organizational Institutionalism***. London.

Durand, R., & Khaire, M. 2017. Where do market categories come from and how? Distinguishing category creation from category emergence. ***Journal of Management***, 43(1): 87–110.
Durand, R., & Kremp, P.-A. 2016. Classical Deviation: Organizational and Individual Status as Antecedents of Conformity. ***Academy of Management Journal***, 59(1): 65–89.
Durand, R., & Paolella, L. 2013. Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory. ***Journal of Management Studies***, 50(6): 1100–1123.
Durand, R., & Thornton, P. H. 2018. Categorizing Institutional Logics, Institutionalizing Categories: A Review of Two Literatures. ***Academy of Management Annals***, 12(2): 631–658.
Foucault, T., Sraer, D., & Thesmar, D. J. 2011. Individual Investors and Volatility. ***The Journal of Finance***, 66(4): 1369–1406.
Glaser, V. L., Krikorian Atkinson, M., & Fiss, P. C. 2019. Goal-Based Categorization: Dynamic Classification in the Display Advertising Industry. ***Organization Studies***, 017084061988336.
Goldberg, A., Hannan, M. T., & Kovacs, B. 2016. What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. ***American Sociological Review***, (2): 215.
Granqvist, N., & Ritvala, T. 2016. Beyond Prototypes: Drivers of Market Categorization in Functional Foods and Nanotechnology. ***Journal of Management Studies***, 53(2): 210–237.
Haans, R. F. J. 2019. What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories. ***Strategic Management Journal***, 40(1): 3–27.
Hannan, M. T., Mens, G. L., Hsu, G., Kovács, B., Negro, G., et al. 2019. ***Concepts and Categories: Foundations for Sociological and Cultural Analysis*** (1 edition). New York: Columbia University Press.
Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., et al. 2019. Topic Modeling in Management Research: Rendering New Theory from Textual Data. ***Academy of Management Annals***, 13(2): 586–632.
Harris, Z. S. 1952. Discourse Analysis. ***Language***, 28(1): 1–30.
Harris, Z. S. 1954. Distributional Structure. ***WORD***, 10(2–3): 146–162.
Hoberg, G., & Phillips, G. 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. ***Review of Financial Studies***, 23(10): 3773–3811.
Hoberg, G., & Phillips, G. 2016. Text-Based Network Industries and Endogenous Product Differentiation. ***Journal of Political Economy***, 124(5): 1423–1465.
Homa, D., Sterling, S., & Trepel, L. 1981. Limitations of exemplar-based generalization and the abstraction of categorical information. ***Journal of Experimental Psychology: Human Learning and Memory***, 7(6): 418–439.
Hsu, G. 2006. Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production. ***Administrative Science Quarterly***, 51(3): 420–450.
Hsu, G., Koçak, Ö., & Hannan, M. T. 2009. Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests. ***American Sociological Review***, 74(1): 150–169.
Hsu, G., Roberts, P. W., & Swaminathan, A. 2012. Evaluative Schemas and the Mediating Role of Critics. ***Organization Science***, 23(1): 83–97.

Ibbotson, R. G., Sindelar, J. L., & Ritter, J. R. 1994. The Market's Problems with the Pricing of Initial Public Offerings. ***Journal of Applied Corporate Finance***, 7(1): 66–74.

Jones, C., & Massa, F. G. 2013. From Novel Practice to Consecrated Exemplar: Unity Temple as a Case of Institutional Evangelizing. ***Organization Studies***, 34(8): 1099–1136.

Karthikeyan, S. I., Jonsson, S., & Wezel, F. C. 2015. The Travails of Identity Change: Competitor Claims and Distinctiveness of British Political Parties, 1970–1992. ***Organization Science***, 27(1): 106–122.

Kennedy, M. T. 2008. Getting Counted: Markets, Media, and Reality. ***American Sociological Review***, 73(2): 270–295.

Kennedy, M. T., & Fiss, P. C. 2013. An Ontological Turn in Categories Research: From Standards of Legitimacy to Evidence of Actuality: An Ontological Turn in Categories Research. ***Journal of Management Studies***, 50(6): 1138–1154.

Kennedy, M. T., Lo, J. Y.-C., & Lounsbury, M. 2010. Category currency: The changing value of conformity as a function of ongoing meaning construction. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 369–397. Emerald Group Publishing Limited.

Khaire, M., & Wadhwani, R. D. 2010. Changing Landscapes: The Construction of Meaning and Value in a New Market Category—Modern Indian Art. ***Academy of Management Journal***, 53(6): 1281–1304.

Kim, B. K., & Jensen, M. 2011. How Product Order Affects Market Identity: Repertoire Ordering in the U.S. Opera Market. ***Administrative Science Quarterly***, 56(2): 238–256.

Kovács, B., & Hannan, M. T. 2010. The consequences of category spanning depend on contrast. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 175–201. Emerald Group Publishing Limited.

Kovács, B., & Hannan, M. T. 2015. Conceptual Spaces and the Consequences of Category Spanning. ***Sociological Science***, 2: 252–286.

Lee, P. M. 2001. What's in a name.com?: The effects of '.com' name changes on stock prices and trading activity. ***Strategic Management Journal***, 22(8): 793–804.

Lenci, A. 2008. Distributional semantics in linguistic and cognitive research. ***Italian Journal of Linguistics***, 20(1): 1–32.

Lenci, A. 2018. Distributional Models of Word Meaning. ***Annual Review of Linguistics***, 4(1): 151–171.

Leung, M. D., & Sharkey, A. J. 2014. Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount. ***Organization Science***, 25(1): 171–184.

Lo, J. Y.-C., Fiss, P. C., Rhee, E. Y., & Kennedy, M. T. 2019. Category Viability: Balanced Levels of Coherence and Distinctiveness. ***Academy of Management Review***. https://doi.org/10.5465/amr.2017.0011.

Loughran, T., & Ritter, J. R. 2002. Why Don't Issuers Get Upset about Leaving Money on the Table in IPOs? ***The Review of Financial Studies***, 15(2): 413–443.

Mervis, C. B., & Rosch, E. 1981. Categorization of Natural Objects. ***Annual Review of Psychology***, 32(1): 89–115.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), ***Advances in Neural Information Processing Systems 26***: 3111–3119. Curran Associates, Inc.

Navis, C., & Glynn, M. A. 2010. How New Market Categories Emerge: Temporal Dynamics of Legitimacy, Identity, and Entrepreneurship in Satellite Radio, 1990–2005. ***Administrative Science Quarterly***, 55(3): 439–471.

Negro, G., & Leung, M. D. 2013. “Actual” and Perceptual Effects of Category Spanning. ***Organization Science***, 24(3): 684–696.

Nigam, A., & Ocasio, W. 2010. Event Attention, Environmental Sensemaking, and Change in Institutional Logics: An Inductive Analysis of the Effects of Public Attention to Clinton’s Health Care Reform Initiative. ***Organization Science***, 21(4): 823–841.

Nosofsky, R. M., & Zaki, S. R. 2002. Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. ***Journal of Experimental Psychology. Learning, Memory, and Cognition***, 28(5): 924–940.

Paolella, L., & Durand, R. 2016. Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market. ***Academy of Management Journal***, 59(1): 330–351.

Paolella, L., & Sharkey, A. 2017. Forging Consensus: An Integrated View of How Categories Shape the Perception of Organizational Identity. In R. Durand, N. Granqvist, & A. Tyllström (Eds.), ***Research in the Sociology of Organizations***, vol. 51: 327–353. Emerald Publishing Limited.

Pennington, J., Socher, R., & Manning, C. D. 2014. Glove: Global vectors for word representation. ***In EMNLP***.

Pontikes, E. G. 2012. Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences’ Evaluations. ***Administrative Science Quarterly***, 57(1): 81–118.

Pontikes, E. G., & Barnett, W. P. 2015. The persistence of lenient market categories. ***Organization Science***, (5): 1415.

Pontikes, E. G., & Kim, R. 2017. Strategic Categorization. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 71–111. Emerald Publishing Limited.

Porac, J. F., Thomas, H., & Baden-Fuller, C. 1989. Competitive Groups as Cognitive Communities: The Case of Scottish Knitwear Manufacturers*. ***Journal of Management Studies***, 26(4): 397–416.

Rao, H., Monin, P., & Durand, R. 2005. Border Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy. ***American Sociological Review***, 70(6): 968–991.

Reed, S. K. 1972. Pattern recognition and categorization. ***Cognitive Psychology***, 3(3): 382–407.

Rosa, J. A., Porac, J. F., Runser-Spanjol, J., & Saxon, M. S. 1999. Sociocognitive Dynamics in a Product Market. ***Journal of Marketing***, 63: 64.

Rosch, E. 1975. Cognitive representations of semantic categories. ***Journal of Experimental Psychology: General***, 104(3): 192–233.

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

Ruef, M., & Patterson, K. 2009. Credit and Classification: The Impact of Industry Boundaries in Nineteenth-Century America. ***Administrative Science Quarterly***, 54(3): 486–520.

Schneiberg, M., & Berk, G. 2010. From Categorical Imperative to Learning by Categories: Cost Accounting and New Categorical Practices in American Manufacturing, 1900–1930. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 255–292. Emerald Group Publishing Limited.

Sharkey, A. J. 2014. Categories and Organizational Status: The Role of Industry Status in the Response to Organizational Deviance. ***American Journal of Sociology***, 119(5): 1380–1433.

Sharpe, W. F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. ***The Journal of Finance***, 19(3): 425.

Smith, E. B. 2011. Identities as lenses: How organizational identity affects audiences' evaluation of organizational performance. ***Administrative Science Quarterly***, 56(1): 61–94.
Stambaugh, R. F., Yu, J., & Yuan, Y. 2015. Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle: Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle. ***The Journal of Finance***, 70(5): 1903–1948.
Vergne, J.-P. 2012. Stigmatized Categories and Public Disapproval of Organizations: A Mixed-Methods Study of the Global Arms Industry, 1996–2007. ***Academy of Management Journal***, 55(5): 1027–1052.
Vergne, J.-P., & Wry, T. 2014. Categorizing Categorization Research: Review, Integration, and Future Directions. ***Journal of Management Studies***, 51(1): 56–94.
Weber, K., Heinze, K. L., & DeSoucey, M. 2008. Forage for thought: Mobilizing codes in the movement for grass-fed meat and dairy products. ***Administrative Science Quarterly***, 53(3): 529–567.
Wittgenstein, L. 1953. ***Philosophical Investigations***. Paris: Blackwell Publishers.
Zhao, E. Y., Ishihara, M., Jennings, P. D., & Lounsbury, M. 2018. Optimal Distinctiveness in the Console Video Game Industry: An Exemplar-Based Model of Proto-Category Evolution. ***Organization Science***, 29(4): 588–611.
Zuckerman, E. W. 1999. The Categorical Imperative: Securities Analysts and the Illegitimacy Discount. ***American Journal of Sociology***, 104(5): 1398–1438.
Zuckerman, E. W. 2004. Structural Incoherence and Stock Market Activity. ***American Sociological Review***, 69(3): 405–432.
Zuckerman, E. W. 2017. The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 31–68. Emerald Publishing Limited

CHAPTER 1
The (Relative) Effects of Typicality on Volatility: A Study Using Word Embeddings

Abstract [167 words]

Studies on market categories generally hold that they tend to stabilize audiences' valuations. Yet, recent results suggest that audiences' valuations may vary substantially even in the presence of established market categories. This apparent contradiction calls for a re-examination of the stabilizing role of market categories. This paper answers this call by studying the relationship between the typicality of publicly listed firms and the volatility of their stock prices. It finds that typical firms experience lower volatility as investors can rely on information encoded in industry prototypes to value them. However, this effect is weaker in ambiguous industry categories, as the information encoded in prototype is more open to interpretations. I measure typicality using a new method that uses word embeddings to represent firms in a shared semantic space. To this purpose, I rely on a collection of over 100,000 annual reports. This paper has implications for the literature on market categories, for the literature on organizational approaches to financial markets, and for computational approaches to organizational phenomena.

Introduction

Market participants group entities together into categories, and how typical an entity is within its category has an impact on audiences' valuations. In general, audiences value typical entities more positively as they are easier to make sense of (Hsu, 2006; Hsu, Koçak, & Hannan, 2009; Kovács & Hannan, 2015). Typical entities also have stronger appeal in the eyes of audiences, providing that audiences value existing categories positively (Hannan et al., 2019). Furthermore, atypical entities are more likely to be overlooked and left out of audiences' consideration set (Zuckerman, 1999). Thus, market categories tend to have a stabilizing role on audiences' valuations: typical entities are more likely to thrive while atypical ones are discounted and weeded out.

Recent research, however, challenge this view and suggest that audiences' valuations can vary substantially even in -or because of- the presence of market categories (Durand & Thornton, 2018; Schneiberg & Berk, 2010). For instance, some audiences with specific goals or cultural aspirations, value atypical entities more positively, not more negatively (Goldberg, Hannan, & Kovacs, 2016; Paolella & Durand, 2016; Zuckerman, 2017). Furthermore, the meaning of categories themselves change over time, leading to re-evaluations of their members (Delmestri & Greenwood, 2016; Kennedy, Lo, & Lounsbury, 2010). Finally, audiences' perceptions and valuations of typical entities is influenced by contextual factors such as the ambiguity of existing categories (Kovács & Hannan, 2010, 2015). The stabilizing role of market categories on audiences' valuations is thus called into question.

To re-ascertain the nature of the relationship between typicality and the stability of audiences' valuations, this paper considers the effects of typicality and industry category ambiguity on the volatility of the stock price of publicly listed firms in the U.S. Typical firms are similar to their category's prototype, which encodes general information on members of the category. This information is most relevant to value typical entities as they resemble the prototype of their category. By contrast, it is less relevant when used to value atypical entities, which tend differ from the prototype. Hence, investors can leverage the information encoded in industry category prototypes to value typical firms while the same information becomes useless or even an impediment to the valuation of atypical firms. Typical firms are thus less exposed to uninformed investors, who might over- or underestimate firms' value, leading to less volatility (De Long, Shleifer, Summers, & Waldmann, 1990). However, this effect depends on the level of categorical ambiguity. In ambiguous industry categories, the prototype is a loose combination of ill-defined features. Thus, the information encoded in the prototype loses some of its relevance to value typical firms. Therefore, categorical ambiguity attenuates the negative relationship between typicality and volatility.

This paper tests these ideas using Natural Language Processing on business overview sections extracted from of a large corpus of annual reports to measure typicality and categorical ambiguity. This allows to represent firms in a shared semantic space, to create vectors representing industry prototypes and to straightforwardly measure typicality as the similarity of a firm's vector to its industry prototype. Categorical ambiguity is measured using the median level of typicality of firms within an industry. The proposed approach has several advantages. It does not rely on analysts' coverage, which might be influenced by firm size or other variables. It leverages semantic information encoded in all the words used to describe a firm's activity and not solely that contained in category labels, which firms use strategically to blur their positioning. It does not require potentially biased human coding or judgment. Results largely support the proposed theory. Firms with a high level of typicality experience lower standard deviation in their monthly or daily returns in the following year. This effect is attenuated by the level of categorical ambiguity. These results are robust to the use of more fine-grained levels of definition for industries and to the use of different measures of volatility and categorical ambiguity.

This article contributes to three main strands of research. It first contributes to the literature on market categories. It shows that typicality is associated with less volatility in stock prices and that the strength of this association depends on the level of ambiguity within the industry. It thus suggests that the stabilizing role of market categories on audiences' valuations is contingent on the presence or absence of ambiguous categories. Second, this article contributes more broadly to organizational approaches to financial markets. It imports new insights from the finance literature on volatility and relate it to the organizational literature on categories to link typicality and ambiguity to volatility. Third, this article contributes to the advancement of computational approaches to organizational phenomena by introducing a new method to measure similarities between organizations using NLP.

Typicality and valuation: current developments and limitations

Categories are the symbolic and material attributes of products, firms, and industries that are both shared among actors and that distinguish these entities from others (Durand & Thornton, 2018). A category is a fuzzy set of entities, where the degree of membership of an entity into the category is a function of its similarity to the category prototype (Durand & Paolella, 2013; Mervis & Rosch, 1981; Rosch & Mervis, 1975). Depending on the context being studied, entities and categories may be films and movie genres, restaurants and types of cuisine, members of an industry and the industry itself (Hsu, 2006; Kennedy, 2008; Porac, Thomas, & Baden-Fuller, 1989; Rao, Monin, & Durand, 2005). The prototype of a category consists of an abstract representation of the 'average' member of a category (Reed, 1972) ; it is an abstraction upon which one relies to identify entities as being members of the category, based on similarity or 'family resemblance' (Mervis & Rosch, 1981; Wittgenstein, 1953). Similarity to the prototype is generally understood as resulting from the overall alignment of the features of a specific entity with the features of the prototype of the entity's category.

Established theory and results suggest that audiences discount atypical entities for two related reasons. First, typical entities are easier to make sense of; audiences can readily interpret information about them using the models and schemas tied to the category (Hsu, Hannan, & Pólos, 2011; Zuckerman, 2004). Second, some categories are intrinsically appealing to audiences, which thus view typical instances of these categories more positively (Hannan et al., 2019). Since audiences generally favour typical entities and shun atypical ones, categories generally have a stabilizing role in markets, ensuring that deviants are weeded out and that audiences converge in their assessments of existing entities (Hannan et al., 2019; Zuckerman, 1999).

A burgeoning strand of research questions these ideas, suggesting that categories can induce variability in audiences' valuations. In some settings, some audiences have an inclination toward atypical offerings. For example, clients of law firms prefer those which offer a wide range of different legal services as these firms are better able to meet their various needs in a complex legal environment (Paolella & Durand, 2016; Paolella & Sharkey, 2017). In the context of movies and restaurants, although audiences exhibiting a high variety in their tastes have a stronger inclination toward typical entities, other audiences have a preference for entities mixing features of different categories (Goldberg, Hannan, et al., 2016). More generally, typical offerings are appealing to audiences only when they seek "minimally satisfying performance" (Zuckerman, 2017).

Categories themselves are not 'set in stone' and their meanings might change over time, which further questions their stabilizing role. For example, Grappa was long seen as a low-brow alcohol but progressively became a high status beverage thanks to the purposeful action of a dedicated group of producers (Delmestri & Greenwood, 2016). Researchers have thus proposed that categories have varying 'currency' or 'viability' which in turn induces variability in audiences' valuations of their members (Kennedy et al., 2010; Lo, Fiss, Rhee, &

Kennedy, 2019). Finally, the effects that category can have on audiences' valuations is contingent on other factors which themselves might be unstable, such as the ambiguity of the category (Kovács & Hannan, 2010) or the status of its members or of the category itself (Phillips & Zuckerman, 2001; Sharkey, 2014). In view of these results, it seems important to re-ascertain the stabilizing role of market categories and the boundary conditions that may impact it.

To do so, this paper studies the impact of publicly listed firms' typicality on the volatility of their stock prices. It reasons that if investors have an easier time making sense of firms which are typical of their industry category thanks to the information encoded in prototypes, then typical firms will be less exposed to uninformed investors, leading to smaller volatility. It introduces recent developments in the finance literature which support this view. It then argues that if the relationship between typicality and volatility hinges on the relevance of the information encoded in industry category prototypes to value typical entities, then it will be attenuated in ambiguous industry categories where the information encoded in prototypes is of poorer quality.

The impacts of typicality and categorical ambiguity on volatility

Although volatility is not often studied in the management literature, it has a strong impact on financial markets and reflects investors' difficulty in valuing a firm (Zuckerman, 2004) as well as – and relatedly – their confidence in its future performance (Bansal & Clelland, 2004; Harrison, Thurgood, Boivie, & Pfarrer, 2019). In finance, the capital asset pricing model identifies two antecedents of the volatility of a firm's returns: 1) the correlation between the firm's returns and the returns of an efficient portfolio of assets and 2) firm-specific (idiosyncratic) volatility in stock prices (Sharpe, 1964). Investors and funds actively manage volatility and base their investment decisions in part on this variable. For example, the asset manager BlackRock proposes several funds aiming specifically at low volatility stocks, as they would tend to lose less in case of market correction. Volatility limits arbitrage as investors who adopt a contrarian strategy on volatile stocks face considerable uncertainty regarding the point at which the stock price trend will revert and allow them to profit from their position (De Long et al., 1990; Pontiff, 2006; Stambaugh, Yu, & Yuan, 2015). Volatility thus plays a key role in financial markets, representing both an opportunity to achieve superior returns and a limit to market efficiency (Zuckerman, 2004).

One important antecedent of volatility is firms' exposure to uninformed investors. Uninformed investors trade on 'noise', i.e. they trade on signals which they falsely believe reveal something on firms' fundamentals (Black, 1986). Provided uninformed investors hold the same optimistic or pessimistic expectations, they can drive prices away from fundamentals for an extended period of time, until their sentiment shifts and prices revert to their mean (De Long et al., 1990). Thus, greater exposition to uninformed trader, leads to a greater volatility in prices. Finance research often assimilates uninformed traders to retail investors, who are assumed to have a reduced ability to distinguish noise from information (Brandt, Brav, Graham, & Kumar, 2010; Foucault, Sraer, & Thesmar, 2011). However, any investor may behave as an uninformed trader as long as she does not have access to relevant information on fundamentals. Thus, in general, volatility is associated with greater mispricing of stocks by uninformed investors (Aabo, Pantzalis, & Park, 2017).

Drawing on these insights from finance research, typical entities experience less volatility in stock prices because they are less exposed to uninformed investors. Based on the socio-cognitive literature on categories, category prototypes encode general information about the most representative members of a category that guide the interpretation of category members and set the expectations of audiences (Cantor, Mischel, & Schwartz, 1982; Reed, 1972; Rosa, Porac, Runser-Spanjol, & Saxon, 1999; Rosch & Mervis, 1975). The information encoded in the prototype is most relevant to value category members which are representative of the category, i.e. typical members. Hence, audiences benefit from an additional source of information to assess typical entities, the prototype of the category itself. It is thanks to this supplementary source of information that typical entities appear easier to make sense of (Hannan et al., 2019; Leung & Sharkey, 2014; Negro & Leung, 2013). The information encoded in the prototype is less relevant to value atypical firms since they depart significantly from it. Furthermore, investors may nonetheless consciously or unconsciously rely on prototypes when assessing atypical firms and as a result trade on noise rather than on relevant information. Hence, *ceteris paribus*, typical firms are generally less exposed to uninformed investors and thus less prone to be over- or undervalued. This results in lower stock price volatility:

> **Hypothesis 1:** *The greater the typicality of a firm, the lower the subsequent volatility in its stock price*

Typicality thus has a direct effect on volatility. However, the proposed account is so far incomplete because it focuses on each firm's positioning relative to the industry prototype (typicality), independently of the positioning of other members of the category relative to the prototype. However, category research repeatedly demonstrated that how category members are distributed within a category have significant impacts on each member. Several constructs in the literature try to capture this notion. Contrast refers to the extant to which members of a category also belongs to other categories (Kovács & Hannan, 2010, 2015). Leniency is a function of both contrast and the number of distinct categories straddled by category members (Pontikes, 2012; Pontikes & Barnett, 2015). Categorical homogeneity (or heterogeneity) refers to the spread of category members around the prototype (Haans, 2019). Categorical coherence refers to the degree of similarity or family resemblance among members (Lo et al., 2019). Other researchers use the term ambiguity to refer to similar constructs (Granqvist, Grodal, & Woolley, 2013; Ruef & Patterson, 2009). References to ambiguity are also made when discussing leniency or contrast.

In the literal sense, ambiguity refers to a situation in which something has multiple interpretations and may therefore cause confusion. Ambiguity thus seems to be the most general and well-suited term to refer to a situation where members of a category are generally dissimilar to its prototype. When members of a category are generally dissimilar from the prototype, it cannot have a well-identified and agreed upon meaning -what constitutes a 'representative' member of the category is unclear. The prototype is necessarily open to interpretations and may lead to confusion. When members of a category are generally similar to the prototype, they are all clustered around it, it has a well-identified and agreed upon meaning -what constitutes a 'representative' member of the category is clear. The prototype is not open to interpretations and does not lead to significant confusion. If we take the prototype of the category to capture its meaning -i.e. what it means to be a member of the category- then ambiguous categories are indeed those where members are generally dissimilar from the prototype.

Ambiguous categories have ill-defined prototypes and unclear boundaries, grouping entities with little in common. For these reasons, ambiguous categories are in general discounted by audiences. For example, book readers and restaurant goers tend to favour products that do not blend conceptually distant categories, while firms belonging to ambiguous industries have less facility obtaining a credit (Kovács & Hannan, 2015; Ruef & Patterson, 2009). However, for the same reasons, some producers prefer ambiguous market categories as they offer more flexibility (Pontikes & Barnett, 2015). More generally, managers of a firm have more freedom in claiming membership in ambiguous categories and do so strategically, based on their perceptions of their labels (Granqvist et al., 2013). Producers spanning ambiguous categories are less discounted by audiences than those spanning unambiguous categories (Kovács & Hannan, 2010). Finally, some audiences with specific goals sometimes prefer ambiguous labels. For instance, venture capitalists are inclined toward firms associated with lenient labels (Pontikes, 2012).

That ambiguous categories have ill-defined prototypes imply that the information that they encode is of a reduced quality. When the ambiguity in a category is high, the prototype of the category summarizes a loose combination of features, and the boundaries of the category are eroding or ill-defined (Granqvist, Grodal, & Woolley, 2013; Pontikes & Barnett, 2015; Rao et al., 2005). In this situation, the prototype does not encode as much relevant information to value typical entities. In the setting of publicly listed firms, this means that typical firms are more prone to be exposed to uninformed investors in ambiguous categories than in unambiguous ones. Hence, we have:

> **Hypothesis 2:** *The ambiguity of a firm's industry attenuates the negative relationship between its typicality and the volatility of its stock price*

Data

To test the ideas developed above, I gathered data on publicly listed firms from 1995 to 2018 from Compustat fundamental annuals database and data on daily security prices from CRSP. I downloaded the annual report – the form 10-K – of all firms in the sample from the Security and Exchange Commission website, as the business overview section of annual reports is used to measure typicality. I managed to automatically extract a total of 82,797 business overview sections from 106,772 annual reports published between 1995 and 2018. I relied on this entire corpus to train the Natural Language Processing model used as part of the measurement of typicality as it needs a vast amount of data to train on. However, after this first step, I focused specifically on firms which are not financial institutions (do not have a SIC code between 6000 and 6999), are not the subsidiary of any other firms, and are listed in the NASDAQ, the American Stock Exchange or the New-York Stock Exchange. I also focused on firm-year observations for the period 1996 to 2018, as annual reports from the preceding year enter into the computation of industry prototypes. Finally, I also focus on firms in an industry with more than five members in a given year.

Measuring typicality using NLP on the business overview section of annual reports

General overview. This paper measures the typicality of publicly listed firms using Natural Language Processing (henceforth NLP) on the business overview section of the form 10-K (or annual report) published every year by publicly listed firms in the U.S. The general idea behind the proposed method is to represent firms in a semantic space so that similar firms are close to one another in the space, and then use vectors associated to firms to construct industry prototypes and measure firms' typicality.

An algorithm written specifically for this purpose downloaded a total of 106,772 annual reports published between 1995 and 2018 from the SEC website. It then automatically extracted the business overview sections from these reports. Due to some reports not being machine readable, the final corpus size is of 82,797 business overview sections. Following some pre-processing, a word embedding model trained on this data. Word embedding models are NLP models which learn how to represent the meaning of words into a semantic space so that words with similar meanings are located in the same region of the space (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013; Pennington, Socher, & Manning, 2014). Each publicly listed firm recorded in Compustat in a given year between 1995 and 2018, with a SIC code and to which a form 10-K could be associated for this year, was represented as a vector by linearly combining the words contained in its business overview section. Prototypes for industries were then created at the level of the two digits SIC code by taking the centroid or average of the vectors of all firms belonging in the industry. Typicality was finally computed using these prototypes.

In the following paragraphs, I explain the advantages of the proposed method, detail how documents were pre-processed, word embeddings trained, and firms represented in a semantic space, before describing how prototypes were constructed and typicality measured.

Limitations of existing methods to measure typicality. Typicality is traditionally measured in two distinct ways: based on analysts' coverage of firms (Zuckerman, 1999, 2004) or based on the category labels associated to firms or products. The intuition behind the first measure is that firms which are more central to their industry will tend to be covered by industry specialists while peripheral members will not. One limitation of this measure is that it depends on covariates that predict this variable. For instance, industry specialists may tend to cover larger firms although they diverge from their industry prototype and fail to cover small firms, irrespective of their typicality. Measures based on categorical affiliations attributed to or claimed by firms have limitations of their own. Measures based on categorical claims made by firms through the use of certain labels or names (Kennedy, 2008; Pontikes, 2012; Pontikes & Barnett, 2015) might not accurately reflect categorical affiliations. A firm which is very typical of its category can actively avoid using the category's label to try to distinguish itself from others or, conversely, atypical firms might use the category label intensely to create a sense of typicality. Measures based on the attribution of category labels to firms or products by third parties (Goldberg, Hannan, et al., 2016; Hsu et al., 2009; Kovács & Hannan, 2015) are less exposed to this issue. However, both approaches assimilate atypicality to category spanning as they hold that the typicality of an entity is inversely proportional to the number of (conceptually distant) categories that it spans. Thus, entities associated to a single label or to conceptually related labels are the most typical of their categories. However, one may be atypical of one's category without being associated with any other category. For example, both the penguin and the platypus are atypical for their categories (resp. birds and mammals) but only the platypus straddle the two categories through its possession of bird-like features (such as having a beak and laying eggs). It would be an appealing feature for a measure of typicality that even entities claiming belonging into a single category but being atypical for it would be identified as such. Measures of typicality based on human or experts' assessment could be an option but would be hard to obtain for a very large sample and can be biased.

This paper proposes an approach to measuring typicality which overcome these hurdles. It leverages the content of *all* the words that are used by firms when describing their activities in the business overview section of their annual reports to assess typicality. Firms using the same words end up being identified as similar to one another, irrespective of analysts' coverage. Similarities and differences in overall word uses determine typicality, so that firms using unusual words in their annual reports will be identified as atypical even if they do not claim membership into multiple categories. The labels used by firms have a reduced importance and hidden similarities are more likely to be revealed than when focusing solely on a restricted set of category labels. Finally, leveraging similarities in word uses to measure typicality allows to measure similarity without factoring in potentially biased human judgments and at a large scale.

Using word embedding models to measure similarity between firms. Following an approach pioneered by Hoberg and Phillips, bag-of-words[1] representations of the business overview section of annual reports can be used to measure similarity between firms and identify product-based industries (Hoberg & Phillips, 2010, 2016). The main idea behind this method is to use the words contained in the business overview section of annual reports to represent firms as vectors in a semantic space to then be able to compute similarities between them. Firms which tend to use the same words in their business overview sections, which are likely similar, will be located close to one another in the space, while firms using different words will be located far from one another in the space.

[1] A bag-of-word is a representation of a document as an unordered count of the words within it.

Under a bag-of-word approach, each business overview section in the available collection is transformed into a vector through the following steps. First, a semantic space with a number of dimensions equal to the size of the vocabulary used for the analysis is created. Each dimension is associated to a unique word in the vocabulary. For example, if one uses a vocabulary of 10,000 words, the semantic space will have 10,000 dimensions, each dimension corresponding to one of these 10,000 words. Second, each business overview section is represented as a vector whose coordinate along each dimension corresponds to the frequency of the word associated to that dimension in the business overview section. It is then possible to measure the similarity between two firms using the vectors representing their business overview sections. In this paper, I use cosine similarity to measure similarity between documents, which is a common measure used in NLP for this purpose (Jurafsky & Martin, 2009). Cosine similarity is a measure of the angle between two vectors and ranges from -1 to 1.

One limitation of the bag-of-word approach is that it does not take into account semantic relatedness between words and thus does not capture accurately similarities between firms. As an illustration, let's consider a universe with only three firms, A, B and C, and with a vocabulary of only three words, 'good', 'positive' and 'bad'. Let's assume that A uses only the word 'good', B uses 'good' half the time and 'bad' half the time and C uses 'good' half the time and 'positive' half the time. A bag-of-words approach would find that the similarity between A and B is the same as the similarity between A and C[2]. Yet, due to the semantic relation between all the three words, one would expect C to be more similar to A than B and B to be dissimilar from A and C. To be able to achieve this kind of fine grain measurement, one would have to map A, B and C to a space where the words 'good' and 'positive' are close to one another, and the word 'bad' far from both of them.

[2] Indeed, cosine similarity((1,0,0),(0.5, 0, 0.5))= cosine similarity((1,0,0),(0.5, 0.5, 0))

This is what an approach based on word embeddings do. Word embeddings are vectors representing words in a semantic space, so that words with related meanings are close to one another in the space. Word embeddings are learnt using what is called a word embedding model, trained on a large corpus. The model learns to represent words in a semantic space either by trying to predict neighbouring words given a focal word (Mikolov, Sutskever, et al., 2013; Pennington et al., 2014) or by predicting a target word given neighbouring words (Mikolov, Chen, et al., 2013). This paper uses the latter approach. Vectors learnt by word embeddings models capture meaningful semantic relations which can be represented using simple vector manipulations (Mikolov, Chen, et al., 2013; Mikolov, Sutskever, et al., 2013; Pennington et al., 2014).

Once a model has learnt word embeddings, they can be used to map the bag-of-words representation of a document to the embedding space. A straightforward way to do this is to represent each document as a linear combination of the embeddings of the words contained in it, weighing each word embedding by the frequency of the associated word in the document.

Pre-processing and training word embedding models. The business overview sections used to train word embedding models were first pre-processed as is common in NLP studies. All inserted tables, images, pdfs, and html code were stripped out of each document. Documents were then tokenized, i.e. reduced to a list of words, and stopwords, i.e. very common words, were excluded. Then, a vocabulary for the entire corpus was created. It includes the 20,000 words which are the most common in the business overview sections of annual reports after excluding stop words. Word embeddings models trained only on words included in the vocabulary. These 20,000 most common words represented 98% of the original words.

The Python library gensim was used to train the word embedding model on the business overviews corpus. When using this Python library, the user is free to set a number of parameters: the size of the window around a target word used to train the model, the number of dimensions of the semantic space in which words will be represented, the objective function used for learning and the extent to which frequent words are downsampled. Following common practice using these models (Mikolov, Chen, et al., 2013; Pennington et al., 2014), the window size was set to 5 and the number of dimensions to 300 while all other parameters where left to the default choice. The model then trained and learnt vectors representing words in the associated vocabulary. Table 1 illustrates the ability of the model to capture semantic relations by showing some selected words along with their 10 most similar words based on word-vectors learnt by the model.

-- *Insert Table 1 about here* --

Representing firms as vectors based on the business overview section. A vector was associated to each firm in the sample for any given year at which the firm published an annual report. The vector associated to a firm in a given year represents its position vis-à-vis other firms in the semantic space in this year based on the content of the business overview section of its annual report. Vectors were constructed as follows:

$$Vector_{f,y} = \sum_{w \in BusinessSection_{f,y}} P_{BusinessSection_{f,y}}(w) * Embbeding_w$$

Where $Vector_{f,y}$ is the vector representing firm f based on the business overview section of its annual report in year y. $BusinessSection_{f,y}$ is the set of words contained in the business overview section of firm f's annual report in year y. $P_{BusinessSection_{f,y}}(w)$ is the frequency of word w in this business overview section and $Embbeding_w$ is the embedding associated to word w in the embeddings model trained on business overview sections.

Creating prototypes and measuring typicality. Industry prototypes were created at the level of the first two digits SIC code associated to each firm in Compustat for each year. The prototype of a given industry for a given year is the centroid of the vectors of all firms that belonged to the industry in the preceding year:

$$Prototype_{I_y} = \frac{1}{|I_{y-1}|} \sum_{g \in I_{y-1}} Vector_{g,y-1}$$

$Prototype_{I_y}$ is the prototype of industry I_y, which is industry I in year y. I_{y-1} is the set of firms g such that g is in the two-digits SIC industry denoted by I in year y-1. $|I_{y\_1}|$ is the cardinal of this set. $Vector_{g,y-1}$ is the vector associated to g in year y-1 based on the business overview section of its annual report in year y-1.

The typicality of each firm is measured as the cosine similarity of the firm's corresponding vector to its corresponding industry prototype:

$$Typicality_{f,y} = \text{cosine similarity } (Vector_{f,y}, Prototype_{I_{f_y}})$$

$Typicality_{f,y}$ is the typicality of f in year y. $Prototype_{I_{f_y}}$ is the prototype of industry I_f, the industry to which firm f belong, in year y. $Vector_{f,y}$ is the vector associated to firm f based on the business overview section of its annual report, in year y.

Table 2 shows the top 10 words which are the most similar to industry prototypes for the 10 most represented industries in the sample in the year 2017. The words which are the most similar to industry prototypes based on cosine similarity are consistently those that are related to the industry. Table 3 shows the five most typical firms as well as the five least typical firms for the business services industry in the year 2017. The most typical firms are all cloud-based services or related to data management, reflecting dominant trends in business services. The most atypical firms exercise activities which are not representative of business services in 2017 – such as leasing aircrafts or containers –, operate on small segments – such as art trading in certain geographic areas – or in multiple sectors. The proposed measure of typicality thus seems able to correctly discriminate between typical and atypical firms.

-- Insert Table 2 and 3 about here --

Models and variables

Dependent variables. The volatility of returns is measured as their standard deviation over a set period of time (Foucault et al., 2011; French & Roll, 1986; Zuckerman, 2004). The main analysis of this paper uses the standard deviation in montlhy returns over the next year as a measure of the volatility of a firm's stock price:

$$Standard\ deviation\ of\ monthly\ returns\ over\ the\ next\ year_{f,y} = \sqrt{\frac{\sum_{m\ in\ month\ of\ y+1}(ret_m - \left(\frac{1}{12}\sum_{m\ in\ month\ of\ y+1} ret_m\right))^2}{12}}$$

I also use the standard deviation in daily returns as a complementary measure, as well as alternative measures of volatility in robustness checks. Daily volatility is more sensible to daily noises that may influence stock prices. All these measures are winsorized at the bottom and top 0.5 % to mitigate the influence of outliers.

Independent variables. The measure of typicality introduced earlier is the main independent variable. Typicality captures the similarity of a firm to its two-digits SIC code industry prototype. The typicality of a firm in a given year is based on its annual report for the year – from which the fundamentals data from Compustat also comes.

In line with hypothesis 2, the effect of typicality is moderated by the ambiguity of its two-digits SIC code industry. I measure ambiguity within a focal firm's two-digits SIC code industry by considering the typicality of all firms recorded as belonging to the industry in the current year, taking the median and then subtract it to 1:

$$Ambiguity\ Industry\ I\ in\ year\ y = 1 - Median_{g\ in\ I\ in\ year\ y}(Typicality_{g,y})$$

The intuition behind this measure is that the higher the median level of typicality in the industry, the more members of the industry are generally similar to the prototype, the less ambiguous is the industry. Both measures are centred and standardized to facilitate the interpretation of the interaction terms in subsequent analysis.

Controls. The size of a firm impact its future returns as bigger firms have less opportunities to grow while the book-to-market ratio of the firm is an indicator of the general risk to which it is exposed and of how optimistic investors are regarding its earnings prospects (Fama & French, 1992). Thus, both relate to volatility. Size is controlled for using the log of market capitalization. The book to market ratio is measured as the book value divided by market capitalization. A firm's profitability reduces the level of risk to which it is exposed and thus its future volatility (Harrison et al., 2019). It is also controlled for using the firm's returns on assets (ROA). The ROA and book-to-market ratio are winsorized at the top and bottom 0.5% to mitigate the influence of outliers. All models use firm fixed effects, which account for the influence of any unobserved firm-specific factors which do not vary over time. Year effects control for possible trend effects.

Descriptive statistics and correlations. Table 4 gives the descriptive statistics for all variables as well as correlations between variables. The typicality variable is skewed to the right, with some firms exhibiting high levels of dissimilarity from their prototype. However, the median level of ambiguity before standardization is 0.2 (on a scale from 0 to 2), suggesting that firms are relatively close to their industry prototypes, irrespective of their industry. There is a strong correlation between the standard deviation in monthly returns and the standard deviation in daily returns, as would be expected.

-- Insert Table 4 about here --

Model. All hypotheses were tested using panel regressions with firm fixed-effects, year effects and standard errors clustered by firms. The model used is of the following form:

$$\begin{aligned} Std\ Dev\ in\ Returns_{f,y} \\ &= \beta_0 + \beta_1 Typicality_{f,y} + \beta_2 Typicality_{f,y} \times Ambiguity\ in\ industry_{f,y} \\ &+ \beta X\ + \alpha_f + \alpha_y + \varepsilon_{f,y} \end{aligned}$$

Where $Std\ Dev\ in\ Returns_{f,y}$ is the volatility of firm f's stock price over year y+1, $Typicality_{f,y}$ is the typicality of firm f in year y, $Ambiguity\ in\ industry_{f,y}$ is the ambiguity of firm f's 2-digits SIC code industry in year y, X is a vector of controls, α_f is the firm fixed-effect, α_y is the year effect and $\varepsilon_{f,y}$ is the error term.

Main results and robustness checks

Table 5 presents the results. Models 1 to 5 use the standard deviation in monthly returns over the next year as a dependent variable and models 6 to 10 use the standard deviation of daily returns over the next year as a dependent variable. Model 1 shows the effects of control variables on the standard deviation in monthly returns. The size of the firm is negatively related with the standard deviation in monthly returns. This implies that bigger firms experience less volatility. Firms with a high book-to-market ratio being exposed to more risks, they experience greater volatility. Profitable firms also enjoy reduced volatility in the subsequent year. Model 2 introduces typicality as an antecedent of the standard deviation in monthly returns. It is found to be associated very significantly ($p < 0.001$) and negatively with the standard deviation in monthly returns. Model 3 looks at the effect of industry ambiguity on the standard deviation of monthly returns. It is positive and very significant ($p < 0.001$). This implies that firms in ambiguous industries generally experience more volatility. Model 4 looks at the simultaneous effects of typicality and ambiguity on volatility. Both are significantly associated with the standard deviation in monthly returns. In model 4, a one standard deviation increase in typicality is associated with a loss of 0.65 percent in the standard deviation of monthly returns. An increase of one standard deviation in ambiguity leads to a gain of 0.46 percent in volatility. In model 5, the interaction of typicality and industry ambiguity is positive and significant ($p<0.01$). When industry ambiguity is one standard deviation above the sample mean, a one standard deviation increase in typicality leads to a loss of 0.52 percent in the standard deviation of monthly returns. When industry ambiguity is one standard deviation below the sample mean, a one standard deviation increase in typicality leads to an even greater loss of 0.83 percent in the standard deviation of monthly returns. These results support hypothesis 1 and 2.

The results obtained with the standard deviation in daily returns over the next year provide further evidence that typicality is associated with reduced volatility. Note that observations are lost due to missing data on daily returns for some firms. Model 6 shows results which are largely similar to those of model 1. In model 7, typicality is associated negatively and very significantly ($p < 0.001$) with the standard deviation in daily returns. Model 8 shows a positive and significant ($p < 0.001$) relationship between industry ambiguity and the standard deviation in daily returns. Model 9 tests for the simultaneous effects of typicality and industry ambiguity on the standard deviation in daily returns. Typicality is negatively and very significantly associated with volatility ($p < 0.001$) while ambiguity is positively and significantly associated with it ($p < 0.001$). Finally, the interaction of typicality and ambiguity is positive and significant in model 10, as expected. These additional results lend further support to hypotheses 1 and 2. Figures 1 and 2 plot the effects of typicality on, respectively, the standard deviation in monthly returns and the standard deviation in daily returns, both on average and as a function of industry ambiguity.

-- *Insert Table 5 and Figure 1 and 2 about here* --

Robustness checks. I carried on robustness checks in unreported analysis. To ensure that results are not dependent on the choice of the measure used for volatility, I used two alternative measure of volatility : the systematic risk, or beta, which reflects how much the stock price of a firm is influenced by market-wide movements, and the idiosyncratic volatility, measured as the standard deviation of the residuals of a regression of daily returns on market returns (Harrison et al., 2019; Li, Rajgopal, & Venkatachalam, 2014). When using idiosyncratic volatility as a dependent variable, typicality has a negative and significant main effect (p-value < 0.001) but the interaction between typicality and ambiguity is not significant. When using beta as a dependent variable, typicality does not have a significant main effect but the interaction coefficient between typicality and ambiguity is positive and significant (p-value < 0.01). A graph of this interaction reveals that for low levels of ambiguity, there is no association between typicality and beta while for higher levels there is a negative association between the two. These results generally support the proposed theory. Another hurdle that the proposed theory encounters is that the strong interaction effects observed between typicality and ambiguity could be due to the presence of an inverted U-shape relationship between typicality and volatility. I tested for this alternative, introducing a quadratic term for typicality. When using the standard deviation of monthly returns as a dependent variable, the quadratic term is negative and significant but the lower bound of the confidence interval for the turning point is outside the data range, casting doubts on the validity of the relationship (Haans, Pieters, & He, 2016). When using the standard deviation of daily returns as a dependent variable, the quadratic term is not significant. Hence, there is little support for an inverted U-shape-based theory.

To ensure that results are not dependent on the use of the median in the moderator construction, I replaced the median level of typicality by the average. Findings remain unchanged to this alternative specification. I also ensured that the results were not dependent on the level of analysis used and created measures of typicality and ambiguity at the level of the 3-digits and 4-digits SIC code industries, excluding industries with less than 5 members in both cases. Results for hypothesis 1 and hypothesis 2 are unchanged when using 3-digits SIC code industries. Hypothesis 1 is supported when using 4-digits SIC code industries while hypothesis 2 is marginally supported (p-value = 0.057) when using the standard deviation in monthly returns as a dependent variable. Note that at this level of analysis, a significant number of observations is lost due to the exclusion of industries with less than 5 members.

Discussion

This paper has three main contributions. First, it contributes to the literature on categories in market. Second, it contributes more broadly to organizational approaches to financial markets. Finally, this paper contributes to the emergence of computational approaches to the study of organizations.

Contribution to the category literature. This study re-ascertain the stabilizing roles of market categories. At the same time, it shows that this stabilizing role is contingent on the presence or absence of ambiguous categories. In so doing, it makes three novel contributions to the category literature. First, this paper places the relevance of the information encoded in category prototypes to value typical category members at the heart of the mechanism linking typicality to the stability of audiences' valuations. It thus relates to both socio-cognitive and institutional approaches to categorization (Durand & Thornton, 2018) which hold that typicality implies more comprehensibility and predictability (Aldrich & Fiol, 1994; DiMaggio & Powell, 1983; Meyer & Rowan, 1977; Suchman, 1995). This paper provides a direct test of this intuition which, while well-established, has rarely been explored empirically.

Second, this article is the first to explicitly tie the effect of typicality on volatility to the ambiguity of categories. Previous research focused on the direct relationship between typicality and volatility (Zuckerman, 2004) or on the variability of producers' own valuation of their products as a function of their typicality (Hsu, Roberts, & Swaminathan, 2012). This article shades lights on how categorical ambiguity moderates the relationship between typicality and the stability of audiences' valuations. Theoretically, it relates this moderating role to the diminished informational relevance of the prototype. Empirically, it sees ambiguity as being the result of a limited propensity of category members to cluster around the prototype (Haans, 2019), rather than depending on the propensity of category members to span categories (Kovács & Hannan, 2010; Pontikes, 2012; Pontikes & Barnett, 2015).

Third, while extant research has been dedicated to study how market categories emerge, are maintained and disappear (Durand & Khaire, 2017; Navis & Glynn, 2010; Weber, Heinze, & DeSoucey, 2008), i.e. to study the volatility of category systems, little research studies how one's positioning within a category impacts subsequent volatility in one's valuation. This oversight is puzzling for two reasons. First, at the level of individual category members, volatility is an important variable to consider in many settings, such as public markets but also venture capital – as exemplified by recent controversies on the discrepancies between VCs' valuation of WeWork or Uber and that of the market as revealed during the pre-IPO phase. Second, at the level of an entire category, the volatility of the value of members of the category reflects the degree of stability of the category itself. For example, Khaire and Whadwani find that as the modern Indian art category stabilized, auction houses estimations of the range of expected value for works belonging to this category became narrower (Khaire & Wadhwani, 2010). Future research could benefit from further exploring these and other aspects of the link between categories and volatility.

Contribution to organizational perspectives on financial phenomenon. This article contributes to organizational approaches to financial markets (Cobb, Wry, & Zhao, 2016; Fleischer, 2009; Ioannou & Serafeim, 2015; Ruef & Patterson, 2009; Smith, 2011; Syakhroza, Paolella, & Munir, 2018; Zuckerman, 1999, 2004). Specifically, this paper develops a theory focusing on the informational consequences of typicality on volatility and produces new results supporting this theory. It proposes that typical firms are less exposed to uninformed investors and thus experience less volatility. Based on this argument, typical stocks experience greater volatility as a general tendency and not only following specific events such as quarterly earnings announcements (Zuckerman, 2004). Results support this hypothesis, expanding existing organizational research on cognitive legitimacy and information asymmetries in financial markets (Pollock & Rindova, 2003).

Furthermore, this article helps bridging organizational and financial perspectives on valuation by identifying how insights on uninformed investors stemming from the finance literature resonate with existing accounts of how prototypes contribute to valuation. While it is important to acknowledge that financial approaches to volatility are diverse and do not all converge on the proposition that greater volatility reflects greater exposition to uninformed investors, recent theorization and evidences give credit to this idea (Aabo et al., 2017; Li et al., 2014; Stambaugh et al., 2015). Although organizational and financial perspectives may diverge significantly in their underlying assumptions and perspectives (Zajac & Westphal, 2004), much is to be learnt from such selective coupling between the two disciplines, as the results presented in this paper illustrate.

Contribution to computational approaches to the study of organizations. While studies on typicality rely on analysts' coverage or category labels to measure typicality, the method used in this paper infers the position of a firm vis-à-vis others using the language contained in its annual reports. This approach to measuring typicality is less sensible to covariates determining analysts' coverage and less influenced by firms' strategic uses of labels as it does not focus specifically on them to measure typicality. It also does not rely on human judgments which might introduce biases in measurements. The proposed approach promotes a view of typicality as being revealed by language uses which go beyond the mere association of some entities with some labels. Firms that belong to the same category will tend to use a similar language to describe their activities, and firms which use an unusual language given their category are atypical firms. Beyond revealing similarities in the 'objective' features of the products and activities of firms, the proposed method locates firms vis-à-vis one another in a *semantic* space, where both 'objective' and semantic knowledge on the meaning of words is encoded. Thus, it accounts for the central role of language into defining what being typical of a category is.

More broadly, this paper follows suite recent investigations of phenomena of interest to strategist, organizational theorists and economic sociologists, such as cultural fit, innovation or distinctiveness (Goldberg, Srivastava, Manian, Monroe, & Potts, 2016; Haans, 2019; Kaplan & Vakili, 2015; Srivastava, Goldberg, Manian, & Potts, 2017), using advanced computational techniques. Computational techniques, and especially those coming from

Natural Language Processing, will become integral part of social scientists' toolbox. In this article, I tried to introduce word embeddings to the reader as a viable tool to represent texts in high dimensional space and then measure similarities and dissimilarities between firms more accurately than when using bag-of-words representations of the content of documents. I hope that this will catch the interest of some and further advance the development of computational approaches to organizational phenomena.

Finally, organization scholars are accustomed to thinking of firms and products as located in competitive 'spaces', and a vast literature tries to link the position of a firm or of a product in such spaces to its performance or valuation by audiences. As this paper illustrates, computational methods offer the opportunity to operationalize such theories and represent firms or products in shared spaces and directly 'observe' their positions relative to others. The approach proposed in this paper is not limited to the typicality setting and might be easily adapted to study other aspects of a firm's positioning. One could compare merging firms (Hoberg & Phillips, 2010), industries undergoing divestments to industries undergoing investments (Durand & Vergne, 2015), or one could create networks of semantic relatedness between firms. Representing firms in a semantic space offers a unique opportunity to measure constructs that were so far hard to measure and thus opens new spaces for research on firms' positioning. This is not to be understood solely in a methodological sense, as the methods that we use to study phenomenon influence our thinking about them.

Limitations and conclusion

This paper suffers from several limitation. First, selection issues might influence the results. Notably, I miss daily or monthly returns for some observations which are thus not used in the analyses of daily and monthly volatility. However, assuming that more typical firms as well as firms with lower variations in value tend to stay in the sample, this might mean that results underestimate rather than overestimate the negative effect of typicality on volatility. In any case, one must avoid making strong causal claims based on the findings contained in this paper. What it uncovers is a set of strong *associations* between typicality and volatility, which correspond to those predicted by the theory developed. It does so through the use of a novel Natural Language Processing technique, which holds many promises. Second, one may claim that two-digits SIC codes imperfectly capture the industry categories which are used by investors to classify publicly listed firms. This concern is mitigated by the analyses carried on at the 3-digits and 4-digits SIC code as part of the robustness checks. These industries are typically small and likely to gather firms which investors would identify as peers. Even if the industry categories that are used do not perfectly reproduce those used by investors, they certainly overlap, and it is likely that firms which are identified as highly typical or atypical of their 2-digits SIC code industry categories are indeed typical or atypical firms in the perceptions of investors. While these limitations need to be acknowledged, they should not lead to overlook the substantive association found between typicality as measured based on business overview sections of annual reports and volatility. Leveraging information encoded in words using NLP is a promising avenue to uncover relationships of interests to social scientists which were so far hard to grasp and might have an important impact on the economy or, more broadly, on society. This paper aims to bring us a little forward into this direction.

REFERENCES

Aabo, T., Pantzalis, C., & Park, J. C. 2017. Idiosyncratic volatility: An indicator of noise trading? ***Journal of Banking & Finance***, 75: 136–151.

Aldrich, H. E., & Fiol, C. M. 1994. Fools Rush in? The Institutional Context of Industry Creation. ***The Academy of Management Review***, 19(4): 645.

Bansal, P., & Clelland, I. 2004. Talking trash: Legitimacy, impression management, and unsystematic risk in the context of natural environment. ***Academy of Management Journal***, (1): 93.

Black, F. 1986. Noise. ***The Journal of Finance***, 41(3): 528–543.

Brandt, M. W., Brav, A., Graham, J. R., & Kumar, A. 2010. The Idiosyncratic Volatility Puzzle: Time Trend or Speculative Episodes? ***The Review of Financial Studies***, 23(2): 863–899.

Cantor, N., Mischel, W., & Schwartz, J. C. 1982. A prototype analysis of psychological situations. ***Cognitive Psychology***, 14(1): 45–77.

Cobb, J. A., Wry, T., & Zhao, E. Y. 2016. Funding Financial Inclusion: Institutional Logics and the Contextual Contingency of Funding for Microfinance Organizations. ***Academy of Management Journal***, 59: 2103–2131.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. 1990. Noise Trader Risk in Financial Markets. ***Journal of Political Economy***, 98(4): 703–738.

Delmestri, G., & Greenwood, R. 2016. How Cinderella Became a Queen: Theorizing Radical Status Change. ***Administrative Science Quarterly***, 61(4): 507–550.

DiMaggio, P. J., & Powell, W. W. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. ***American Sociological Review***, 48(2): 147–160.

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

Durand, R., & Paolella, L. 2013. Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory. ***Journal of Management Studies***, 50(6): 1100–1123.

Durand, R., & Thornton, P. H. 2018. Categorizing Institutional Logics, Institutionalizing Categories: A Review of Two Literatures. ***Academy of Management Annals***, 12(2): 631–658.

Durand, R., & Vergne, J.-P. 2015. Asset divestment as a response to media attacks in stigmatized industries: Asset Divestment as a Response to Media Attacks. ***Strategic Management Journal***, 36(8): 1205–1223.

Fama, E. F., & French, K. R. 1992. The Cross-Section of Expected Stock Returns. ***The Journal of Finance***, 47(2): 427–465.

Fleischer, A. 2009. Ambiguity and the equity of rating systems: United States brokerage firms, 1995–2000. ***Administrative Science Quarterly***, 54(4): 555–574.

Foucault, T., Sraer, D., & Thesmar, D. J. 2011. Individual Investors and Volatility. ***The Journal of Finance***, 66(4): 1369–1406.

French, K. R., & Roll, R. 1986. Stock return variances: The arrival of information and the reaction of traders. ***Journal of Financial Economics***, 17(1): 5–26.

Goldberg, A., Hannan, M. T., & Kovacs, B. 2016. What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. ***American Sociological Review***, (2): 215.

Goldberg, A., Srivastava, S. B., Manian, V. G., Monroe, W., & Potts, C. 2016. Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. ***American Sociological Review***, 81(6): 1190–1222.

Granqvist, N., Grodal, S., & Woolley, J. L. 2013. Hedging Your Bets: Explaining Executives' Market Labeling Strategies in Nanotechnology. ***Organization Science***, 24(2): 395–413.

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

Haans, R. F. J., Pieters, C., & He, Z.-L. 2016. Thinking about U: Theorizing and testing U- and inverted U-shaped relationships in strategy research. ***Strategic Management Journal***, 37(7): 1177–1195.

Hannan, M. T., Mens, G. L., Hsu, G., Kovács, B., Negro, G., et al. 2019. ***Concepts and Categories: Foundations for Sociological and Cultural Analysis*** (1 edition). New York: Columbia University Press.

Harrison, J., Thurgood, G. R., Boivie, S., & Pfarrer, M. 2019. Perception Is Reality: How CEOs' Observed Personality Influences Market Perceptions of Firm Risk and Shareholder Returns. ***Academy of Management Journal***. https://doi.org/10.5465/amj.2018.0626.

Hoberg, G., & Phillips, G. 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. ***Review of Financial Studies***, 23(10): 3773–3811.

Hoberg, G., & Phillips, G. 2016. Text-Based Network Industries and Endogenous Product Differentiation. ***Journal of Political Economy***, 124(5): 1423–1465.

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

Hsu, G., Hannan, M. T., & Pólos, L. 2011. Typecasting, Legitimation, and Form Emergence: A Formal Theory. ***Sociological Theory***, 29(2): 97–123.

Hsu, G., Koçak, Ö., & Hannan, M. T. 2009. Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests. ***American Sociological Review***, 74(1): 150–169.

Hsu, G., Roberts, P. W., & Swaminathan, A. 2012. Evaluative Schemas and the Mediating Role of Critics. ***Organization Science***, 23(1): 83–97.

Ioannou, I., & Serafeim, G. 2015. The impact of corporate social responsibility on investment recommendations: Analysts' perceptions and shifting institutional logics. ***Strategic Management Journal***, 36(7): 1053–1081.

Jurafsky, D., & Martin, J. H. 2009. ***Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition / Daniel Jurafsky, James H. Martin***. Upper Saddle River, N.J. : Pearson Prentice Hall, c2009.

Kaplan, S., & Vakili, K. 2015. The double-edged sword of recombination in breakthrough innovation. ***Strategic Management Journal***, 36(10): 1435–1457.

Kennedy, M. T. 2008. Getting Counted: Markets, Media, and Reality. ***American Sociological Review***, 73(2): 270–295.

Kennedy, M. T., Lo, J. Y.-C., & Lounsbury, M. 2010. Category currency: The changing value of conformity as a function of ongoing meaning construction. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 369–397. Emerald Group Publishing Limited.

Khaire, M., & Wadhwani, R. D. 2010. Changing Landscapes: The Construction of Meaning and Value in a New Market Category—Modern Indian Art. ***Academy of Management Journal***, 53(6): 1281–1304.
Kovács, B., & Hannan, M. T. 2010. The consequences of category spanning depend on contrast. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 175–201. Emerald Group Publishing Limited.
Kovács, B., & Hannan, M. T. 2015. Conceptual Spaces and the Consequences of Category Spanning. ***Sociological Science***, 2: 252–286.
Leung, M. D., & Sharkey, A. J. 2014. Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount. ***Organization Science***, 25(1): 171–184.
Li, B., Rajgopal, S., & Venkatachalam, M. 2014. R^2 and Idiosyncratic Risk Are Not Interchangeable. ***The Accounting Review***, 89(6): 2261–2295.
Lo, J. Y.-C., Fiss, P. C., Rhee, E. Y., & Kennedy, M. T. 2019. Category Viability: Balanced Levels of Coherence and Distinctiveness. ***Academy of Management Review***. https://doi.org/10.5465/amr.2017.0011.
Mervis, C. B., & Rosch, E. 1981. Categorization of Natural Objects. ***Annual Review of Psychology***, 32(1): 89–115.
Meyer, J. W., & Rowan, B. 1977. Institutionalized organizations: Formal structure as myth and ceremony. ***American Journal of Sociology***, 340–363.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. 2013. Efficient Estimation of Word Representations in Vector Space. ***ArXiv:1301.3781 [Cs]***. http://arxiv.org/abs/1301.3781.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), ***Advances in Neural Information Processing Systems 26***: 3111–3119. Curran Associates, Inc.
Navis, C., & Glynn, M. A. 2010. How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. ***Administrative Science Quarterly***, 55(3): 439–471.
Negro, G., & Leung, M. D. 2013. "Actual" and Perceptual Effects of Category Spanning. ***Organization Science***, 24(3): 684–696.
Paolella, L., & Durand, R. 2016. Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market. ***Academy of Management Journal***, 59(1): 330–351.
Paolella, L., & Sharkey, A. 2017. Forging Consensus: An Integrated View of How Categories Shape the Perception of Organizational Identity. In R. Durand, N. Granqvist, & A. Tyllström (Eds.), ***Research in the Sociology of Organizations***, vol. 51: 327–353. Emerald Publishing Limited.
Pennington, J., Socher, R., & Manning, C. D. 2014. Glove: Global vectors for word representation. ***In EMNLP***.
Phillips, D. J., & Zuckerman, E. W. 2001. Middle-Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets1. ***American Journal of Sociology***, 107(2): 379–429.
Pollock, T. G., & Rindova, V. P. 2003. Media legitimation effects in the market for initial public offerings. ***Academy of Management Journal***, 46(5): 631–642.
Pontiff, J. 2006. Costly arbitrage and the myth of idiosyncratic risk. ***Journal of Accounting and Economics***, 42(1–2): 35–52.
Pontikes, E. G. 2012. Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations. ***Administrative Science Quarterly***, 57(1): 81–118.

Pontikes, E. G., & Barnett, W. P. 2015. The persistence of lenient market categories. ***Organization Science***, (5): 1415.
Porac, J. F., Thomas, H., & Baden-Fuller, C. 1989. Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers*. ***Journal of Management Studies***, 26(4): 397–416.
Rao, H., Monin, P., & Durand, R. 2005. Border Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy. ***American Sociological Review***, 70(6): 968–991.
Reed, S. K. 1972. Pattern recognition and categorization. ***Cognitive Psychology***, 3(3): 382–407.
Rosa, J. A., Porac, J. F., Runser-Spanjol, J., & Saxon, M. S. 1999. Sociocognitive Dynamics in a Product Market. ***Journal of Marketing***, 63: 64.
Rosch, E., & Mervis, C. B. 1975. Family resemblances: Studies in the internal structure of categories. ***Cognitive Psychology***, 7(4): 573–605.
Ruef, M., & Patterson, K. 2009. Credit and classification: The impact of industry boundaries in nineteenth-century America. ***Administrative Science Quarterly***, 54(3): 486–520.
Schneiberg, M., & Berk, G. 2010. From Categorical Imperative to Learning by Categories: Cost Accounting and New Categorical Practices in American Manufacturing, 1900–1930. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 255–292. Emerald Group Publishing Limited.
Sharkey, A. J. 2014. Categories and Organizational Status: The Role of Industry Status in the Response to Organizational Deviance. ***American Journal of Sociology***, 119(5): 1380–1433.
Sharpe, W. F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. ***The Journal of Finance***, 19(3): 425.
Smith, E. B. 2011. Identities as lenses: How organizational identity affects audiences' evaluation of organizational performance. ***Administrative Science Quarterly***, 56(1): 61–94.
Srivastava, S. B., Goldberg, A., Manian, V. G., & Potts, C. 2017. Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations. ***Management Science***. https://doi.org/10.1287/mnsc.2016.2671.
Stambaugh, R. F., Yu, J., & Yuan, Y. 2015. Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle. ***The Journal of Finance***, 70(5): 1903–1948.
Suchman, M. C. 1995. Managing Legitimacy: Strategic and Institutional Approaches. ***The Academy of Management Review***, 20(3): 571–610.
Syakhroza, M. A., Paolella, L., & Munir, K. 2018. Holier than Thou? Identity Buffers and Adoption of Controversial Practices in the Islamic Banking Category. ***Academy of Management Journal***, 62(4): 1252–1277.
Weber, K., Heinze, K. L., & DeSoucey, M. 2008. Forage for thought: Mobilizing codes in the movement for grass-fed meat and dairy products. ***Administrative Science Quarterly***, 53(3): 529–567.
Wittgenstein, L. 1953. ***Philosophical Investigations***. Paris: Blackwell Publishers.
Zajac, E. J., & Westphal, J. D. 2004. The Social Construction of Market Value: Institutionalization and Learning Perspectives on Stock Market Reactions. ***American Sociological Review***, 69(3): 433–457.
Zuckerman, E. W. 1999. The Categorical Imperative: Securities Analysts and the Illegitimacy Discount. ***American Journal of Sociology***, 104(5): 1398–1438.
Zuckerman, E. W. 2004. Structural Incoherence and Stock Market Activity. ***American Sociological Review***, 69(3): 405–432.

Zuckerman, E. W. 2017. The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 31–68. Emerald Publishing Limited.

FIGURES

Figure 1. Main and moderated effects of typicality on the standard deviation in monthly returns

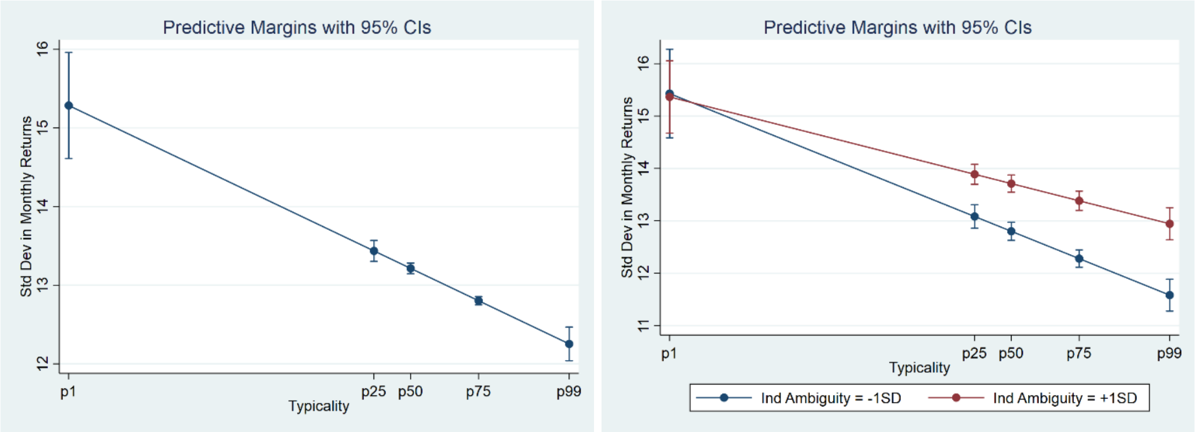
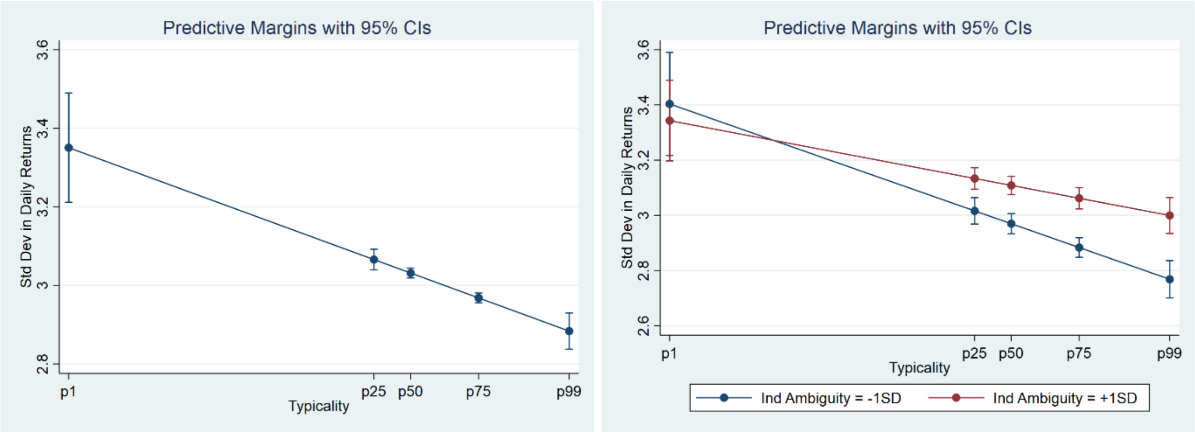


Figure 2. Main and moderated effects of typicality on the standard deviation in daily returns



TABLES

Table 1. Most similar words to selected words as evaluated using the word embedding model trained on business overview sections

managers	manager, executives, professionals, presidents, teams, specialists, management, salespeople, coordinators, staff
ceo	cfo, coo, founder, chief, cofounder, chairman, executive, emeritus, mr, nonexecutive
stakeholders	constituencies, constituents, stakeholder, advocates, empowering, perspectives, leaders, rewarding, empower, sustainability
owners	owner, operators, operator, sellers, developers, holders, purchasers, builders, buyers, lessees
debt	indebtedness, borrowings , debentures, notes, borrowing, financing, revolving, debts, financings, indentures
profit	profits, margins, margin, profitability, revenues, revenue, gross, income, earnings, net
losses	loss, liabilities, chargeoffs, exposures, impairments, , expenses, recoveries, amounts, writedowns, liability
novartis	pfizer, gsk, merck, wyeth, astrazeneca, janssen, roche, sanofiaventis, schering, glaxosmithkline
chrysler	daimlerchrysler, gm, ford, nissan, daimler, volkswagen, bmw, toyota, chevrolet, honda
facebook	twitter, google, yahoo, youtube, apps, app, websites, web, aol, ebay

Table 2. 15 most similar words to industry prototypes for the 10 most represented industries in 2017

SIC	Industry Name	10 Words Most Similar to The Industry Prototype in 2017
73	Business Services	clients, cloudbased, could, ecommerce, internetbased, saas, customers, may, effectively, ebusiness, business, websites, advertisers, businesses, internetenabled
28	Chemical and Allied Products	commercialization, candidates, collaborators, drug, trials, clinical, drugs, candidate, indications, fda, preclinical, biologic, formulation, commercialize, formulations
36	Electronic and other Electrical Equipment and Components, except Computer Equipment	products, oems, customers, technologies, could, suppliers, oem, chipsets, vendors, hardware, semiconductors, ics, brocade, ic, technological
38	Measuring, Analyzing, and Controlling Instruments; Photographic, Medical and Optical Goods; Watches and Clocks	products, technologies, product, processes, manufacturing, could, technology, manufacturers, diagnostic, may, manufacture, collaborators, costly, delays, technological
35	Industrial and Commercial Machinery and Computer Equipment	products, could, customers, suppliers, oems, brocade, vendors, infrastructure, business, disruptions, oem, difficulties, hardware, businesses, technologies
49	Electric, Gas and Sanitary Services	electricity, pnm, gas, sce, electric, coal, utilities, wpsc, tep, pacificorp, energy, pse, sppc, cleco, psco
13	Oil and Gas Extraction	oil, drilling, exploration, gas, wells, coal, natural, shale, mining, hydrocarbons, basin, future, eog, midstream, hydrocarbon
37	Transportation Equipment	suppliers, aftermarket, oems, could, oem, operations, offhighway, customers, automotive, disruptions, heavyduty, parts, mro, aerospace, business
20	Manufacturing	snack, coffee, beverage, fruit, foodservice, brands, bakery, pasta, beverages, dairy, mattel, meat, beer, seasonings, products
48	Communications	networks, wireless, broadband, television, broadcast, broadcasters, programming, directv, subscribers, terrestrial, internet, satellite, cable, isp, streaming

Table 3. Five firms most similar and least similar to the industry prototype for the business services industry (SIC code 73) in 2017

Five Most Typical Firms (from most to least typical)
MINDBODY Inc. We are the leading provider of cloud-based business management software for the wellness services industry and a rapidly growing marketplace for wellness services. As of December 31, 2017, our customers employed over 372,000 wellness practitioners serving approximately 41 million consumers in more than 100 countries. Our integrated software and payments platform helps wellness business owners run, market and build their businesses, while engaging consumers by aggregating available classes and appointments, and enabling rapid discovery, booking and payment.
Box, Inc. provides a cloud content management platform that enables organizations of various sizes to manage and share their content from anywhere or any device. The company's Software-as-a-Service platform enables users to collaborate on content internally and with external parties, automate content-driven business processes, develop custom applications, and implement data protection, security, and compliance features.
Salesforce is a leading provider of enterprise software, delivered through the cloud, with a focus on customer relationship management, or CRM. We introduced our first CRM solution in 2000, and we have since expanded our service offerings into new areas and industries with new editions, features and platform capabilities.
Bazaarvoice was founded on the premise that the collective voice of the consumer is the most powerful marketing tool in the world. Our solutions and services allow our retailer and brand clients to understand that consumer voice and the role it plays in influencing purchasing decisions, both online and offline. Our solutions capture, manage and display consumer-generated content including ratings and reviews, questions and answers, customer stories, and social posts, photos, and videos.
Xactly Corp. We are a leading provider of enterprise-class, cloud-based incentive compensation solutions for employee and sales performance management. We address a critical business need: to incentivize employees and align their behaviors with company goals.
Five Most Atypical Firms (from most to least typical)
Boston Omaha Corporation commenced its current business operations in June 2015 and currently operates two separate lines of business: outdoor billboards, and surety insurance and related insurance brokerage activities. We also hold minority interests in homebuilding and commercial real estate brokerage activities.
Takung Art Co. Through Hong Kong Takung, Shanghai Takung and Tianjin Takung, we offer on-line listing and trading services that allow artists/art dealers/owners to access a much bigger art trading market where they can engage with a wide range of investors that they might not encounter without our platform.
Moxian Inc. We are in the O2O ("Online-to-Offline") business. While there are many definitions of O2O, with respect to our business, O2O means providing an online platform for small and medium sized enterprises ("SMEs") with physical stores to conduct business online, interact with existing customers.
AeroCentury Corp. Since its formation, the Company has been engaged in the business of investing in used regional aircraft equipment leased to foreign and domestic regional air carriers. The Company's principal business objective is to increase stockholder value by acquiring aircraft assets and managing those assets in order to provide a return on investment through lease revenue and, eventually, sale proceeds.
General Finance Corp. Founded in 2005, we are a leading specialty rental services company offering portable storage, modular space and liquid containment solutions, with a diverse and expanding lease fleet of 80,712 units as of June 30, 2017.

Table 4. Descriptive statistics and correlations

	Std Dev in Monthly Returns	Std Dev in Daily Returns	Typicality	Industry Ambiguity	Log of Mkt Cap	Book-to-Market Ratio		Mean	Std Dev	Min	Med	Max
Std Dev in Monthly Returns	1,00							13,10	8,67	2,54	10,89	62,00
Std Dev in Daily Returns	0,84	1,00						3,29	1,89	0,81	2,82	12,71
Typicality	0,03	0,02	1,00					0,00	1,00	-6,92	0,29	1,39
Industry Ambiguity	-0,04	-0,06	-0,24	1,00				0,00	1,00	-3,13	-0,04	9,15
Log of Mkt Cap	-0,40	-0,51	0,00	0,00	1,00			6,24	2,03	-4,45	6,20	13,89
Book-to-Market Ratio	0,12	0,13	0,01	-0,02	-0,34	1,00		0,53	0,62	-2,52	0,42	4,17
ROA	-0,38	-0,42	-0,03	0,01	0,28	-0,04		-0,04	0,30	-2,37	0,03	0,37

Table 5. OLS regressions of monthly volatility (models 1 to 5) and daily volatility (models 6 to 10) on typicality

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6	(7) Model 7	(8) Model 8	(9) Model 9	(10) Model 10
Typicality		-0.697***		-0.653***	-0.676***		-0.113***		-0.108***	-0.113***
		(0.096)		(0.098)	(0.099)		(0.020)		(0.020)	(0.020)
Industry Ambiguity			0.523***	0.458***	0.507***			0.079***	0.068***	0.079***
			(0.078)	(0.077)	(0.079)			(0.016)	(0.016)	(0.016)
Typ#IndAmbig					0.154**					0.030*
					(0.051)					(0.012)
Log of Mkt Cap	-1.494***	-1.479***	-1.509***	-1.495***	-1.494***	-0.312***	-0.311***	-0.316***	-0.314***	-0.314***
	(0.104)	(0.103)	(0.104)	(0.104)	(0.104)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Book-to-Market ratio	0.521**	0.541**	0.529**	0.543**	0.544**	0.145***	0.145***	0.142***	0.144***	0.144***
	(0.169)	(0.167)	(0.168)	(0.168)	(0.168)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
ROA	-5.737***	-5.686***	-5.683***	-5.649***	-5.653***	-1.288***	-1.283***	-1.280***	-1.275***	-1.276***
	(0.358)	(0.356)	(0.358)	(0.357)	(0.357)	(0.067)	(0.067)	(0.067)	(0.067)	(0.067)
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Effect	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	40,530	40,515	40,137	40,137	40,137	37,115	37,102	36,761	36,761	36,761
Adjusted R-squared	0.510	0.511	0.511	0.512	0.512	0.656	0.656	0.656	0.657	0.657

Robust standard errors in parentheses; *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.1

CHAPTER 2
Organizational Appeal and Market Valuation: A Natural Language Processing Study of IPO First-Day Returns

Abstract [195 words]

Redressing the imbalance in extant research that associates organization appeal with typicality (or similarity to existing prototypes), we distinguish another dimension of appeal -attractiveness, or similarity to recent successes- and study their impact on market valuation. In the context of initial public offerings (IPOs), we expect that typicality reduces information asymmetry, limits the underpricing of stocks by underwriters, and thus results in lower first-day returns. By contrast, since attractive stocks are expected to possess distinctive competences and to achieve superior future performance, they tend to have higher first-day returns. Using a sample of 2,038 U.S. IPOs from 1998 to 2015, we operationalize typicality and attractiveness by applying a novel natural language processing approach to 140,000 financial documents published by established and issuing firms. Whereas typicality does not have a significant direct effect on first-day returns, it has a negative association with first-day returns when investor sentiment is high. Attractiveness has a positive effect on first-day returns; this effect is enhanced when investor sentiment is high. These findings contribute to the literature on organizational appeal and market categories, bring methodological developments in the study of organizational similarity, and bridge financial and socio-cognitive approaches to firm valuation.

Introduction

A well-established strand of research holds that audiences value organizations relative to their conformity to established categories. According to the prevailing wisdom, built on insights from both cognitive psychology and institutional theory, when organizations' features align with the typical features of their category, audiences perceive those organizations as being more predictable, more acceptable, and, ultimately, more appealing (Deephouse, 1996; DiMaggio & Powell, 1983; Meyer & Rowan, 1977; Suchman, 1995). In contrast, when organizations mix features from different categories, they are considered to be less "pure" members of each category and hence are, overall, less appealing (Hsu, Hannan, & Pólos, 2011). Thus, typical organizations are more appealing and derive a higher market valuation than their less typical peers.

Although this framework fits many settings, it suffers from two limitations. First, it focuses on audiences relying primarily on one mode of evaluation: comparing the features of organizations with those of relatively stable prototypes of existing categories (Hannan et al., 2019). Yet, audiences sometimes experience fleeting attractions toward features that may or may not be typical of established categories (Abrahamson and Fairchild, 1999; Durand and Khaire, 2017) -this second facet of appeal has been largely ignored, leading to a dramatic imbalance in how organizational appeal has been defined and measured, by essentially linking appeal to membership in a category (Hsu, Koçak, & Hannan, 2009). Second, prior studies have analyzed the consequences of appeal-as-typicality primarily on evaluation (e.g., assessments by customers and critics), and less so on market valuation per se (i.e. pricing of an entity in dollars). Thus, a more comprehensive theory of organizational appeal and its relationships with market valuation is needed.

We address these limitations by developing a model of organizational appeal that comprises two dimensions: *typicality*, which we define as similarity to a category prototype; and *attractiveness*, which we define as similarity to organizations that have recently been successful. We expect that when an organization is typical of its category there is less opportunity for information asymmetry and market valuation is easier. Attractive organizations resemble recent successes and thus appear as possessing features which will lead to higher future performance (Zhao et al., 2018). Hence, market valuation tends to be higher for attractive organizations because audiences perceive them as possessing special competencies that contribute to their competitiveness and future performance.

We develop and test our theory using data on 2,038 initial public offerings (IPOs) in the United States from 1998 to 2015, focusing on how typicality and attractiveness impact first-day returns. The IPO setting is appealing to test our proposed theory because investors are likely to (1) value issuing firms by comparing them to the prototypes of their industry and (2) be influenced in their judgments by their knowledge of recent IPOs that have experienced a prominent surge in their valuation on their first day of trading. As typical firms are less exposed to information asymmetry, underwriters have a reduced propensity to underprice them, resulting in lower first-day returns. By contrast, since attractive stocks are expected to achieve superior future performance, they tend to have higher first-day returns. To ascertain the strength of both these effects, we introduce investor sentiment as a touchstone factor. When market sentiment is high, excitement around stocks with high information asymmetry is exacerbated, leading typical stocks to experience even lower first-day returns. During these periods, investors are also more inclined toward stocks expected to deliver superior future performance, and thus attractive stocks have even higher first-day returns.

To test these hypotheses, we measure typicality and attractiveness by representing the issuing firms and a benchmark of 17,542 established firms drawn from Compustat in a shared high-dimensional space. We gathered all annual reports and IPO prospectuses published by both the issuing firms in our sample and established firms in our benchmark—in total 38,256 IPO-related documents and 100,263 annual reports. We used a natural language processing technique called Doc2Vec (Lau & Baldwin, 2016; Le & Mikolov, 2014; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) to represent the collected publications as vectors in a shared space. We then created vectors for industry prototypes and operationalized typicality as the similarity of an issuing firm's vector to the prototype of its main industry category. We operationalized attractiveness as the average similarity of an issuing firm to the five most successful IPOs in the preceding year.

Our findings do not support the expectation of a negative association between typicality and first-day returns but significantly support the expectation of a positive association between attractiveness and first-day returns. Notably, this effect is comparable to those of much more established variables such as venture capital support or the hotness of the IPO market (i.e. a one standard deviation increase in attractiveness leads to an increase in first-day returns of 4.64% to be compared with a gain of 6.82% in first-day returns for a one standard deviation increase in the hotness of the IPO market and a gain of 4.85% in first-day returns for being backed by venture capitalists). We find support for the hypotheses that when investor sentiment is high, typical firms tend to experience lower first-day returns (marginal support) and attractive firms higher first-day returns (strong support). In a series of robustness checks and supplemental analyses, we show that the results are robust to measurements of attractiveness based on past within-industry successes rather than cross-industry successes, to the use of a measure of typicality that takes into account the leniency of the category (Pontikes & Barnett, 2015), and to the use of different levels of definitions for industries when measuring typicality.

We make three main contributions with this paper. First, we theorize and uncover the concurrent effects of the two dimensions of appeal, typicality and attractiveness, on IPOs' first-day returns and market valuation. Second, we propose a new method to measure typicality and attractiveness which is both theoretically and practically an improvement on pre-existing measurements, as it enables researchers to jointly examine the effects of typicality and attractiveness based on a single collection of documents. Third, we bridge financial and socio-cognitive approaches to IPOs' first-day returns by promoting typicality and attractiveness as impacting first-day returns through their connection to mechanisms that are both existing (information asymmetry) and often ignored (alleged presence of special competencies leading to expectations of higher future performance).

Organizational appeal as typicality: current developments and limitations

One of the main tools that audiences use to make sense of and value organizations are established systems of categories (for reviews see Cattani et al. 2017, Durand and Paolella 2013, Vergne and Wry 2014). Categories, understood as groupings of organizations based on their resemblance to a prototype –an abstract representation of the 'average member' of a category–, shape how audiences construct and interpret information on organizations in markets (Durand & Paolella, 2013; Hsu et al., 2011). The typicality of an organization depends on its similarity to its category's prototype and signals the organization as belonging to a clear grouping of organizations. Typicality is generally negatively linked to meddling with categorical boundaries: atypical entities are thus defined as both those deviating from typified representations of their peers (Smith, 2011) and those combining features from categories that are distant from one another in a conceptual space (Goldberg, Hannan, & Kovacs, 2016; Kovács & Hannan, 2015). In other research, atypicality is simply assimilated to category-spanning, as mixing features from multiple categories blurs an entity's type (Hsu et al., 2011, 2009).

A first dimension of organizational appeal follows from this perspective. As Hsu et al. (2009, p. 153) write, "Category membership can be linked to the intrinsic appeal of a producer/product to an audience member—that is, the degree to which it fits her taste (Hannan et al. 2007)." Because agents tend to value more favourably those offerings that meet their expectations for a category, appeal is positively associated with typicality. Driving this association between appeal and typicality is the certainty that the features of typical organizations and the performance they will deliver are both within the expectations set by established prototypes. Little ambiguity surrounds the activities of typical organizations, and information about them can be readily and easily interpreted (Zuckerman, 2004). Audiences feel assured that typical organizations will deliver, at least, a minimally satisfying level of performance (Zuckerman, 2017). By contrast, atypical entities are unsettling for audiences, who have a harder time identifying, interpreting, and valuing them (Hsu, 2006).

Two factors lead us to consider this prevailing approach as incomplete. First, in some contexts, audiences ascribe high value to atypical organizations. For example, venture capitalists who seek to invest in "the next big thing" prefer atypical organizations, which have the potential to disrupt established categories (Pontikes, 2012); law firms that straddle categorical boundaries are more appealing to clients facing high stakes and a complex environment (Paolella & Durand, 2016); and atypical hedge funds are appealing to investors, provided they signal their ability to deliver superior performance (Smith, 2011). Hence, while typicality is fundamental, the extant research seems to have an over-reliance on typicality as the main or only component of appeal. Thus, more recent research has underlined that when evaluating organizations, audiences may rely on alternative models of valuation, based, for example, on the construction of *ad hoc* categories or the alignment of organizations with a *theory of value* that defines which kinds of organizations may be more likely to help achieve a prespecified set of goals (Durand & Paolella, 2013; Paolella & Durand, 2016; Zuckerman, 2017).

Second, beyond the controllable aspects of their resources and their market positioning within a given category, organizations may exhibit features that are not necessarily characteristic of their category but that seem attractive to audiences as they perceive them as signalling superior competencies due to current hypes and trends (Abrahamson & Fairchild, 1999; Granqvist, Grodal, & Woolley, 2013; Lee, 2001; Zhao et al., 2018). Notably, similarity with recent successes signals such an alignment. Hence, while a first dimension of appeal results from an audience's judgments on whether an organization's traits relate to a category (i.e., show typicality), a second dimension of appeal results from an organization's alignment with recent successes. This second dimension of appeal, which we label *attractiveness*, has been marginalized in prior research. Although scholars have extensively studied the consequences of appeal-as-typicality on valuation and performance, no balanced, systematic treatment of a producer's two-sided appeal has yet emerged. We tackle this challenge by specifying the role of attractiveness in valuation and its divergence from the role of typicality.

Specifying the role of attractiveness as a second dimension of appeal

Attractiveness characterizes impermanent features associated with success or hype that lead to an entity being perceived as possessing special competencies suggestive of superior future performance (Abrahamson & Fairchild, 1999; Zuckerman, 2017). These features may correspond to generic characterizations that cut across industries or salient labels related to firms' specific resources and products (e.g., organic, AI, carbon capture, dot-com) (Granqvist et al., 2013). For instance, PitchBook, the world data provider on venture capital and IPOs, created "slices"—i.e., generic classes of firms that encompass firms from multiple industries, such as edtech or agritech. An appeal-as-typicality approach would have difficulty explaining why firms belonging to these "slices" could be appealing: they are not representative of their industry prototype and they possess features that cut across industry boundaries. However, an approach based on attractiveness explains their appeal: due to recent successes within these "slices", investors currently believe that firms belonging to these slices possess special competencies that will enable them to achieve superior performance. Whereas typical features are consubstantial to a category, attractive features are fluid and change over time. Their effects are therefore less controllable and less predictable. For example, during the dot-com bubble, firms with dot-com in their names generally enjoyed superior returns, due to widespread excitement for activities tied to the Internet (Lee, 2001); however, this effect disappeared when the bubble burst.

Audiences interpret a firm's similarity to recent successes—i.e., its attractiveness—as a sign that it possesses competencies that will enable it to emulate these successes. Evidence of this behavior among audiences recently surfaced in Zhao et al.'s (2018) study of category emergence in the video game industry. The authors showed that categories emerge around exemplary hit games in the video games industry and that in the early stage of category emergence, strong similarity to the exemplary hit games yielded higher appeal from audiences. Zhao et al. (2018) further argued that exemplary hit games play an important role by serving as a highly salient benchmark to identify new games as having a high potential. As the effects of attractiveness are dependent on which firms are identified as successes at a given point in time (i.e., on the influence of temporary hypes and fads), audiences find certain kinds of organizations to be attractive in a non-durable way (Abrahamson and Fairchild, 1999).

Organizational appeal is thus the function of two components: typicality, the degree to which an organization conforms to existing prototypes, and attractiveness, the degree to which an organization is similar to recent successes. Typicality generates appeal through the feeling of certainty it offers to audiences. Attractiveness generates appeal by suggesting higher future performance. Due to the different mechanisms through which they generate appeal, typicality and attractiveness do not necessarily have identical effects on audiences' valuations. We develop the conflicting impacts on valuation of these two components of appeal in the context of IPOs.

Linking typicality and attractiveness to insights on IPOs from the finance literature

During an IPO, an issuing firm enters the public market by selling a portion of its shares previously owned by private investors to institutional investors who then start trading the shares to other investors on the public market. A small set of investment banks, or underwriters, which form the underwriting syndicate, are usually in charge of pricing the shares, based on their analysis of the firm and their assessment of institutional investors' interest in the shares. Prior to introducing the offering to institutional investors, underwriters determine a price range for the issuing firm's shares. Underwriters and the issuing firms then present the offering to institutional investors through a roadshow and private meetings, and to all investors through the initial pre-IPO prospectus, also known as Form S-1, a mandatory document required by the U.S. Securities and Exchange Commission (SEC) and available on its website. Form S-1 is arguably among retail investors' primary source of information on the issuing firm, as suggested by the number of non-robot requests for this document from the SEC website (Loughran & McDonald, 2017); however, retail investors may also gather information on the IPO from the financial press, or from recorded interventions by underwriters or the issuing firm's managers.

Once the underwriters have settled on an offer price, those institutional investors who have expressed their interest in buying the issuing firms' shares are allocated shares, which they purchase at the offer price. On the first day of trading, the shares begin trading at the market price, which usually rise well above their offer price, generating high first-day returns. The finance literature has generally interpreted this phenomenon as a sign that IPOs are "underpriced," -i.e., that their offer price has been set below their expected market price- and has suggested one of the main causes of underpricing is information asymmetry.

The information asymmetry argument relies on the observation that uninformed investors face a winner's curse: they risk being allocated shares that are priced too high because informed investors did not invest in them (Rock, 1986). IPOs are thus routinely underpriced to entice investments from uninformed investors or investors exposed to bad signals (Biais & Faugeron-Crouzet, 2002). Routinely underpricing IPOs also ensures that institutional investors have no incentives to invest in acquiring information to detect overpriced shares (Gondat-Larralde & James, 2008). Similarly, underpricing by only partially adjusting the offer price of high-demand issues can compensate investors for revealing positive information about the issuing firm's value (Hanley, 1993). Managers of issuing firms seem to endorse the interpretation that underpricing compensates investors for investing in shares with uncertain value. For example, a survey of 336 chief financial officers (CFOs) in the United States found that 60% of CFOs agreed with the statement that underpricing compensates institutional investors for investing in IPOs which informed investors might have dodged (Brau & Fawcett, 2006).

The first dimension of organizational appeal, typicality, is highly compatible with the information asymmetry argument. Notably, one of the main attributes of typical organizations is that they are easier to understand by using immediately available knowledge (Aldrich & Fiol, 1994), and can be readily evaluated based on the models of valuation associated with their category (Zuckerman, 2004). As typical IPOs lie close to their industry prototypes and exhibit the typical traits of their industry at a given point in time, they are easier to value by identifying comparable peers and relying on industry-specific assumptions -such as industry multiples, one of the most common techniques for the valuation of IPOs (Kim & Ritter, 1999; Paleari, Signori, & Vismara, 2014; Purnanandam & Swaminathan, 2004; Roosenboom, 2012). Thus, information asymmetry among investors is lower for typical firms, leading to less underpricing and lower first-day returns. Hence:

> **Hypothesis 1a:** *The higher the typicality of an issuing firm, the lower its first-day returns.*

Beyond arguments relying on information asymmetry and underpricing to explain first-day returns[3], some scholars have proposed an alternative explanation of first-day returns that has nearly disappeared from the finance literature. This research suggests that in the presence of a

[3] Note that while we focus on the information asymmetry literature, another major line of arguments in finance ties the effects of underpricing to agency problems. Under this view, underpricing is due to agency conflicts involving underwriters, investors, and issuers. In the 1990s, issuers tended to overlook underpricing in exchange for side-payments or for coverage by prestigious analysts (Loughran and Ritter, 2004). Some investors engage in long-term contracts with brokers and pay fixed per-share commissions in exchange for access to brokers' premium services, such as access to underpriced IPO shares (Goldstein, Irvine, Kandel, & Wiener, 2009). Others send abnormal commissions to brokers in charge of profitable IPOs in their efforts to be allocated shares (Goldstein, Irvine, & Puckett, 2011). Accounts based on prospect theory propose that when both private and public information suggest the market price of an IPO will be higher than the expectations set by the initial price range, underwriters will leverage issuers' reduced sensibility to the amount of money "left on the table," due to a concurrent increase in the wealth resulting from the IPO, to only partially adjust the price of IPO shares (Loughran & Ritter, 2002). An argument based on this view would also find that typical issuing firms enjoy reduced first-day returns as, the value of the stock being more agreed upon among all parties, there is less room for manipulation of the offer price. Here, we focus on the information asymmetry argument, as it directly aligns with typicality as leading to more certainty regarding the value of the firm. We comment on the agency approach in the Discussion section as a complementary line of research.

community of investors who hold highly positive opinions about an issuing firm's future performance, this firm's first-day returns might be very high even without underpricing (Derrien, 2005; Houge, Loughran, Suchanek, & Yan, 2001; Miller, 1977). The financial press often conveys similar views on first-day returns, interpreting high levels of first-day returns as a sign that investors have high expectations for the future performance of the firm and low or negative first-day returns as a sign that investors have poor expectations. Under the information asymmetry argument, underwriters underprice shares of firms submitted to information asymmetry to overcome investors' reluctance to bid for them. While this first view presents high first-day returns as resulting from a lack of appeal due to the presence of information asymmetry among investors, the second view introduced above presents first-day returns as resulting from superior appeal due to investors expecting high future performance. Our proposition to separate typicality and attractiveness as two distinct dimensions of appeal allows to resolve the apparent contradiction between these two views. Whereas appeal-as-typicality aligns its effects with information asymmetry and reduces first-day returns, appeal-as-attractiveness relies on a different mechanism (i.e., revealing a presumed superior future performance), and leads, *ceteris paribus*, to higher first-day returns. The two dimensions of appeal thus have opposite consequences on first-day returns.

According to research on categorization, audiences value entities based on their similarity to either established prototypes or salient features associated with success (Massini, Lewin, & Greve, 2005; Zhao et al., 2018). In the IPO market, each IPO is unique and shares similarity with its industry peers (typicality) but pertains also to the broader set of all prior IPOs. As such, each IPO is more or less similar to other recent IPO successes, both within and outside of their industry category. We define 'successful' IPOs (in the eyes of investors) as those experiencing high levels of first-day returns. Hence, investors will interpret an issuing firm's similarity to recently successful peers – its attractiveness – as a sign that it also possesses features indicative of superior future performance. Therefore, attractiveness leads to more optimistic expectations among investors regarding an issuing firm's future performance, pushing first-day returns upward. Hence:

> **Hypothesis 1b:** *The higher the attractiveness of an issuing firm, the higher its first-day returns.*

Asserting the role of appeal through investor sentiment

Thus far, we have (1) tied the effects on first-day returns of issuing firms' typicality to reduced information asymmetry and (2) tied the effects on first-day returns of issuing firms' attractiveness to an expectation of superior performance. To further assert the presence of these mechanisms, we seek a touchstone factor, independent of the firm, to provide evidence of the influence of either mechanism. How investors interpret firms' attributes at a given point in time is tied to their collective beliefs at this moment—i.e., what finance scholars have called their "sentiment" (Ibbotson, Sindelar, & Ritter, 1994), which can either inhibit or reinforce the influence of typicality and attractiveness on first-day returns.

Financial experts' anecdotal evidence supports the claim that investor sentiment affects judgments in terms of what is considered to be an appealing IPO and can indirectly influence first-day returns. To illustrate, we report advice by Howard Marks and Warren Buffet, two leading figures in the investment community, as written in letters to shareholders in 2000, at the burst of the dot-com bubble (1999–2001). In Marks's February 2000 letter to clients of his fund, Oaktree, the investment guru described the "lottery ticket mentality" of investors during the bubble. He emphasized that the usual indicators of business performance were being disregarded and that investors instead were betting on attractive concepts, which they were not necessarily able to assess or understand, out of fear of missing “the next big thing”:

> In this valuation parameter vacuum, a “lottery ticket mentality” seems to govern the purchase decision. The model for investments in the tech and dot-com companies isn’t the likelihood of a 20% or 30% annual return based on projected earnings and p/e [price–earnings] ratios, but a shot at a 1,000% gain based on a concept. The pitch might be “We’re looking for first-round financing for a company valued at $30 million that we think we can IPO in two years at $2 billion.” Or maybe it’s “The IPO will be priced at $20. It may end the day at $100 and be at $200 in six months.” Would you play? Could you stand the risk of saying no and being wrong? The pressure to buy can be immense. (Marks 2000, pp. 10–11).

Similarly, in his 2000 letter to the shareholders of Berkshire Hathaway Inc., CEO and Chairman Warren Buffet dismissed investors’ relentless investing in companies with highly uncertain prospects. He went on to underline that as stock prices continued rising due to the bubble, investors’ enthusiasm for uncertain stocks also grew and, although they all knew the party would eventually stop, they felt compelled to continue investing in so-called attractive stocks. He then added that, unlike other investors, he aimed to continue investing in business he understood and whose future performance he was able to gauge with a reasonable margin of error:

> At Berkshire, we make no attempt to pick the few winners that will emerge from an ocean of unproven enterprises. We’re not smart enough to do that, and we know it. Instead, we try to apply Aesop’s 2,600-year-old equation to opportunities in which we have reasonable confidence as to how many birds are in the bush and when they will emerge (a formulation that my grandsons would probably update to “A girl in a convertible is worth five in the phonebook.”). Obviously, we can never precisely predict the timing of cash flows in and out of a business or their exact amount. We try, therefore, to keep our estimates conservative and to focus on industries where business surprises are unlikely to wreak havoc on owners. Even so, we make many mistakes: I’m the fellow, remember, who thought he understood the future economics of trading stamps, textiles, shoes and second-tier department stores. (Buffet 2000, p. 14).

These qualitative insights that two leading figures of the investment community published at the heat of the Internet bubble suggest that during the dot-com bubble investors found appeal in firms that may not have been typical of their industry but shared features that investors found attractive.

These accounts of investors' behavior during the dot-com bubble can be generalized to describe how investors react during periods of high sentiment. During these periods, investors face a strong social pressure to evade typical investments and to instead prefer attractive ones, either because they let their sentiment drive their decisions or because they imitate other investors, in fear of being left out. In other words, investors are not only less sensible to typicality but also more sensible to attractiveness (Baker & Wurgler, 2006, 2007). The greater excitement around stocks with high information asymmetry leads to a disregard for more typical firms. Thus, when investor sentiment is high, typical stocks experience even lower first-day returns. Furthermore, investors become overly optimistic so that those firms that are perceived as attractive are also perceived as being able to provide returns even superior than what would have been expected in calmer periods. Thus, when investor sentiment is high, shares in attractive firms are even more appealing, pushing further higher first-day returns of issuing firms. Hence the following hypotheses:

> **Hypothesis 2a:** *The level of investors' sentiment reinforces the negative relationship between firms' typicality and their first-day returns.*
> **Hypothesis 2b:** *The level of investors' sentiment reinforces the positive relationship between firms' attractiveness and their first-day returns.*

Data and methods

We examine typicality and attractiveness together by using a method that produces commensurate metrics and relies on natural language processing. A key step in our measurement of typicality and attractiveness is to represent, in a shared semantic space, both the issuing firms in our sample and established firms that were publicly listed at a given point in time, by using a model called Doc2Vec (Dai, Olah, & Le, 2015; Lau & Baldwin, 2016; Le & Mikolov, 2014). In the following section, we first describe the data used in this paper and then discuss each step of our method to represent firms in a semantic space using Doc2Vec.

Data. We collected data on IPOs that occurred in the United States between 1996 and 2015, which we obtained from SDC Platinum new issues database. We then cleaned this initial data using Pr. Jay Ritter's database of IPOs since 1975, available on Pr. Jay Ritter's website and as described by Loughran and Ritter (2004: Appendix B). We supplemented this cleaned data with both stock-level data for the first day of trading from the Center for Research in Security Prices (CRSP) and fundamentals data for the fiscal year prior to the year of the IPO from Compustat. We excluded from the sample those IPOs with an offer price below $5 (i.e., penny stocks), and those that represented financial institutions, closed-end funds, American depository receipts, and real estate investment trusts. We also required that all IPOs within the sample have a valid Central Index Key, which the SEC provides for identifying and locating financial documents in its database. Using this Central Index Key, we then collected from the SEC website the initial and final prospectus for each IPO and any amendments to those prospectuses. All in all, we collected 38,256 IPO-related documents.

We needed a representative sample of established, publicly traded firms to measure whether IPOs were typical of their industries. To this end, we leveraged the fact that firms listed in Compustat are associated with both a standard industrial classification (SIC) code, which identifies a firm's industry, and a Central Index Key, which enabled us to retrieve annual reports from the SEC website. To create a benchmark of established publicly traded firms against which to compare IPOs, we thus downloaded from the SEC website the annual reports of each firm present in the Compustat database between 1995 and 2015 and for which we had a valid Central Index Key. We gathered a total of 100,263 annual reports from 17,542 firms.

Preprocessing of documents. As is common in related research using 10-Ks and IPO prospectuses (e.g., Hanley & Hoberg, 2010; Hoberg & Phillips, 2016), we used several steps to preprocess each document we had collected. First, we extracted the main part of the document (i.e., the annual report or the IPO prospectus) and omitted the tables and appendixes. From the main text, we removed all inserted pdfs, images, other inserted files, and html code. We then tokenized the documents and removed punctuation, digits, and stopwords -i.e., frequently occurring words such as *the* and *a*. Finally, we lowered all tokens. At the end of this preprocessing, each document in our collection consisted of an ordered list of tokens, without digits, punctuation, or stopwords.

Presenting Doc2Vec. Doc2Vec is a model that researchers can use to represent documents in a collection as vectors in a shared space. Doc2Vec was first introduced by Le and Mikolov (2014), who provided a high-level discussion of how it works in addition to graphical illustrations detailing the functioning of the model. We now explain the general intuition behind the model, largely based on Le and Mikolov (2014), to which we refer the interested reader.

Doc2Vec comes in two flavors: the "distributed memory" version (henceforth DM) and the "distributed bag-of-words" version (henceforth DBOW). The DM version of the model is based on a continuous bag-of-words model, which learns the representations of words in a vector space by trying to predict a focal word on the basis of a few surrounding context words (Mikolov et al., 2013). A continuous bag-of-word model consists of a sliding window over a collection of documents, and at each time step, the model's goal is to predict a target word within the window, based on neighbouring words, assuming that the order of those words does not matter. Thus, while continuous bag-of-words models are not sensitive to word order *per se* -in the tradition of most similar models, such as topic models- they are sensitive to the *proximity* between words.

The DM version of Doc2Vec is a simple extension of a continuous bag-of-word model wherein a document-specific vector is learned by sliding a window over a collection of documents and trying to predict a target word based on both words within the window *and* the identity of the document. The document vector that is learned through this process summarizes the document-specific information shared across contexts; thus, it acts as a “distributed memory” (Le & Mikolov, 2014). In the DBOW version of Doc2Vec, computation is simplified by predicting words randomly drawn from randomly picked windows over the document, solely relying on the document identity (Le & Mikolov, 2014). In both versions, the document vector that is learned can be thought of as a representation of the semantic content of the entire document.

There is unclear evidence regarding which version of Doc2Vec performs better. Le and Mikolov suggest that the DM version is more appealing, as it is closer to the word order and outperforms the DBOW version. Yet, they recommend concatenating vectors learned from both versions of the model for better downstream performance in terms of sentiment analysis or information retrieval (Le & Mikolov, 2014). In a more recent examination of Doc2Vec, the DBOW version was found to be less computationally intensive and to perform better than the DM version in certain classification tasks (Dai et al., 2015; Lau & Baldwin, 2016). The results we present rely on the DBOW version.

Learning word vectors using Doc2Vec. When training Doc2Vec on a collection of documents, researchers must set a number of parameters. For example, researchers can fix the dimensions of the vectors that will be learned, how frequently a word needs to show up in the complete corpora to be included in the computation, the size of the sliding window over documents, the number of epochs (i.e., the number of times the model processes the complete collection of documents during training), and the learning rate. The results we present in this paper emerged from the following specifications: we fixed the number of dimensions at 300 and the minimum count of words at 10; and adopted a window size of 10, with the number of epochs at 20, and an initial learning rate of 0.025, which was fixed to decrease linearly at each epoch to reach a minimum of 0.0001 at the last epoch. By omitting words that occurred fewer than 10 times in our complete corpora, we ended up with a total vocabulary size of 378,736 word-types.

We tried alternative specifications -e.g., fixing the number of dimensions at 100 and 200, and fixing the number of epochs at 10- and using these specifications, we ran the same analysis that led to the results described in this paper. Our main findings were not substantially altered by the use of different specifications (see Appendix A). The implementation of Doc2Vec that we used came from the Python library Gensim. After training Doc2Vec on our collection of documents, we obtained a vector representing each document in our corpus. We then used these vectors to represent IPOs and established firms in a shared vector space.

*Computing prototype and IPO vectors, and assessing their face validity***.** We represented each issuing firm at the moment of its IPO as the average of the vectors representing the different versions of Form S-1 associated with it. By assimilating established firms in a given year to the vector associated to their annual report, we were able to represent the prototype of a given industry in a given year, which we needed to compute typicality. For each year and industry, we considered the set of all established firms recorded as belonging to the industry in the past three years and represented each of them using the vector associated to their most recent annual reports.[4] We then computed the prototype of the focal industry for the focal year as the centroid of these vectors:

[4] This procedure ensured that firms that are recorded as belonging to the industry for two or three years of the last three years are not counted several times.

$$Issuing\ firm\ i = \frac{1}{\text{Number of docs for IPO i}} \sum_{doc\ for\ IPO\ i} \begin{bmatrix} \vdots \end{bmatrix}^{doc}$$

$$Prototype\ industry\ p\ in\ year\ y = \frac{1}{|Industry\ p^y|} \sum_{f \in Industry\ p^y} \begin{bmatrix} \vdots \end{bmatrix}^{\text{Most recent } 10-K\ for\ f\ in\ the\ past\ three\ years}$$

Where i is an issuing firm, f is an established firm, $Industry\ p^y$ is industry p in year y—conceived as the set of firms recorded as belonging in the industry in at least one of the past three years—and $|Industry\ p^y|$ the cardinal of industry p. We took an established firm's industry to be the industry identified in its SIC code (measured at the level of the first two digits of the SIC code).

To establish the face-validity of our vectors, we ran a series of tests. First, at the level of the firms, we controlled that vectors capture meaningful similarities between firms. Table 1 presents the five established companies most similar to three selected IPOs: one of the "Big Four" tech companies, Facebook; one company in the apparel industry, New York & Company, Inc.; and one pharmaceuticals company, Conatus Pharmaceuticals Inc. We associated each firm with the first sentences of the business overview section of its annual report. Firms identified as being the most similar to a focal IPO are indeed related to it in terms of their activities.

-- *Insert Table 1 about here* --

Second, to address the concern that our use of two different types of documents might impact our results, we gauged whether the position of an issuing firm in the space, as measured using its IPO prospectus, was a good predictor of its position in the space in the following year, as measured using its annual reports. In 89% of cases across all years, the established firm in year $y + 1$ that is the most similar to a focal issuing firm in year y is the same firm. In 94% of cases, the most or second most similar firm is the same firm. Therefore, our use of annual reports and IPO prospectuses did not interfere with the ability of our approach to capture the position of a given firm (either established or issuing) *vis-à-vis* other firms. This result also suggests that the position of a firm remains relatively stable in the short term.

To provide readers further insights into the relationships between firms captured by our approach, we indicate in Table 2 the four established firms that, for the years 2000, 2007, and 2014, are the most similar to the industry prototypes for the two most represented industries among IPOs in our sample—chemical & allied products, and business services. We associated each firm with the first sentences of the business overview section of its annual report. The firms that are closest to the prototype are representative of their industry.

-- Insert Table 2 about here --

Variables, models and descriptive statistics

Dependent Variable. Our main dependent variable is the level of *first-day returns* for a focal issuing firm, measured as the difference between market price at the end of the day of trading and the offer price, multiplied by 100, and divided by the offer price (e.g., Ibbotson et al., 1994; Loughran & Ritter, 2004; Pollock & Rindova, 2003; Pollock, Rindova, & Maggitti, 2008).

Measuring Typicality. We operationalized the *typicality* of an IPO as its similarity to the prototype of its main industry category, as defined using the two-first digits of the SIC code associated with the issuing firm by Compustat.[5] The formula that we used to measure the similarity of an issuing firm to its industry prototype is as follows:

$$Typicality(Issuing\ firm\ i) = cosine\ similarity(Issuing\ firm\ i, Prototype\ for\ industry\ of\ firm\ i)$$

This formula is in line with the conceptualization of typicality/atypicality as similarity/dissimilarity from a prototype, conceived as a central tendency or "average" over the

[5] Since we created industry prototypes using established firms identified as members of the industry in Compustat, we used the SIC code, as associated with issuing firms by Compustat, not by Thomson. However, note that these SIC codes agreed 83.6% of the time. In unreported analyses, we ran our models using the Thomson SIC codes and when the Thomson SIC code and Compustat SIC code agreed; and the results remained consistent.

features of the members of a given category (Durand & Paolella, 2013; Haans, 2019). To render the interpretation more natural, in our final analysis we multiplied *typicality* by 100. Finally, we centered this variable to simplify the interpretation of the interaction coefficient in our statistical analysis.

Measuring attractiveness. We operationalized *attractiveness* as the average similarity of a focal issuing firm to the top five most successful IPOs in the preceding year, as measured by their first-day returns. We thus measured attractiveness as the extent to which a focal entity was similar to recently successful entities. The formula that we used is as follows:

$$Attractiveness(Issuing\ firm\ i) = \frac{1}{5} \sum_{j \in Top\ 5\ ipos\ in\ year\ y-1} cosine\ similarity(Issuing\ firm\ i, Issuing\ firm\ j)$$

To render the interpretation more natural, in our final analysis we multiplied *attractiveness* by 100. We also centered this variable to simplify the interpretation of the interaction coefficient in our statistical analysis. Table 3 shows the five IPOs that had the highest level of underpricing in year $y - 1$ along with the five most attractive IPOs in year y for the years 2000, 2007, and 2014. As can be seen, attractive firms tend to belong to certain specific industries but do not necessarily all belong to the same industry, which supports the idea that *attractiveness* measures investors' temporary inclination toward certain kinds of firms at a given point in time.[6]

Moderator and controls. We measured *investor sentiment* using survey data from the American Association of Individual Investors (AAII). Each week the AAII asks its members to indicate where they think the stock market will be in six months: up, down, or the same. Based

[6] Note that in robustness checks, we also developed a measure of attractiveness based on recent *within* industry successes and find similar results.

on their responses, investors are then labelled as bullish, bearish, or neutral, resulting in the weekly percentages of investors feeling bullish, bearish, or neutral. We use the difference between the percentage of bullish investors and the percentage of bearish investors as a proxy of investor sentiment (Brown & Cliff, 2004; DeVault, Sias, & Starks, 2019).[7] Similar to the treatment of our two independent variables, we centred this variable.

We used several control variables based on the literature on IPOs. Because our results could be driven by agency conflicts involving underwriters, investors, and issuers, we controlled for revisions in the final offer price relative to the initial middle of filing price range. *Offer price revision* captures underwriters' attempts to manage both the perceived cost of leaving money on the table for the issuer and investors' expectations that the share would be underpriced (Hanley, 1993; Loughran & Ritter, 2002; Ritter & Welch, 2002). It was operationalized as the percentage gain (loss) of the offer price relative to the initial middle of filing price range. We controlled for the "hotness" of the market for IPOs at the moment of the IPO by using the percentage of firms whose offer price was above the midpoint of the price range, as provided in their initial prospectus (Ibbotson et al., 1994). We controlled for the IPO *market hotness* independently of our measurement of *investor sentiment* to try to alleviate concerns that our results could be driven by timing effects, whereby issuers wait for the IPO market to be hot to initiate their IPO. We also controlled for whether the IPO received venture capital support before the IPO, as the presence of venture capitalists may also influence the setting of the offer price (Arthurs, Hoskisson, Busenitz, & Johnson, 2008). We further controlled for each firm's size and age. For the former, we used the log of the firm's total assets in the fiscal year preceding the year of the

[7] We used other specifications for *sentiment* and obtained the same results. For instance, we measured *sentiment* at the moment of the IPO as the percentage of firms whose offer price was above the midpoint of the price range given in their initial prospectus (Ibbotson et al., 1994).

IPO, as reported in Compustat; and for the latter, the log of the number of years since its creation plus one.

We added industry fixed-effects, which were based on the first two digits of an IPO's main SIC code; year effects; and fixed-effects for the lead underwriter and for the stock exchange on which the stock is traded. Industry fixed-effects control for unobserved industry-specific effects that are stable over time. The inclusion of lead underwriter fixed-effects is an additional way of controlling for potential agency conflicts at the level of the underwriter; for underwriter's level characteristics, such as the underwriter's prestige (Carter & Manaster, 1990) or its connections with institutional investors (Goldstein, Irvine, Kandel, & Wiener, 2009; Goldstein, Irvine, & Puckett, 2011); and for each lead underwriter's general propensity to underprice shares. Stock exchange fixed-effects ensured that our results were not driven by systematic differences in IPO returns as a function of the market in which the focal firm was listed. Year effects controlled for unobserved year-specific trends.

Models. In all our models, we use used clustered errors by industry to mitigate concerns that errors might be correlated within industries. To test for Hypotheses 1a, 1b, 2a, and 2b, we tested different versions of the following model:

$$First - day\ returns = \beta_1 Typicality + \beta_2 Attractiveness + \beta_3 Typicality \times Sentiment + \beta_4 Attractiveness \times Sentiment + \beta_5 Sentiment + \beta Controls + \varepsilon$$

Descriptive statistics and correlation table. Our final dataset corresponds to 2,038 IPOs that occurred between 1998 and 2015. Table 4 presents descriptive statistics and the correlation matrix for our variables. The mean for all our main independent variables and our moderator -*investor sentiment*- is 0 since we centered these variables to facilitate later interpretation of the interaction terms. The *first-day returns* variable skews heavily to the right, as is usual in studies on IPOs. In our sample, 56% of firms received venture capitalist support. Our main independent variables were weakly correlated, with a correlation of 0.09.

-- *Insert Table 4 about here* --

Main results and robustness checks

The results of our main analysis are presented in Table 5, referencing Models 1 to 7. Model 1 contains only *investor sentiment* and our control variables. As can be seen, in this model, *investor sentiment* has no direct effect on *first-day returns*. This is not surprising given that we control for the hotness of the IPO market, which is often used as a measure of the level of investor sentiment on the IPO market specifically – as opposed to the general level of investor sentiment. The *offer price revision* has a highly significant positive effect on *first-day returns*. A one standard deviation increase in *price revision* (a gain of 14.4%) is associated with a gain of 14.95% in *first-day returns*. When accounting for *price revision,* most other control variables have expected effects on *first-day returns* but are only weakly significant. Only *market hotness* appears strongly significant, with a gain of one standard deviation in *market hotness* leading to a gain of 6.98% in *first-day returns*.

Models 2 to 7 introduce successively our main variables of interest. As expected, the overall effect of *typicality* on *first-day returns* is constantly negative across models but never reaches significance, thereby providing no support for Hypothesis 1a. In model 5, the effect of the interaction between *typicality* and *investor sentiment* is negative but weakly significant ($p = 0.078$), providing only weak support for Hypothesis 2a.

The overall effect of *attractiveness* on *first-day returns* is positive and significant in models 3 to 7, supporting Hypothesis 1b. In model 4, a one standard deviation increase in *attractiveness* leads to an increase in *first-day returns* of 4.64%. In models 6 and 7, the effect of the interaction between *attractiveness* and *investor sentiment* is positive and significant (at $p < 0.01$), supporting Hypothesis 2b. In Model 6, a one standard deviation increase in *attractiveness* when *investor sentiment* is one standard deviation above the sample mean leads to a gain in *first-day returns* of 7.32%, with nearly half of this effect (2.96%) due to the moderating effect of *investor sentiment*. By contrast, a one standard deviation increase in *attractiveness* when *investor sentiment* is one standard deviation below the sample mean leads to a smaller gain of 1.41% in *first-day returns* due to the moderating effect of *investor sentiment.* Figure 1 plots the effects of both *typicality* and *attractiveness* on *first-day returns* using the results from models 4.

The effects of *attractiveness*, when taking into account the moderating effects of *investor sentiment* is economically strong. From Model 6, a one standard deviation increase in *attractiveness* when *investor sentiment* is one standard deviation above the sample mean leads to a gain in first-day returns of 7.32%. As a point of comparison, a one standard deviation increase in *market hotness* leads to a gain of 6.82% in first-day returns while being backed by venture capitalists leads to a gain of 4.85% in first-day returns. Figure 2 plots the effects of both *typicality* and *attractiveness* on *first-day returns* as a function of *investor sentiment* using the results from model 5 and model 6.

-- *Insert Table 5 and Figure 1 and 2 about here* --

Robustness checks. We conducted a range of supplementary analyses beyond the different calibrations of the Doc2Vec procedure. First, the effect of the typicality of a firm depends on the typicality of other industry members. Indeed, a typicality of 0.5 (for example) does not hold the same meaning in an industry where members' similarity to the prototype is generally above 0.8 or in an industry where all members' similarity to the prototype is below 0.2. To account for the *leniency* of the category (Pontikes & Barnett, 2015), we used a new measure of similarity to the industry prototype, which we call *relative typicality*. In these tests, we divided the similarity of an issuing firm to its industry prototype by the average similarity of established firms in the industry to the prototype. Results of this analysis are shown in Table 6 in models 8 to 11. Our overall results continue to be supported when using this measure of the *typicality* of an issuing firm.

-- *Insert Table 6 about here* --

In unreported robustness checks, we used other alternative specifications of our two main independent variables. We measured *typicality* at the level of the 4- and 3-digit SIC codes to see whether our results for typicality (or lack thereof) were influenced by the coarseness of the industry classification used. Our results are robust to these alternative measures of *typicality*.

While the gist of our theory, and we think its interest, is in arguing that investors are sensible to the similarity of a focal issuing firm to recently successful IPOs irrespective of their belonging to the same industry, we acknowledge that attractiveness can be approached at the industry level and that investors may compare issuing firms to recently successful IPOs in the same industry. We therefore created a measure of *industry-specific attractiveness* at the 2-digit SIC code level. Hypotheses 1b is still supported under this alternative specification but not Hypothesis 2b.

Discussion

We underlined that the literature on typicality has generally approached appeal as resulting from conformity to relatively stable prototypes, while ignoring a second component of appeal—attractiveness, or similarity to recent successes. We explored this insight in the context of IPOs, proposing that typical firms exhibit less information asymmetry while attractive firms are expected to deliver higher future performance. We found little support for the direct effect of typicality on first-day returns (hypothesis 1a). In line with hypothesis 1b, we found attractiveness to have a positive impact on first-day returns. When investor sentiment is high, typicality has a negative effect on first-day returns (marginally significant), due to a reduced need for underpricing, while the positive effect of attractiveness on first-day returns is increased due to investors' greater interest in attractive firms (highly significant). The effects that we uncover have a magnitude on par with much more established effects, such as the hotness of the IPO market or of being backed by venture capitalists.

Contribution to the literature on appeal. Based on these results, our proposed framework offers the possibility to think of typicality and attractiveness as two distinct components of a firm's appeal to investors, and to redress the imbalance in prior research that focused on typicality. While typical entities can be appealing due to the greater certainty attached to their value, attractive firms can be appealing due to their perceived superior competencies. These two forms of appeal may have different consequences for valuation, depending on the context being studied. Notably, the importance of one or the other dimension in valuations, and whether or not it will increase or reduce appeal, likely depends on audiences' inclinations toward either the safety of typical solutions, including the taken-for-grantedness of categories used in the market, or the expected superior performance of attractive solutions (Paolella & Durand, 2016; Pontikes, 2012; Smith, 2011; Zhao et al., 2018; Zuckerman, 2017). To advance toward a more comprehensive theory of audiences' valuation of organizations based on the two components of appeal, future research could try to identify contextual factors that would render attractiveness appealing/unappealing to audiences.

Contribution to computational approaches to organizations. In the study of organizations, natural language processing techniques are attracting more and more attention. Topic models have made a notable entry into management scholars' methodological toolbox and have been used to gain supplementary insights into the structure of corpora and to measure certain constructs of interest, such as innovation or cultural heterogeneity (Corritore, Goldberg, & Srivastava, 2019; Croidieu & Kim, 2017; DiMaggio, Nag, & Blei, 2013; Kaplan & Vakili, 2015). Scholars have also used the bag-of-words model, in which documents are represented as unordered counts of each word within them, or dictionary-based approaches that rely on a lexicon of words associated with such sentiments as positivity, negativity, or uncertainty (Hoberg & Phillips, 2016; Loughran & McDonald, 2013; Tetlock, Saar-Tschansky, & Macskassy, 2008).

We propose a method to represent firms as vectors in a shared high-dimensional space by leveraging the power of a natural language processing technique called Doc2Vec (Dai et al., 2015; Lau & Baldwin, 2016; Le & Mikolov, 2014). Using a single corpus of IPO prospectuses and annual reports, we represent both issuing firms and established firms in a shared space, "observe" industry prototypes, and then measure both typicality and attractiveness to assess their respective effects. Whereas, thus far, researchers have been painstakingly reconstructing such data by hand, our novel method allows the computation of industry prototypes in continuous time and with high reliability, providing, to our knowledge, the most direct test of extant theories on typicality. Concomitantly, this method enables the measurement of not only the similarity of any focal firm to any other firm but also of their attractiveness. While this approach is comparable with other attempts to represent categories and their members in high-dimensional spaces in the cognitive literature (Nosofsky, Sanders, & McDaniel, 2018; Verbeemen, Vanpaemel, Pattyn, Storms, & Verguts, 2007), to our knowledge, this is among the first time such an approach has been proposed in the management literature and used to test both well-established and novel propositions on typicality and attractiveness. Our proposed computation of industry prototypes and attractiveness can be transposed to other settings with few modifications, offering the opportunity to standardize our measurements of typicality and attractiveness across studies and fields.[8]

Contribution to bridging organizational and financial approaches to IPOs. Our theory tries to bridge the financial and socio-cognitive literatures by identifying typicality as related to information asymmetry among investors and separating its effects from attractiveness. We emphasized the information asymmetry argument in our discussion of the main effect of typicality on first-day returns as it resonates well with accounts of the effects of typicality in the category literature. Our non-finding suggests that information asymmetry created by departing from the central tendency within one's industry is not associated with underpricing. This implies that underwriters do not consider typicality as a potential source of information asymmetry when pricing the shares of issuing firms.

Our results can be used to cross-fertilize socio-cognitive and financial approaches to IPOs. On the one hand, our results resonate with behavioral accounts that present investors as being more inclined to take risks in periods of high returns (Baker & Wurgler, 2006, 2007). On the other hand, they are in line with the proposition that attractiveness plays an important role for external audiences (Zhao et al., 2018). Thus, these two socio-cognitive mechanisms are likely at

[8] For example, a researcher could leverage movie review websites (Goldberg et al., 2016; Hsu et al., 2009) and use either the content of reviews or movies' descriptions to represent them in a shared space, and then represent the prototype of a movie genre as the centroid of the vectors of films belonging to this genre, while measuring attractiveness as similarity to movies with the highest box office revenue. Similar applications could have relevance for nearly any other setting, including leveraging not only annual reports, press releases, and companies' sustainability reports but also review websites, e-commerce platforms, forums, or any other widely distributed and available textual content for which it is possible to locate organizations and products in a shared space and then directly measure similarity relationships of interest.

play when considering the phenomenon of high first-day returns, as driven by investors' evaluations of issuing firms. Future research might explore how typicality and attractiveness relate to each other and whether attractiveness affects underwriters' setting of the offer price.

Limitations and conclusion

Our paper suffers from three main limitations. First, the endogeneity pervasive to the IPO setting prevents us from making strong causal claims regarding the effects of typicality and attractiveness on first-day returns. The correlations that we unveil seem robust to many different specifications, and the effects we uncover are strong enough to warrant further exploration. Note that our use of numerous fixed-effects controls for possible trend effects and for unobserved heterogeneity among underwriters, industries, and stock exchanges. Second, we created industry prototypes based on SIC codes, which may not accurately capture the industry boundaries that investors use. Yet, firms that do not fit well into the SIC code classifications could arguably be precisely those firms that are considered atypical, so that our measure of typicality would still hold some relevance. Third, the context of IPOs may not generalize well, as investors interested in IPOs may, compared with more traditional investors, present both a higher inclination for attractive firms and a reduced preference for typicality. Future research could investigate the generalizability of our results to other investment settings. Beyond the specific hypotheses tied to the IPO context, this limitation does not invalidate our broader claims: that a firm's appeal to audiences is shaped by both the firm's typicality and its attractiveness. Valuation is not cold science but is bathed in a shared space of meanings and representations that our paper helps unveil, leaving much more room for future discoveries.

REFERENCES

Abrahamson, E., & Fairchild, G. 1999. Management fashion: Lifecycles, triggers, and collective learning processes. ***Administrative Science Quarterly***, 44(4): 708–740.

Aldrich, H. E., & Fiol, C. M. 1994. Fools rush in? The institutional context of industry creation. ***Academy of Management Review***, 19(4): 645–670.

Arthurs, J. D., Hoskisson, R. E., Busenitz, L. W., & Johnson, R. A. 2008. Managerial agents watching other agents: Multiple agency conflicts regarding underpricing in IPO firms. ***Academy of Management Journal***, 51(2): 277–294.

Baker, M., & Wurgler, J. 2006. Investor Sentiment and the Cross-Section of Stock Returns. ***The Journal of Finance***, 61(4): 1645–1680.

Baker, M., & Wurgler, J. 2007. Investor Sentiment in the Stock Market. ***The Journal of Economic Perspectives***, 21(2): 129–151.

Biais, B., & Faugeron-Crouzet, A. M. 2002. IPO Auctions: English, Dutch, … French, and Internet. ***Journal of Financial Intermediation***, 11(1): 9–36.

Brau, J. C., & Fawcett, S. E. 2006. Initial Public Offerings: An Analysis of Theory and Practice. ***The Journal of Finance***, 61(1): 399–436.

Brown, G. W., & Cliff, M. T. 2004. Investor sentiment and the near-term stock market. ***Journal of Empirical Finance***, 11(1): 1–27.

Carter, R., & Manaster, S. 1990. Initial Public Offerings and Underwriter Reputation. ***The Journal of Finance***, 45(4): 1045.

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

Corritore, M., Goldberg, A., & Srivastava, S. B. 2019. Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance. ***Administrative Science Quarterly***, 0001839219844175.

Croidieu, G., & Kim, P. H. 2017. Labor of Love: Amateurs and Lay-expertise Legitimation in the Early U.S. Radio Field. ***Administrative Science Quarterly***, 0001839216686531.

Dai, A. M., Olah, C., & Le, Q. V. 2015. Document Embedding with Paragraph Vectors. ***NIPS 2014 Deep Learning Workshop***. http://arxiv.org/abs/1507.07998.

Deephouse, D. L. 1996. Does isomorphism legitimate? ***Academy of Management Journal***, 39(4): 1024–1039.

Derrien, F. 2005. IPO Pricing in “Hot” Market Conditions: Who Leaves Money on the Table? ***The Journal of Finance***, 60(1): 487–521.

DeVault, L., Sias, R., & Starks, L. 2019. Sentiment Metrics and Investor Demand. ***The Journal of Finance***, 74(2): 985–1024.

DiMaggio, P. J., & Powell, W. W. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. ***American Sociological Review***, 48(2): 147–160.

DiMaggio, P., Nag, M., & Blei, D. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. ***Poetics***, 41(6): 570–606.

Durand, R., & Paolella, L. 2013. Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory. ***Journal of Management Studies***, 50(6): 1100–1123.

Goldberg, A., Hannan, M. T., & Kovacs, B. 2016. What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. ***American Sociological Review***, (2): 215.

Goldstein, M. A., Irvine, P., Kandel, E., & Wiener, Z. 2009. Brokerage Commissions and Institutional Trading Patterns. ***Review of Financial Studies***, 22(12): 5175–5212.

Goldstein, M. A., Irvine, P., & Puckett, A. 2011. Purchasing IPOs with Commissions. ***Journal of Financial and Quantitative Analysis***, 46(05): 1193–1225.

Gondat-Larralde, C., & James, K. R. 2008. IPO pricing and share allocation: The importance of being ignorant. ***The Journal of Finance***, 63(1): 449–478.

Granqvist, N., Grodal, S., & Woolley, J. L. 2013. Hedging Your Bets: Explaining Executives' Market Labeling Strategies in Nanotechnology. ***Organization Science***, 24(2): 395–413.

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

Hanley, K. W. 1993. The underpricing of initial public offerings and the partial adjustment phenomenon. ***Journal of Financial Economics***, 34(2): 231–250.

Hanley, K. W., & Hoberg, G. 2010. The Information Content of IPO Prospectuses. ***Review of Financial Studies***, 23(7): 2821–2864.

Hannan, M. T., Mens, G. L., Hsu, G., Kovács, B., Negro, G., et al. 2019. ***Concepts and Categories: Foundations for Sociological and Cultural Analysis*** (1 edition). New York: Columbia University Press.

Hoberg, G., & Phillips, G. 2016. Text-Based Network Industries and Endogenous Product Differentiation. ***Journal of Political Economy***, 124(5): 1423–1465.

Houge, T., Loughran, T., Suchanek, G., & Yan, X. 2001. Divergence of Opinion, Uncertainty, and the Quality of Initial Public Offerings. ***Financial Management***, 30(4): 5–23.

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

Hsu, G., Hannan, M. T., & Pólos, L. 2011. Typecasting, Legitimation, and Form Emergence: A Formal Theory*. ***Sociological Theory***, 29(2): 97–123.

Hsu, G., Koçak, Ö., & Hannan, M. T. 2009. Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests. ***American Sociological Review***, 74(1): 150–169.

Ibbotson, R. G., Sindelar, J. L., & Ritter, J. R. 1994. The Market's Problems with the Pricing of Initial Public Offerings. ***Journal of Applied Corporate Finance***, 7(1): 66–74.

Jurafsky, D., & Martin, J. H. 2009. ***Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition / Daniel Jurafsky, James H. Martin***. Upper Saddle River, N.J. : Pearson Prentice Hall, c2009.

Kaplan, S., & Vakili, K. 2015. The double-edged sword of recombination in breakthrough innovation. ***Strategic Management Journal***, 36(10): 1435–1457.

Kim, M., & Ritter, J. 1999. Valuing IPOs. ***Journal of Financial Economics***, 53(3): 409–437.

Kovács, B., & Hannan, M. T. 2015. Conceptual Spaces and the Consequences of Category Spanning. ***Sociological Science***, 2: 252–286.

Lau, J. H., & Baldwin, T. 2016. An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. ***Proceedings of the 1st Workshop on Representation Learning for NLP***, 78–86. Berlin, Germany: Association for Computational Linguistics.

Le, Q., & Mikolov, T. 2014. Distributed Representations of Sentences and Documents. ***PMLR***, 1188–1196. Presented at the International Conference on Machine Learning.

Lee, P. M. 2001. What's in a name.com?: The effects of '.com' name changes on stock prices and trading activity. ***Strategic Management Journal***, 22(8): 793–804.

Loughran, T., & McDonald, B. 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. ***Journal of Financial Economics***, 109: 307–326.

Loughran, T., & McDonald, B. 2017. The Use of EDGAR Filings by Investors. ***Journal of Behavioral Finance***, 18(2): 231–248.

Loughran, T., & Ritter, J. 2004. Why has IPO underpricing changed over time? ***Financial Management***, 5–37.

Loughran, T., & Ritter, J. R. 2002. Why Don't Issuers Get Upset about Leaving Money on the Table in IPOs? ***The Review of Financial Studies***, 15(2): 413–443.

Massini, S., Lewin, A. Y., & Greve, H. R. 2005. Innovators and imitators: Organizational reference groups and adoption of organizational routines. ***Research Policy***, 34(10): 1550–1569.

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

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), ***Advances in Neural Information Processing Systems 26***: 3111–3119. Curran Associates, Inc.

Miller, E. M. 1977. Risk, Uncertainty, and Divergence of Opinion. ***The Journal of Finance***, 32(4): 1151–1168.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. 2018. Tests of an exemplar-memory model of classification learning in a high-dimensional natural-science category domain. ***Journal of Experimental Psychology: General***, 147(3): 328–353.

Paleari, S., Signori, A., & Vismara, S. 2014. How Do Underwriters Select Peers When Valuing IPOs? ***Financial Management***, (4): 731.

Paolella, L., & Durand, R. 2016. Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market. ***Academy of Management Journal***, 59(1): 330–351.

Pollock, T. G., & Rindova, V. P. 2003. Media Legitimation Effects in the Market for Initial Public Offerings. ***Academy of Management Journal***, 46(5): 631–642.

Pollock, T. G., Rindova, V. P., & Maggitti, P. G. 2008. Market watch: Information and availability cascades among the media and investors in the US IPO market. ***Academy of Management Journal***, 51(2): 335–358.

Pontikes, E. G. 2012. Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations. ***Administrative Science Quarterly***, 57(1): 81–118.

Pontikes, E. G., & Barnett, W. P. 2015. The persistence of lenient market categories. ***Organization Science***, (5): 1415.

Purnanandam, A. K., & Swaminathan, B. 2004. Are IPOs Really Underpriced? ***The Review of Financial Studies***, 17(3): 811–848.

Ritter, J. R., & Welch, I. 2002. A review of IPO activity, pricing, and allocations. ***The Journal of Finance***, 57(4): 1795–1828.

Rock, K. 1986. Why new issues are underpriced. ***Journal of Financial Economics***, 15(1–2): 187–212.

Roosenboom, P. 2012. Valuing and pricing IPOs. ***Journal of Banking & Finance***, 36(6): 1653–1664.
Smith, E. B. 2011. Identities as lenses: How organizational identity affects audiences' evaluation of organizational performance. ***Administrative Science Quarterly***, 56(1): 61–94.
Suchman, M. C. 1995. Managing Legitimacy: Strategic and Institutional Approaches. ***The Academy of Management Review***, 20(3): 571–610.
Tetlock, P. C., Saar-Tschansky, M., & Macskassy, S. 2008. More than words: Quantifying language to measure firms' fundamentals. ***The Journal of Finance***, 63(3): 1437–1467.
Verbeemen, T., Vanpaemel, W., Pattyn, S., Storms, G., & Verguts, T. 2007. Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach☆. ***Journal of Memory and Language***, 56(4): 537–554.
Zhao, E. Y., Ishihara, M., Jennings, P. D., & Lounsbury, M. 2018. Optimal Distinctiveness in the Console Video Game Industry: An Exemplar-Based Model of Proto-Category Evolution. ***Organization Science***, 29(4): 588–611.
Zuckerman, E. W. 2004. Structural Incoherence and Stock Market Activity. ***American Sociological Review***, 69(3): 405–432.
Zuckerman, E. W. 2017. The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 31–68. Emerald Publishing Limited.

FIGURES

Figure 1. Effects of typicality and attractiveness on first-day returns (model 4)

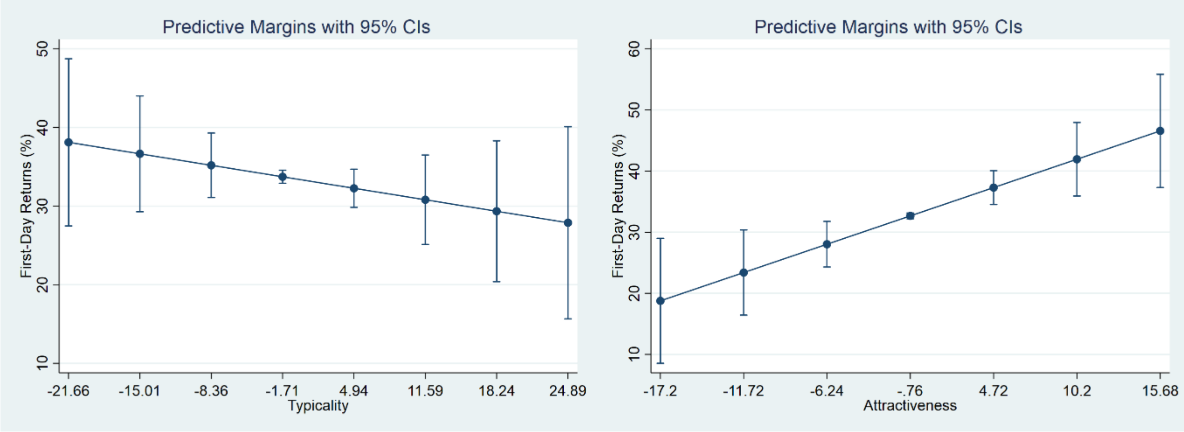
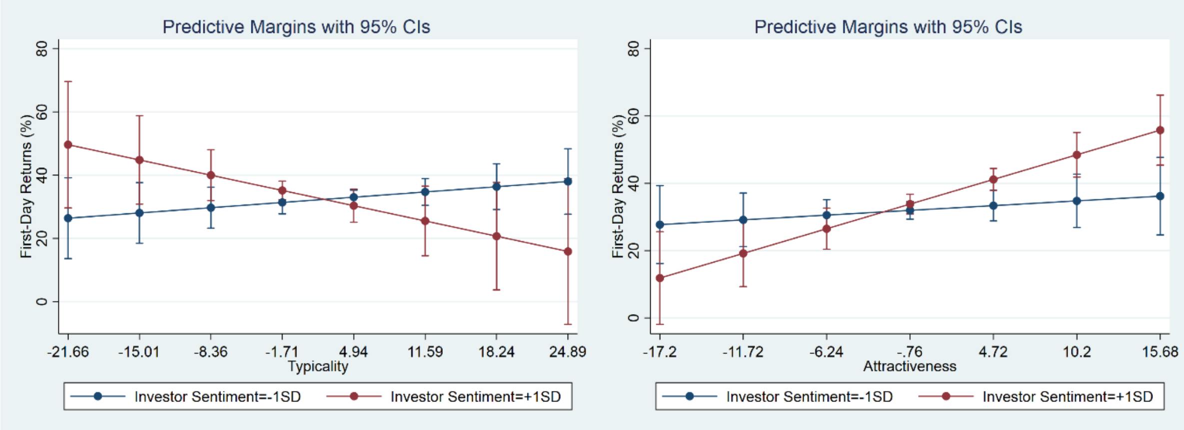


Figure 2. Effects of typicality and attractiveness on first-day returns as a function of investor sentiment (model 5 and 6)



TABLES

Table 1. Three selected IPOs and their five most similar peers (in descending order)

IPO	**Facebook Inc**	Our mission is to make the world more open and connected. People use Facebook to stay connected with their friends and family, to discover what is going on in the world around them, and to share and express what matters to them to the people they care about.	**New York & Company Inc**	We are a leading specialty retailer of fashion-oriented, moderately-priced women's apparel, serving our customers for over 86 years.	**Conatus Pharmaceuticals Inc**	We are a biotechnology company focused on the development and commercialization of novel medicines to treat liver disease.
First Most Similar Peer	Zynga Inc	Zynga Inc. is the world’s leading provider of social game services with 240 million average MAUs, in over 175 countries. We develop, market and operate online social games as live services played over the Internet and on social networking sites and mobile platforms.	Galyan's Trading Company	Galyan’s Trading Company, Inc. is a specialty retailer that offers a broad range of outdoor and athletic equipment, apparel, footwear and accessories, as well as casual apparel and footwear.	Vanda Pharmaceuticals Inc	Vanda Pharmaceuticals Inc. is a biopharmaceutical company focused on the development and commercialization of products for the treatment of central nervous system disorders.
Second Most Similar Peer	Jive Software	We were incorporated in Delaware in February 2001, with a mission to change the way that work gets done. We believe that our social business software unleashes creativity, drives innovation and improves productivity by increasing engagement within the enterprise	Aéropostale Inc	Aeropostale, Inc. is a mall-based specialty retailer of casual apparel and accessories that targets both young women and young men aged 11 to 20.	ADMA Biologics Inc	ADMA Biologics is a specialty immune globulin company that develops, manufactures and intends to market plasma-based biologics for the treatment and prevention of certain infectious diseases.
Third Most Similar Pee	Broadsoft Inc	We are the leading global provider of software and services that enable mobile, fixed-line and cable service providers to deliver Unified Communications and other voice and multimedia services over their Internet protocol, IP, based networks.	Wilsons Leather	We are the leading specialty retailer of quality leather outerwear, accessories and apparel in the United States.	Cytokinetics Inc	We are a clinical-stage biopharmaceutical company focused on the discovery and development of novel small molecule therapeutics that modulate muscle function for the potential treatment of serious diseases and medical conditions.
Fourth Most Similar Peer	TripAdvisor Inc	We are the world’s largest online travel company, empowering users to plan and have the perfect trip. Our travel research platform aggregates reviews and opinions from our community	Charlotte Russe Inc	Charlotte Russe Holdings, Inc. is a rapidly growing, mall-based specialty retailer of fashionable, value-priced apparel and accessories targeting young women between the ages of 15 and 35.	Amicus Inc	We are a biopharmaceutical company focused on the discovery, development and commercialization of small molecule drugs known as pharmacological chaperones.
Fifth Most Similar Peer	Aruba Networks Inc	Aruba Networks is a leading provider of next-generation network access solutions for the mobile enterprise.	American Eagle Outfitters Inc	American Eagle Outfitters, Inc., a Delaware corporation, is a leading lifestyle retailer that designs, markets and sells our own brand of relaxed, casual clothing for 15 to 25 year olds, providing high-quality merchandise at affordable prices.	Lantheus Medical Imaging	We are a global leader in developing, manufacturing and distributing innovative diagnostic medical imaging agents and products that assist clinicians in the diagnosis of cardiovascular diseases such as coronary artery disease, congestive heart failure and stroke, peripheral vascular disease and other diseases.

Table 2. Established firms most similar to the selected industry prototypes for 2000, 2007, and 2014 (in descending order)

2000		2007		2014	
SIC Code 73: Business Services					
Netegrity, Inc.	Netegrity, Inc. provides enterprise security software solutions.	Versant Corp.	Versant Corp. develops and delivers database management software for complex and mission-critical applications.	Artec Global Media Inc.	Artec Global Media, Inc. provides online marketing and reporting solutions.
BindView Development Corp.	BindView Development designs and develops software suites for security and compliance management services.	Lawson Software Americas Inc.	Lawson Software Americas, Inc. develops and provides enterprise resource planning software to manufacturing, distribution, maintenance, and service sector industries.	Interactive Intelligence Group, Inc.	Interactive Intelligence Group, Inc. provides software and cloud services for customer engagement, communications, and collaboration worldwide.
Optika Inc.	Optika Inc. provides enterprise content management (ECM) for imaging, workflow, collaboration, and records management software.	TIBCO Software Inc.	TIBCO Software Inc. provides infrastructure and business intelligence software worldwide.	AllDigital Holdings, Inc.	AllDigital Holdings, Inc. provides digital broadcasting solutions to develop, operate, and support complex digital service and digital broadcasting workflow implementations across various devices.
Eprise Corporation	Eprise Corporation develops content management software products and services that enable businesses to create and publish web content.	NetManage Inc.	NetManage Inc., along with its subsidiaries, develops and markets software and service solutions that enable customers to access their corporate business applications, processes, and data.	eGain Corporation	eGain Corporation provides customer service infrastructure solutions for companies involved in electronic commerce.
SIC Code 28: Chemical and Allied Products					
Alteon Inc.	Alteon Inc. drugs are designed to inhibit or block damage caused by advanced glycosylation end-products (AGE), which are the result of elevated levels of glucose.	Kosan Biosciences Inc.	Kosan Biosciences Inc. develops anticancer agents through clinical procedures.	TG Therapeutics, Inc.	TG Therapeutics, Inc. is a clinical-stage biopharmaceutical company focused on the acquisition, development, and commercialization of innovative pharmaceutical products for the treatment of cancer and other underserved therapeutic needs.
OXiGENE Inc.	OXiGENE Inc. is a biopharmaceutical company primarily focused on the development of vascular disrupting agents (VDAs) for the treatment of cancer.	Neurobiological Technologies Inc.	Neurobiological Technologies, Inc. (NTI) is a biopharmaceutical company focused on the clinical development and regulatory approval of neuroscience drugs.	MEI Pharma Inc.,	MEI Pharma, Inc. is an oncology company focused on the clinical development of novel therapeutics targeting cancer metabolism.
Aronex Pharmaceuticals, Inc.	Aronex Pharmaceuticals, Inc. identifies and develops proprietary innovative medicines to treat cancer and infectious diseases.	La Jolla Pharmaceutical Company	La Jolla Pharmaceutical Company, a biopharmaceutical company, focuses on the discovery, development, and commercialization of therapeutics for life-threatening diseases.	Puma Biotechnology, Inc.	Puma Biotechnology, Inc., a biopharmaceutical company, focuses on the development and commercialization of products to enhance cancer care in the United States.
Repligen Corporation	Repligen Corporation develops, manufactures, and sells products used to enhance the interconnected phases of the biological drug manufacturing process worldwide.	Genta Inc.	Genta Incorporated, a biopharmaceutical company, engages in the identification, development, and commercialization of novel drugs for the treatment of cancer and related diseases.	Oncothyreon, Inc.	Oncothyreon, Inc. is a biotechnology company specializing in the development of innovative therapeutic products for the treatment of cancer.

Table 3. The 5 firms with the highest level of first-day returns in 2000, 2007, and 2014 and the top 5 most attractive firms in 2001, 2008 and 2015

Successful Firms in Year Y			Attractive Firms in Year Y+1	
Firm Name	**SIC Code**	**First-Day Returns**	**Firm Name**	**SIC Code**
2000			**2001**	
VA Linux Systems Inc.	5961	697.5	Network Engines Inc.	7373
Foundry Networks Inc.	3576	525	Onvia.com Inc.	7370
Freemarkets Inc.	7389	483.33	MatrixOne Inc.	7372
Cobalt Networks Inc.	7373	482.39	Coolsavings.com Inc	7370
MarketWatch.com Inc.	7370	473.53	Extensity Inc.	7372
2007			**2008**	
Chipotle Mexican Grill Inc.	5812	100	3PAR Inc.	3572
Isilon Systems	3572	77.69	Data Domain Inc.	3572
Acme Packet Inc.	3576	67.47	Veraz Networks Inc.	3576
Riverbed Technology Inc.	3576	56.92	Compellent Technologies Inc.	7373
Heelys	3140	55.24	Netezza Corp.	3570
2014			**2015**	
Sprouts Farmers Market LLC	5411	122.83	The Habit Restaurants Inc.	5812
Potbelly Corp.	5812	119.79	Zoe's Kitchen Inc.	5812
Noodles & Co.	5812	104.17	Connecture Inc.	7370
Benefitfocus Inc.	7370	102.08	Papa Murphy's Holdings Inc.	5812
Foundation Medicine	2836	96.39	El Pollo Loco Holdings Inc.	5812

Table 4. Descriptive statistics and correlations

	Mean	s.d.	Minimum	Median	Maximum	Underp.	Typical.	Attractiveness	Bull–Bear Spread	Price Rev.	Mkt Hotness	VC Back.	Log of Age	Log of Assets
Underpricing	32.82	62.44	–63.64	13.12	697.5	1.00								
Typicality	0	6.65	–21.65	–0.28	26.5	–0.16	1.00							
Attractiveness	0	5.48	–17.25	–0.46	20.55	0.19	0.09	1.00						
Investor Sentiment	0	0.17	–0.49	–0.01	0.42	0.17	–0.15	–0.07	1.00					
Price Revision	0.86	14.4	–65	0	73.91	0.41	–0.11	0.14	0.16	1.00				
Market Hotness	45.88	22.53	0	45	100	0.36	–0.16	0.07	0.33	0.41	1.00			
VC Backing	0.58	0.49	0	1	1	0.19	0.00	0.28	0.07	0.11	0.11	1.00		
Log of Age	2.3	0.93	0	2.2	5.08	–0.20	0.13	–0.12	–0.08	–0.14	–0.17	–0.34	1.00	
Log of Assets	4.16	1.9	2.65	3.91	11.82	–0.20	0.14	–0.19	–0.14	–0.08	–0.18	–0.45	0.43	1.00

Table 5. OLS regressions of first-day returns on typicality and attractiveness

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Typicality		–0.151		–0.220	–0.238	–0.227	–0.248
		(0.242)		(0.250)	(0.260)	(0.247)	(0.257)
Attractiveness			0.823**	0.846**	0.858**	0.797*	0.805**
			(0.286)	(0.302)	(0.299)	(0.301)	(0.297)
Typ#Inv Sent					–2.866+		–3.139+
					(1.592)		(1.566)
Att#Inv Sent						3.174**	3.508**
						(1.146)	(1.116)
Investor Sentiment	7.111	6.902	7.868	7.586	6.267	7.908	6.498
	(8.389)	(8.399)	(8.067)	(8.050)	(8.167)	(7.921)	(8.890)
Price Revision	1.038***	1.037***	1.026***	1.024***	1.023***	1.036***	1.037***
	(0.134)	(0.136)	(0.138)	(0.140)	(0.141)	(0.143)	(0.143)
Market Hotness	0.306***	0.306***	0.303***	0.303***	0.295***	0.303***	0.294***
	(0.080)	(0.080)	(0.076)	(0.076)	(0.074)	(0.074)	(0.071)
VC Backing	5.368+	5.525+	4.254	4.452	4.757+	4.850+	5.226+
	(2.815)	(2.803)	(2.701)	(2.691)	(2.576)	(2.826)	(2.705)
Log of Age	–2.412+	–2.323+	–2.212+	–2.078	–2.079	–1.925	–1.911
	(1.339)	(1.373)	(1.263)	(1.281)	(1.262)	(1.223)	(1.196)
Log of Assets	–1.495	–1.483	–1.018	–0.987	–0.787	–0.941	–0.717
	(1.311)	(1.280)	(1.334)	(1.292)	(1.214)	(1.296)	(1.221)
Industry FE	YES	YES	YES	YES	YES	YES	YES
Year Effects	YES	YES	YES	YES	YES	YES	YES
Bookrunner FE	YES	YES	YES	YES	YES	YES	YES
Stock Exch. FE	YES	YES	YES	YES	YES	YES	YES
Observations	1.989	1.989	1.989	1.989	1.989	1.989	1.989
Adjusted R–squared	0.294	0.294	0.297	0.297	0.299	0.299	0.301

Robust standard errors are shown in parentheses; *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.1

Table 6. OLS regressions of first-day returns on relative typicality and attractiveness

VARIABLES	Model 8	Model 9	Model 10	Model 11
Rel Typicality	–0.105	–0.135	–0.137	–0.141
	(0.115)	(0.118)	(0.120)	(0.120)
Attractivity		0.852**	0.852**	0.792*
		(0.303)	(0.301)	(0.299)
RelTyp#Inv Sent			–1.108	–1.425+
			(0.780)	(0.749)
Att#Inv Sent				3.916**
				(1.129)
Investor Sentiment	6.827	7.530	6.560	6.683
	(8.372)	(8.030)	(7.966)	(8.255)
Price Revision	1.036***	1.023***	1.023***	1.038***
	(0.137)	(0.141)	(0.141)	(0.143)
Market Hotness	0.306***	0.304***	0.300***	0.299***
	(0.080)	(0.076)	(0.073)	(0.070)
VC Backing	5.655*	4.583+	4.599+	5.095+
	(2.793)	(2.680)	(2.662)	(2.770)
Log of Age	–2.297+	–2.058	–2.099+	–1.924
	(1.363)	(1.270)	(1.250)	(1.183)
Log of Assets	–1.478	–0.978	–0.830	–0.730
	(1.267)	(1.281)	(1.202)	(1.204)
Industry FE	YES	YES	YES	YES
Year Effects	YES	YES	YES	YES
Bookrunner FE	YES	YES	YES	YES
Stock Exch. FE	YES	YES	YES	YES
Observations	1,989	1,989	1,989	1,989
Adjusted R-squared	0.294	0.297	0.299	0.301

Robust standard errors are shown in parentheses; *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.1

APPENDIX A.
Tests of Dimensionality

Table A1. OLS regressions of first-day returns on typicality and attractiveness when using 100 dimensions (model A7) or 200 dimensions (model B7) for document vectors

VARIABLES	Model A7	Model B7
Typicality	–0.122	–0.134
	(0.241)	(0.195)
Attractiveness	0.625*	0.421*
	(0.268)	(0.193)
Typ#Inv Sent	–3.354*	–2.401*
	(1.401)	(1.128)
Att#Inv Sent	3.553***	1.656*
	(0.798)	(0.795)
Investor Sentiment	6.806	6.879
	(9.810)	(8.969)
Controls	YES	YES
Industry FE	YES	YES
Year Effects	YES	YES
Bookrunner FE	YES	YES
Stock Exch. FE	YES	YES
Observations	1,989	1,989
Adjusted R-squared	0.302	0.298

Robust standard errors are shown in parentheses; *** p < 0.001, ** p < 0.01, * p < 0.05, + p < 0.1

In Table A1, we show the results obtained for Models 7 to 8 using vectors of 200 and 100 dimensions instead of 300, but using the same specifications otherwise. As can be seen, the results are not substantially altered.

CHAPTER 3
Valuing Organizations: An Integrated Theory

Abstract [137 words]

Three different models exist in the literature (the prototype-based, goal-based and exemplar-based models) that tie an audience's use of categories and its valuation of organizations. These models disagree about the relationship between an organization's typicality and its valuation. We integrate these different models in a single theory and propose that a focal audience's valuations depend on whether pre-existing prototypes align with its center of interest. We then show how audience heterogeneity and the breadth of audiences' centers of interest influence the relationship between typicality and valuation in non-trivial ways. Finally, we deduce how the distribution of audiences' centers of interest influences the formation of new categories. Our proposed theory has consequences for theoretical models of valuation, for studies on audiences' heterogeneity, for the literature on category formation and more broadly for research on optimal distinctiveness and legitimacy.

Introduction

Audiences constantly value organizations and their products using categories. Venture capitalists categorize start-ups to assess their potential (Wry et al, 2014), CSR agencies categorize and value the environmental and social performance of organizations (Hawn & Ioannou, 2016) and wine critics assess wines based on their modes of production (Negro, Hannan, & Fassiotto, 2015). Due to their ubiquity and relevance, research on categories has boomed and counts hundreds of contributions a year (Cattani, Porac, & Thomas, 2017; Durand & Thornton, 2018). To value organizations and their products, audiences use contextually relevant categories, such as ‘fintech’, ‘green business’ or ‘biodynamic’. They favor entities belonging to positively valued categories and generally penalize entities which are hard to categorize (Hannan et al., 2019; Hsu, 2006; Hsu, Koçak, & Hannan, 2009). Thus, categorization is intrinsically linked to valuation and models of audiences’ categorization of organizations are also models of audiences’ valuation of organizations.

Most often, audiences value organizations and their products using pre-existing, well-established categories (Hannan et al., 2019). These pre-existing categories are defined by a prototype, i.e. an abstract representation of the most representative member of a category (Mervis & Rosch, 1981; Rosch, 1975). The more an entity is similar to the prototype of a category -the higher its typicality-, the higher its valuation (Hsu et al., 2009; Leung & Sharkey, 2014; Negro & Leung, 2013). This prototype-based model of valuation accounts for audiences’ valuations in many but not all contexts. Sometimes, audiences dynamically create *ad hoc* categories to achieve specific goals (Glaser, Krikorian Atkinson, & Fiss, 2019; Granqvist & Ritvala, 2016; Paolella & Durand, 2016; Pontikes & Kim, 2017). In such conditions, audiences value more positively entities which are similar to their representation of the ideal tool for the achievement of their goals, irrespective of their similarity with pre-existing prototypes. Audiences also often use salient exemplars, such as recent organizational successes or consecrated members of a category, as a yardstick to value other entities (Barlow, Verhaal, & Angus, 2019; Zhao, Ishihara, Jennings, & Lounsbury, 2018). The literature thus identifies three different models of audiences' valuation of organizations and their products: the prototype-based model, the goal-based model, and the exemplar-based model.

While these different models offer precious insights on how audiences value organizations, little theoretical efforts have been dedicated so far to articulating all of them together in a coherent framework. Yet, if we were to integrate these different perspectives on audiences' valuation, we would be able to reconcile the conflicting findings that fostered their development and to generate enriched predictions regarding important organizational phenomena. We thus tackle the task of trying to understand when and why audiences sometimes behave as prototype-based, goal-based or exemplar-based evaluators and what the consequences are of these different behaviours.

All three models of valuation assume that before any valuation of a specific entity takes place, audiences find most appealing the combination of features defining the prototype, ideal or exemplar that they currently use as a basis for their valuation, which we name their *center of interest*. Hence, an audience's center of interest is the focus of the audience's attention and the basis used by audiences to assess empirical reality (i.e. products, organizations) with respect to what they expect (a prototype, an ideal, or an exemplar). What then determines an audience's valuation of a focal entity is the similarity between the entity and the combination of features characterizing the audience's current center of interest, i.e. whether the entity is *aligned with the audience's center of interest*.

Using this as a starting point, we develop in the paper that when a pre-existing prototype aligns with an audience's center of interest, they will tend to value typical entity more positively and thus behave as a prototype-based evaluator. When there is no pre-existing prototype aligned with an audience's center of interest, the audience can rely on conceptual combination and behave as a goal-based evaluator, or on a few salient features and behave as an exemplar-based evaluator. We then show that depending on the breadth of audiences' centers of interest and audiences' propensity to have the same or different centers of interest, the relationship between typicality and valuation can either be positive, negative or an inverted U-shape. We finally deduce which distributions of audiences' centers of interest are most likely to lead to the formation of a new category in the absence of pre-existing categories.

We make three main contributions to the category literature with this paper. First, we articulate all three existing models of valuation, proposing conditions that determine a focal audience's valuations as a function of its center of interest. Second, we explain and demonstrate that audiences' propensity to have the same or different centers of interest impacts the relationship between typicality and valuation. As such, we determine the conditions that account for the (seemingly) contradictory empirical results found so far (Askin & Mauskapf, 2017; Hannan et al., 2019; Paolella & Durand, 2016; Zhao et al., 2018). Third, we deduce which distributions of audiences' centers of interest are the most likely to lead to the formation of a new category when pre-existing categories are absent, thus contributing to the literature on category formation (Durand & Khaire, 2017; Kennedy, 2008; Navis & Glynn, 2010). Finally, our proposed theory has broader implications for the literature on optimal distinctiveness (Navis & Glynn, 2011; Zuckerman, 2016) and legitimacy (Bitektine, 2011; Suchman, 1995).

From prototypes to goals to exemplars: introducing the three models of valuation developed in organization theory

The prototype-based model of valuation is heavily influenced by research on category learning in social-psychology. According to this research, there exists natural categories which reflects the correlational structure of the world (Mervis & Rosch, 1981; Rosch, 1975). We learn categories by observing their members and abstracting from their features a summary representation of all exemplars observed, i.e. by learning a prototype for the focal category (Reed, 1972). We then categorize newly observed entities as members of a known category based on their similarity with its prototype. Entities which are similar to a category's prototype constitute typical members of the category while entities which are dissimilar from it are atypical -they lack 'family resemblance' (Rosch & Mervis, 1975). For example, the robin is a typical bird because it has many features that we associate with the prototypical bird -it is small, it sings nicely, it flies, it nests in trees, etc.- while the penguin is an atypical bird because many of its features make it dissimilar from our idea of the prototypical bird -it is rather large, it swims but does not fly, it lives on the ground, etc. Categorizing an entity as a member of a category leads people to make inferences on the unobserved features that it possesses and help them set their expectations – in fact it is arguably the main purpose of categorization (Murphy, 2016). For example, categorizing an event as a party sets expectations regarding who will come, what activities will be involved and how to behave during the event (Cantor, Mischel, & Schwartz, 1982).

According to the prototype-based model of valuation, audiences categorize organizations and their products just like they do with any other entities, by drawing on a pool of pre-existing, contextually relevant categories (Hannan et al., 2019). Within the prototype-based model of valuation, categorization impacts valuation through two main channels. First, entities exhibiting features which do not resemble those of prototypes known by audiences are harder to categorize; this generates disfluency in the processing of organizations' features and leads audiences to penalize atypical entities (Hannan et al., 2019; Hsu et al., 2009). Second, audiences ascribe positive or negative valence to pre-existing categories which impacts the valuation of members of these categories (Kennedy, Lo, & Lounsbury, 2010). In that regard, a common assumption in the literature is that audiences have a positive view of pre-existing categories. In other words, entities which are typical members of pre-existing categories have some intrinsic appeal in the eyes of audiences, which thus value them more positively (Hannan et al., 2019; Hsu, 2006).

These two mechanisms lead to the main prediction of the prototype-based model, which is that audiences value typical entities more positively. This proposition has received large and consistent support in numerous settings. Movies and books fitting into existing genres tend to receive more positive critics, while auctions on ebay are less likely to result in a sell when the auctioned item span categories (Hsu, 2006; Hsu et al., 2009; Kovács & Hannan, 2010). Wines produced by wineries spanning multiple styles of winemaking receive lower ratings when tasters know the winery spans styles (Negro & Leung, 2013). Prospective borrowers who apply for a loan on an online social network receive less money when they affiliate with multiple social groups on their profile (Leung & Sharkey, 2014).

Recent research has deepened the prototype-based model of valuation for organization and management studies. Studies using data from Netflix and Yelp show that some movie-viewers and restaurant-goers assign better ratings to atypical movies and restaurants than to typical ones (Goldberg, Hannan, & Kovacs, 2016). This suggests that audiences have heterogenous cultural preferences in terms of typicality without challenging the building blocks upon which the prototype-based model of valuation is built. Furthermore, two complementary models of valuation emerged, the goal-based and exemplar-based models of valuation, which both stress that audiences do not solely categorize organizations and their products using pre-existing categories.

The goal-based model of valuation proposes that audiences sometimes use *ad hoc* categories to find entities which can help them achieve their current goals (Durand & Paolella, 2013). Goal-based categories are not defined by a prototype but instead by an ideal which depends on the goal which one seeks to achieve (Barsalou, 1985). For example, an ideal for the goal-based category 'foods to eat on a diet' is 'zero calories'. Ideals are formed through a process of conceptual combination, which consists in combining knowledge stored in memory to identify which features will be the most useful in achieving one's intended goal(s) (Barsalou, 1991). Since audiences do not share the same goals at all point in time and audiences idiosyncratically derive goal-based categories with little coordination, goal-based categories guarantee stable market exchanges only if audiences and organizations at least agree on the dimensions that can be combined to derive new goal-based categories – i.e. reduce the heterogeneity of the feature space (Glaser et al., 2019).

Within the goal-based model of valuation, audiences have a theory of value which specifies their goals as well as the ideal means to achieve them (Zuckerman, 2017). Hence, audiences value organizations whose features align with the ideals defined by their current goals more positively (Durand & Paolella, 2013). In other words, it is not similarity to pre-existing prototypes which is conducive of a higher valuation but similarity to audiences' ideals. One important consequence is that in context where audiences' goals lead them to define ideals combining features from instances of multiple categories, audiences will value atypical organizations more positively than typical ones. For instance, Paolella and Durand (2016) find that clients of law firms, who face complex legal situations requiring expertise in many different fields, tend to prefer atypical law firms which do not specialize in a single category of legal services. Thus, the goal-based model of valuation does not pre-suppose the absence of pre-existing categories, it proposes that in certain conditions audiences are more likely to rely on goal-based categories defined by ideals than on pre-existing categories defined by prototypes.

By contrast, the exemplar-based model of valuation was introduced in organization theory for contexts void of pre-existing categories. In such a setting, audiences cannot rely on pre-existing categories, so that salient exemplars provide a much-needed yardstick for audiences, as well as an attractive combination of features (Zhao et al., 2018). The exemplar-based model of valuation thus initially proposed that, absent pre-existing categories, entities which are similar to salient exemplars attract more demand from audiences and are valued more positively. A recent addition suggests that this mechanism is operational even in the presence of pre-existing categories and that the benefits of being similar to successful exemplars fade away if one is also similar to established prototypes (Barlow et al., 2019) [9].

The exemplar-based model of valuation is inspired by research on category learning which suggests that humans remember all previously observed members of a category -i.e. exemplars- and categorize new entities based on their similarity with these previously observed instances (Beatu & Shultz, 2010; Cohen & Basu, 1987; Homa, Sterling, & Trepel, 1981). However, organizational research generally focuses on *salient* exemplars rather than all possible exemplars of a pre-existing category. Four mechanisms can render an exemplar salient to audiences. First, its outstanding success can lead it to stand out from its peers (Barlow et al., 2019; Zhao et al., 2018). Second, its features may overemphasize some of the key features associated with its category (Durand & Kremp, 2016). Third, an exemplar can become salient if its features offer a concrete representation of extant theorization within a field or offer a basis for new theorization (Nigam & Ocasio, 2010). Fourth, an exemplar can be consecrated as a salient member of its category through the active involvement of dedicated audiences (Jones & Massa, 2013).

We summarize the main features of all three models of valuation in Table 1 and move on to integrating these three models in a single coherent framework.

[9] Note that Zhao and colleagues consider the possibility of an inverted U-shape relationship between similarity to salient exemplars and valuation once ‘proto-category’ starts to emerge. However, Barlow and colleagues advocate for a linear relationship between similarity to the exemplar and valuation which is attenuated when there is also high similarity to a pre-existing prototype. We retain the proposition of a linear relationship between similarity to the exemplar and valuation and later offer an alternative take on the articulation of prototype-based and exemplar-based valuations

-- Insert Table 1 about here --

Introducing audiences' centers of interest as the main determinant of their valuations

As illustrated in Table 1, all three models of valuation start with the premise that before any valuation of a specific entity takes place, audiences find certain combinations of features - those defining prototypes, ideals or exemplars- especially appealing. This could be because of their tastes, e.g. a book-reader may have a preference for certain genres such as Science-Fiction or Fantasy (Kovács & Hannan, 2015). This could be because they are seeking to achieve a specific goal, e.g. find a 'restaurant to take someone on a first-date'. This could be because they have a specific exemplar in mind, e.g. an historian of architecture might evaluate ecclesiastic buildings of the 20th century using Unity Temple by Frank Lloyd Wright as a yardstick (Jones & Massa, 2013).

Furthermore, all models hold that entities which are the most similar to the combination of features defined *ex ante* by a focal audience as especially appealing receive the best valuations from this audience (cf. Table 1). For example, a recent formalization of the prototype-based model of valuation proposes that members of a category receive a portion of the value ascribed to it by audiences as a function of their typicality (Hannan et al., 2019). It follows that in all three models the combination of features that is the *most* appealing to a focal audience is precisely the one they define *ex ante* to be the basis of her valuation. Hence, all three models are specific instances of a single model where each audience holds a specific combination of features to be the most appealing and then value observed entities based on their similarity to this combination. In other words, while the three models of valuation present audiences as behaving in different ways, they ultimately offer three different perspectives on a single mechanism.

To integrate all three models of valuation we propose to define the combination of features which is the most appealing to a focal audience in a given context as the audience's current *center of interest*. We say that an entity aligns with an audience's center of interest when the entity is similar to the combination of features defining the audience's center of interest. Audiences value entities aligned with their center of interest more positively. We choose the phrase *center of interest* because it integrates different dimensions of the three models of valuation quite nicely. First, what usually qualifies as 'a center of interest' in daily conversations can correspond to either a pre-existing category, a goal-based category or a salient exemplar. For example, depending on context, one's center of interest can be jazz music, songs to listen to while working or a playlist of opus akin to Beethoven's fifth symphony. Second, the notion of *center of interest* echoes the spatial metaphor which supports most models of valuation (Askin & Mauskapf, 2017; Barlow et al., 2019; Haans, 2019; Kovács & Hannan, 2015; Paolella & Durand, 2016; Pontikes & Hannan, 2014; Zhao et al., 2018). This metaphor generally holds that audiences 'locate' entities in a conceptual or feature 'space' and that similarity to a focal 'point' -thought of in terms of its 'distance' from it- determines valuation. Third, saying that something is someone's center of interest does not necessarily entail that the person's attention is geared toward objects she finds interesting or that she is actively seeking those. Indeed, according to the Merriam-Webster dictionary, an interest is "a feeling that accompanies or causes special attention to something or someone". Thus, saying that an audience finds an entity interesting -i.e. aligned with its center of interest- does not presuppose that its attention is geared only toward finding interesting objects. One's attention may be driven to interesting objects while actively seeking them (as in the goal-based model of valuation) or as a result of finding them interesting (as in the prototype-based or exemplar-based models of valuation).

We assume for this first part of our framework that there exist pre-existing categories shared among audiences. We believe this situation to be the most frequent in the organizational world, as audiences generally try to make sense of their situation using pre-existing categories to in turn define their line of actions (Hannan et al., 2019). Even when creating new categories, audiences and organizations often recombine the features of existing ones (Khaire & Wadhwani, 2010; Rao, Monin, & Durand, 2005). We later release this assumption to discuss category formation and change.

The prototype-based model of valuation assumes that audiences' centers of interest align with pre-existing prototypes: audiences prefer to eat typical food in typical restaurants, to read typical books, to watch typical movies, to drink typical wine from typical producers, and to lend money to typical people (Hsu, 2006; Hsu et al., 2009; Kovács & Hannan, 2015; Leung & Sharkey, 2014; Negro & Leung, 2013). However, the goal-based model and the exemplar-based models of valuation do not make such assumptions. An audience may purposively define its ideal to be atypical (Paolella & Sharkey, 2017; Pontikes, 2012) or rely on salient exemplars to identify canonical members of a category rather than the average ones (Durand & Kremp, 2016). Yet, it is also possible that audiences' ideals correspond to pre-existing prototypes (Zuckerman, 2017). For example, the ideal 'fun movie to watch with kids' might very well be the prototypical family comedy. It is also possible that an audience structures her valuation using an exemplar which is very similar to its category's prototype. As an illustration, while the category of GTA-like video games is well-established and defined by a prototype abstracted from the features of many different games (those of the GTA series of course, but also of the Watchdog series, the Assassin's Creed series, the Mafia series, the Driver series, the Saints row series, etc.), it would still be hard to distinguish the features of this prototype from those of the GTA games.

Since ideals and exemplars can sometimes be similar to pre-existing prototypes, the conceptual distinction between prototype-based valuations on the one hand and goal-based and exemplar-based valuations on the other hand can sometimes become blurry or even collapse. Furthermore, it is unclear what would be the value of maintaining this distinction in such cases. Consider an audience using an ideal which is highly similar to a pre-existing prototype (as in the 'fun movie to watch with friends' case) or an exemplar nearly identical to its category's prototype (as in the GTA-like case) as a basis for her valuation. In both cases, it matters little to know whether the audience's center of interest consists in an ideal reached through conceptual combination or a salient exemplar. In effect, they use a basis akin to a pre-existing prototype for their valuation and as a result value typical entities more positively. We thus propose that when a pre-existing prototype aligns with an audience's center of interest, they will behave as a prototype-based evaluator (independently of the cognitive process involved). Conversely, if an audience value typical entities more positively, it follows that a pre-existing prototype aligns with their center of interest. If it weren't the case, entities aligned with their center of interest but dissimilar from pre-existing prototypes would be valued more positively than typical entities. Hence:

> **Assumption 1.** *A pre-existing prototype aligns with a focal audience's center of interest if and only if they behave as a prototype-based evaluator*

It follows from Assumption 1 that if a pre-existing prototype does not align with an audience's center of interest, they behave either as a goal-based or an exemplar-based evaluator. One key distinction between the goal-based and the exemplar-based model is the breadth of the features defining ideals versus that defining exemplars. The goal-based model of valuation presents audiences as actively combining features to generate categories which will help them solve their goals (Barsalou, 1985, 1991). This mechanism of conceptual combination is generally triggered when audiences are confronted with a complex situation, requiring flexible solutions fit to their every needs (Durand & Paolella, 2013; Paolella & Durand, 2016). For instance, in the online display advertising industry, advertisers dynamically create very elaborate categories to target highly specific prospects such as "the single, 20-something-year-old Asian male who graduated from University of Toronto" (Glaser et al., 2019). Thus, goal-based categories help audiences valuing organizations not by reducing the complexity of the environment but by embracing it through the combination of multiple features.

By contrast, audiences rely on the few core features associated with salient exemplars to simplify the valuation of newly encountered entities. While an exemplar is generally rendered salient first by its recognition as a commercial and/or critical success, it becomes recognizable through its association with a few core features identified and constructed as such in the public discourse surrounding it (Jones & Massa, 2013; Nigam & Ocasio, 2010; Zhao et al., 2018). Audiences can then use these features as a basis to value other entities. Thus, if an audience's center of interest does not align with a pre-existing prototype, the breadth of her center of interest – i.e. the breadth of the combination of features that they find most appealing – determines whether they behave as a goal-based or an exemplar-based evaluator:

> **Proposition 1a.** *If a focal audience's center of interest does not align with pre-existing prototypes and is broad, they will behave as a goal-based evaluator*
> **Proposition 1b.** *If a focal audience's center of interest does not align with pre-existing prototypes and is narrow, they will behave as an exemplar-based evaluator*

We summarize in Figure 1 how the two factors we discussed interact to determine audiences' valuations. We so far focused on the valuations of individual audiences. We now move on to consider multiple audiences and how the alignment of their centers of interest with pre-existing prototypes shapes the relationship between typicality and valuation.

-- Insert Figure 1 about here –

Revisiting the relationship between typicality and valuation as a function of audiences' propensity to have the same or different centers of interest

The prototype-based model of valuation generally assumes that audiences are homogeneous, i.e. that their members all value typical entities more positively due to their attachment to established categories. One important recent development in the literature on categories and valuation is the observation that audiences can have heterogeneous preferences for typicality in some context (Goldberg et al., 2016; Pontikes, 2012; Smith, 2011). This results in situations where some audiences value typical entities more positively while others value them more negatively. However, our preceding discussion suggests that audiences' valuations depend on both the alignment of prototypes with their centers of interest (i.e. their inclination for typical offerings) and on the breadth of their interest. Furthermore, Assumption 1 and Propositions 1a and 1b suggest that different audiences may adopt different behaviours as evaluators depending on their centers of interest.

This latter proposition resonates with insights from both the goal-based and the exemplar-based models of valuation. Within the goal-based model of valuation, audiences use highly idiosyncratic, audience-specific goal-based categories, so that audiences could be relying on very different ideals when categorizing entities and thus have very different centers of interest (Glaser et al., 2019). However, it is also possible that audiences sharing a common goal will reach similar ideals and as a result share the same centers of interest (Zuckerman, 2017). Within the exemplar-based model of valuation, there is no *a priori* reason to assume that all audiences would share an interest in the same exemplars nor is there reason to assume that they would not sometimes do. Thus, to be able to account for the relationship between typicality and valuation, it is necessary to consider both 1) the breadth of audiences centers of interest and 2) audiences' propensity to have the same or different centers of interest.

We illustrate our reasoning as we develop it in this section in Figures 2a, 2b, 2c, 3a and 3b. For the purpose of this illustration, we consider a setting with two pre-existing categories in which six different audiences are valuing organizations. Categories, audiences' centers of interest and organizations can be associated with four different features. In each figure, we cover a large number of patterns of associations of organizations with features. In so doing, our goal is to be relatively exhaustive in our illustration of how organizations' associations with features affect their valuation by audiences.

Panel A, B and C are identical in all figures. In Panel A of each figure, we present the two categories used in our illustration, category 1 and 2. The association of each category with each feature represents its prototype. The prototype of category 1 is most strongly associated with feature 1 and partially with feature 2, and the prototype of category 2 is most strongly associated with feature 4 and partially with feature 3. In Panel B of each figure, we present to our reader the organizations that we use to illustrate our arguments. Like categories, organizations are associated with features. The sum of each organization's associations with features equals 1 as organizations with limited resources can either invest heavily in a few features or spread themselves thin to acquire many different features. We consider all patterns of organizations' associations with features in which a focal organization's association with a given feature has one decimal digit[10]. In Panel B, we also indicate how we compute the typicality of each organization based on its similarity to the prototypes of category 1 and 2. An organization's similarity to the prototype of a given category is computed as 2 minus the sum over features of the absolute value of the difference between the organization's association with each feature and that of the category's prototype. Thus, if an organization's association with each feature is identical to that of the category's prototype, its similarity to the category's prototype is 2. By contrast, if an organization is associated with none of the features associated with a category's prototype, its similarity to the category's prototype is 0. An organization's typicality is then the maximum of its similarity to the prototypes of category 1 and category 2. Hence, organizations which have a high level of typicality are very

[10] In other words, we consider all organizations such that a focal organization's association with a given feature is either 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 or 1, with the additional constraint that the sum of the organization's associations with features equals 1. This leads to a total of 286 organizations.

similar to the prototype of a pre-existing category while organizations which are atypical are not similar at all to the prototype of any pre-existing category.

In Panel C of each figure, we present examples of organizations along with their level of typicality. Given their large number, we could not show all the organizations included in our illustrations. We thus show only selected examples, to give to our reader a sense of the different patterns of organizations' associations with features that we consider. Organization 1 is only associated with feature 4, which is a core feature of category 2. Thus, its typicality is high, at 1.60. Organization 3 has the same association with features than the prototype of category 2. Thus, its typicality is 2. Organization 34 is associated with three different features, with a moderate association with the features of category 2, so its typicality remains above 1, at 1,2. Organization 142 is associated weakly with all features. Its typicality is thus intermediary, at 1. Organization 11 and 66 are associated with only one feature, which is not a core feature of any category. Thus, their typicality is weak, at 0.4. Organizations 191, 224, 282 and 286 mirror organizations 1, 2, 34 and 142 but features 1 and 2 plays the roles of features 4 and 3.

In Panel D of each figure, we present audiences' centers of interest. Panel D changes from one figure to another as we examine different distributions of audiences' centers of interest. Like for categories and organizations, the strength of the association between an audience's center of interest and a feature is given by a number between 0 and 1. The sum of an audience's center of interest's associations with features equals 1 to reflect the fact that the broader the audience's center of interest, the less the audience has an interest in any particular feature.

In Panel E of each figure, we present how we compute the value of each organization in the eyes of each audience. Panel E does not change from one figure to another. In line with the idea that an organization's value in the eyes of an audience is a linear function of its typicality (Hannan et al., 2019), we take an organization's value in the eyes of an audience to be equal to its alignment with the audience's center of interest. We compute the alignment of an organization with each audience's center of interest using the same formula than to compute an organization's similarity to a category's prototype.

Finally, in Panel F of each figure, we plot the relationship between typicality and the average value of organizations across all audiences. We fit a trend line using ordinary least square regression to be able to observe the general direction of the relationship between typicality and valuation given audiences' centers of interest (Panel D) and all the organizations included in our analysis (Panel B and C).

The figures that we produce are a simple illustration. They are intended to show to our reader which kind of organizations tend to gain or lose value when audiences' centers of interest vary, considering numerous possible associations of organizations with features. These figures by no means constitute a demonstration of our arguments nor a test of our theory. They illustrate our main points as well as the way one can think about the relationship between typicality, valuation and audiences' centers of interest when using our theory.

We first consider situations in which audiences cluster in small groups, each group gathering audiences with very similar and narrow centers of interest. Importantly, audiences' centers of interest may differ significantly from one group to another. This kind of situations corresponds to settings where audiences have different tastes as is for example the case in markets for cultural products such as movies or books: some audiences prefer action movies to drama while others appreciate above all romantic comedies or gangster movies (Hsu, 2006). This kind of situations may also correspond to settings where different audiences use different exemplars to categorize organizations. For example, the office rental company WeWork before its failed IPO was considered by some the Uber of shared offices while others compared it to Regus, a leading and well-established company in the co-working industry. Most importantly, this kind of situations can correspond to settings where some audiences define their centers of interest in terms of pre-existing prototypes while others seek specific combinations of features.

Figures 2a, 2c and 2b illustrates the three general situations described in the preceding paragraph. In Panel D of Figure 2a, audiences are clustered in two groups of the same size sharing similar centers of interest. The prototype of category 1 closely aligns with the centers of interest of audiences 1 to 3 while the prototype of category 2 closely aligns with the centers of interest of audiences 4 to 6. Thus, pre-existing prototypes align with audiences' centers of interest. By contrast, in Panel D of Figure 2b, we present a case where audiences' centers of interest do not align with pre-existing prototypes, although audiences are still clustered in two small groups sharing similar centers of interest. Finally, in Panel D of Figure 2c, we show a case in which audiences are divided between prototype-based and exemplar-based evaluators: the centers of interest of audiences 1 and 2 align with the prototype of category 1, the centers of interest of audiences 5 and 6 align with the prototype of category 2 but the centers of interest of audiences 3 and 4 align with neither.

-- *Insert Figure 2a, 2b and 2c about here* --

We start with the two extreme cases in which audiences cluster in small groups whose centers of interest either all align with pre-existing prototypes (as illustrated in Figure 2a) or all lie away from them (as illustrated in Figure 2b). The first case corresponds to a situation in which all audiences are prototype-based evaluators and is the one assumed by and studied in extant research on prototype-based categories (Hsu et al., 2009; Leung & Sharkey, 2014; Negro & Leung, 2013). In this case, the relationship between typicality and valuation is positive. The second case corresponds to a situation in which the proportion of exemplar-based evaluators is one and there are no prototype-based evaluators. In such a situation, the centers of interest of all audiences lie away from pre-existing prototypes (Proposition 1c).

Thus, typical entities generally do not align with audiences' centres of interest and hence tend to be valued more negatively. Moreover, atypical entities are more likely to align with audiences' centers of interest as they both lie away from prototypes. Hence, audiences are also more likely to value atypical entities more positively. This is for example the case in the Google Play platform, where audiences' centers of interest gravitate toward salient apps rather than prototypes, whose features are blurry and generic in this setting (Barlow et al., 2019). Since in this situation audiences tend to value typical entities more negatively and atypical entities more positively, the relationship between typicality and valuation is negative.

This argument is illustrated in Figures 2a and 2b. In Panel F of Figure 2a, we see that the relationship between typicality and valuation is clearly positive: organizations with a high level of typicality are valued more positively while atypical organizations are valued more negatively. In Panel F of Figure 2b, the pattern is reversed: atypical organizations are valued more positively while typical organizations are valued more negatively.

Overall, we make the following propositions:

> **Proposition 2a.** *If audiences cluster in small groups whose narrow centers of interest align with pre-existing prototypes, there is a positive relationship between typicality and valuation*
> **Proposition 2b.** *If audiences cluster in small groups whose narrow centers of interest do not align with pre-existing prototypes, there is a negative relationship between typicality and valuation*

Our discussion of the two extreme cases in which there are only prototype-based or only exemplar-based evaluators among audiences suggest that the proportion of exemplar-based evaluators is an important factor to consider. When the proportion of exemplar-based evaluators rises, there will be less audiences whose centres of interest align with pre-existing prototypes, and thus typical entities will lose some of their value. By contrast, there will be more audiences whose centres of interest lie away from prototypes, which will increase the value of atypical entities. Thus, when audiences cluster in small groups sharing similar centers of interest, a rising proportion of audiences whose centers of interest do not align with pre-existing prototypes, i.e of exemplar-based evaluators, attenuates the positive relationship between typicality and valuation and ultimately reverses it.

Our reasoning is illustrated in Figure 2c. In Panel D of Figure 2c, audiences are clustered in four different groups sharing similar centers of interest. The centers of interest of audiences 1, 2, 5 and 6 are the same than in Figure 2a. Thus, the prototype of category 1 closely aligns with the centers of interest of audiences 1 and 2 while the prototype of category 2 closely aligns with the centers of interest of audiences 5 and 6. The centers of interest of audience 3 and 4 are the same than in Figure 2b. They thus do not align with pre-existing prototypes. Panel F of Figure 2c shows that in this configuration, the positive relationship between typicality and valuation is attenuated relative to the one presented in Panel F of Figure 2a. The orange dots capture organizations' average valuations given audiences centres of interest as depicted in Panel D of Figure 2a while the blue dots represent organizations' average valuations given audiences' centers of interest as depicted in Panel D of Figure 2c. Similarly, the orange trend line shows the relationship between typicality and valuation given audiences centres of interest as depicted in Panel D of Figure 2a while the blue trend line captures the relationship between typicality and valuation given audiences' centers of interest as depicted in Panel D of Figure 2c. The slope of the curve is clearly flatter in this latter case.

All in all, we make the following proposition:

> **Proposition 2c.** *In a situation where audiences are either prototype-based or exemplar-based evaluators, the positive relationship between typicality and valuation will be attenuated and eventually reversed as the proportion of exemplar-based evaluators grows*

We now add goal-based evaluators into the picture. We first consider a situation in which all audiences share highly similar and broad centers of interest, i.e. a situation in which all audiences are goal-based evaluators (by Proposition 1b). To better understand the nature of the relationship between typicality and valuation in such a situation, we distinguish three types of organizations: those which are typical of a category, those which are atypical and associated with a broad combination of features, and those which are atypical and associated with a narrow combination of features. Atypical entities associated with a broad combination of features have a weak association with any particular feature. Thus, their association with features not associated with the prototype of a pre-existing category reduces their typicality only to a small extant. They are also likely to be weakly associated with at least some of the features associated with the prototype of any category. By contrast, atypical entities defined by a narrow combination of features have a strong association with features that are not associated with pre-existing prototypes and thus are much more sharply recognized as atypical. Hence, atypical entities associated with a broad combination of features tend to be less atypical than atypical entities associated with a narrow combination of features.

Since all audiences behave as goal-based evaluators, we do not expect typical entities to be more positively valued (Paolella & Durand, 2016; Paolella & Sharkey, 2017). Moreover, like in the case with only exemplar-based evaluators, since audiences' centers of interest do not align with pre-existing prototypes, audiences generally value typical entities less positively. Meanwhile, atypical entities associated with a broad combination of features are likely to be associated with most of the features defining audiences' centers of interest. Thus, they tend to align with audiences' centers of interest and audiences tend to value them more positively. Finally, atypical entities associated with a narrow combination of features lack many of the features associated with audiences' centers of interest. Hence, they do not align with audiences' centers of interest and audiences tend to value them more negatively. Per our preceding arguments, these entities are also more atypical than entities associated with a broad combination of features.

To wrap up, we thus have a situation where audiences value typical and very atypical entities more negatively while they value mildly atypical entities -those associated with a broad combination of features- more positively. This results in an inverted U-shape relationship between typicality and valuation. The study of this case suggests that introducing goal-based evaluators generally increase the value of entities associated with a broad combination of features, i.e. of mildly atypical entities. Overall, we have:

> **Proposition 3a.** *In a situation where all audiences behave as goal-based evaluators, there is an inverted U-shape relationship between typicality and valuation*
> **Proposition 3b.** *The value of mildly atypical entities relative to typical and very atypical ones grows as the proportion of goal-based evaluators among audiences grows*

Figure 3a illustrates Proposition 3a. The centers of interest of all audiences is broad and weakly associated with all features (Panel D). As a result, mildly typical organizations -i.e. organizations associated with a large number of features- tend to be valued more positively, as shown in Panel F.

-- Insert Figure 3a about here --

Finally, Figure 3b illustrates Proposition 3b. Audiences are clustered in three different groups sharing similar centers of interest. The centers of interest of audiences 1, 2, 5 and 6 are the same than in Figure 2a. Thus, the prototype of category 1 closely aligns with the centers of interest of audiences 1 and 2 while the prototype of category 2 closely aligns with the centers of interest of audiences 5 and 6. The centers of interest of audience 3 and 4 are the same than in Figure 3a. In Panel F, we show in orange the valuation of organizations and the relationship between typicality and valuation given audiences' centers of interest as defined in Panel D of Figure 2a. We show in blue the valuation of organizations and the relationship between typicality and valuation resulting from audiences' centers of interest as depicted in Panel D of Figure 3b. In this configuration, mildly atypical organizations gain in value relative to the two others, altering the curve of the relationship between typicality and valuation relative to the one presented in Panel F of Figure 2a.

-- Insert Figure 3b about here --

Accounting for audiences' centers of interest in the formation of shared categories

Our framework is so far applicable to settings where there are pre-existing categories shared among audiences to categorize entities. However, in order to be able to account for the development of new categories, we relax this assumption and now consider a situation in which there are no pre-existing categories shared among audiences. This situation corresponds to the early days of a category's formation, when an innovation results in the apparition of new entities that have to be introduced to audiences and legitimized (Aldrich & Fiol, 1994; Grodal, Gotsopoulos, & Suarez, 2014; Navis & Glynn, 2010, 2011) or when a set of already existing entities are redefined as members of a new categories by interested agents (Durand & Khaire, 2017; Khaire & Wadhwani, 2010). We specifically focus on the formation of new categories of organizations or products made by organizations and spell out the conditions under which, absent pre-existing categories, certain types of organizations may be more successful than others, leading to the abstraction of prototypes and the formation of categories.

Models of category formation generally emphasize the active involvement of both organizations and audiences into shaping new categories. In the early days of category formation, organizations emphasize the features that they or their products share in the eyes of audiences (Navis & Glynn, 2010). They also contribute to define categorical boundaries by selectively naming other organizations involved in the emerging category in their communication with audiences (Kennedy, 2008). Meanwhile, interested audiences can get heavily involved in shaping new categories, or even foster their creation (Durand & Khaire, 2017; Khaire & Wadhwani, 2010). Notably, audiences' endorsement of specific exemplars function as a signal that a given combination of features is aligned with audiences' centers of interest and encourages organizations to adopt it (Zhao et al., 2018). We place this latter insight at the center of our model of category formation. In the absence of pre-existing categories, organizations are unsure about which combinations of features to adopt. They thus use salient exemplars -i.e. exemplars knowing an outstanding success with audiences- as an indication of where audiences' centers of interest lie (Zhao et al., 2018). New categories thus form around successful exemplars as organizations adopt their features in response to audiences' acclaims.

In a situation without pre-existing categories, audiences' interests do not align with pre-existing prototypes by default. Thus, by Assumption 1 audiences behave either as exemplar-based or goal-based evaluators. Furthermore, audiences can either share the same centers of interest or have very different centers of interest. We start with a situation in which audiences all share the same narrow center of interest as our base case. In this case, entities whose features align with audiences' centers of interest receive a strong endorsement from all audiences. They are thus especially salient to organizations who will then adopt their features. As a result, a new category will form around audiences' common center of interest. This situation corresponds to the situation described by Zhao and colleagues (Zhao et al., 2018).

However, there is no guarantee that audiences will share the same narrow centers of interest. Audiences may cluster in small groups with different centers of interest. In such a situation, different combinations of features are endorsed by different audiences. Organizations thus have a harder time identifying successful others and understanding where audiences' centers of interest lie. Category formation is thus a more difficult and less certain process.

Finally, audiences may have broad rather than narrow centers of interest – i.e. behave as goal-based rather than exemplar-based evaluators. If all audiences share a broad center of interest, entities associated with a broad combination of features are generally valued more positively (cf. our discussion of Propositions 3a and 3b). However, such entities are a poor indicator of which features to adopt for organizations as they are weakly associated with many different features. Many organizations may also not have the resources needed to adopt all these features. Thus, in this situation, category formation is unlikely. More generally, the presence of goal-based evaluators among audiences dilute the success of organizations aligned with the centers of interest of exemplar-based evaluators, rendering them less salient, thus complicating the process of category formation.

All in all, we have:

Proposition 4a. *Category formation is most likely when audiences all behave as exemplar-based evaluators sharing a narrow center of interest.*
Proposition 4b. *The greater the proportion of exemplar-based evaluators with different centers of interest, the lower the likelihood of category formation*
Proposition 4c. *The greater the proportion of goal-based evaluators, the lower the likelihood of category formation*

Discussion

Contribution to the category literature. This paper makes three main contributions to the category literature, a soaring research stream in management and organization studies. First, we articulate and reconcile all three models of valuation in a single theory proposing that before any valuation takes place, audiences already have centers of interests and then value entities aligned with their centers of interest more positively. Depending on the alignment of pre-existing prototypes with audiences' centers of interest and on the breadth of their centers of interest, audiences may behave as either prototype-based, goal-based or exemplar-based evaluators. However, the general mechanism by which they produce their valuations remains the same. Adopting this higher-level lens on valuation allows to go beyond the juxtaposition of different models of valuation suited to different contexts. It opens the possibility of considering how these different models of valuation might co-exist and what the consequences are of being confronted with audiences producing valuations based on different centers of interest. Therefore, an outcome of our paper is that the different and sometimes contradictory empirical findings do not prove wrong the other categorization perspectives. To the contrary, there exist meta-conditions that explain why in some cases valuation of atypical organizations is positive and negative under other conditions. The three main perspective are therefore complementary and we articulate the reasons for these compatibilities. Future research might consider how audiences may shift between different centers of interest over time.

Second, this article revisits the fundamental question of the heterogeneity of audiences and of its impact on the relationship between typicality and valuation. In the literature on categories and valuation, the heterogeneity of audiences refer mostly to their having different inclinations toward typical entities (Goldberg et al., 2016; Pontikes, 2012; Smith, 2011). Continuing this perspective, recent research suggests that a single audience can both value typical firms more positively due to their greater interpretability and value atypical firms more positively if they are better suited for their needs (Paolella & Sharkey, 2017). This paper broadens this discussion and redefines audiences' heterogeneity in terms of their propensity to have the same or different centers of interest. It shows that both this factor and the breadth of audiences' centers of interest interact to shape the relationship between typicality and valuation. Our propositions can be read both as empirical predictions and as a call to be sensible to the importance of considering the distribution of audiences' centers of interest when studying the relationship between typicality and valuation.

Third, this article provides an account of which distributions of audiences' centers of interest are most likely to favour or hinder the formation of new categories. Existing studies have focused on different stages of categories' life, from emergence or creation to maintenance and change (Aldrich & Fiol, 1994; Grodal et al., 2014; Khaire & Wadhwani, 2010; Navis & Glynn, 2010). Our theory specifies which distributions of audiences' centers of interest render the successes of some entities salient to organizations which can then adopt the features of these successful exemplars. In particular, we focus on the heterogeneity of audiences' centers of interests and the role it plays in the formation of categories. We believe this introduces an interesting shift in current discussions that primarily focus on the clarity of the meaning of categories as a determinant of audiences' valuation of organizations (Hannan et al., 2019; Kovács & Hannan, 2010). We point out that categories may rise and fall not only because they are usable -i.e. have a clear meaning- but also because audiences find them interesting (aligned with their centers of interest) (Kennedy et al., 2010; Lo, Fiss, Rhee, & Kennedy, 2019). We leave open for future research what could be the driver of the formation of audiences' centers of interest but we believe that social dynamics such as discourse, power or status play a key role in shaping them (Grodal & Kahl, 2017; Sharkey, 2014; Syakhroza, Paolella, & Munir, 2018).

Beyond categories. We believe that our theory also contributes to various research streams beyond the category literature. Recently, scholars have emphasized that organizations have to balance the need to 'blend in' relative to their peers and to 'stand out' in the eyes of audiences (Navis & Glynn, 2011; Zuckerman, 2016). Such a perspective implicitly relies on a prototype-based account of audiences' valuations. Indeed, one is assumed to 'stand out' when one is distinct from pre-existing prototypes while one 'blends in' by being similar to them (Askin & Mauskapf, 2017; Haans, 2019). Thus, in this view, gaining audiences' attention depends solely on one's position relative to pre-existing prototypes. Our theory broadens this perspective as we no longer assume that audiences' attention is shaped by pre-existing prototypes. Instead, we propose that audiences have centers of interest which may or may not align with pre-existing prototypes. Thus, per our theory, 'standing out' may either be incompatible with 'blending in' -when pre-existing prototypes do not align with audiences' centers of interest- or be perfectly compatible with it -when pre-existing prototypes align with audiences' centers of interest. In other words, whether organizations need to find a balance between ‘blending in’ and ‘standing out’ and to which extent depends on where audiences’ centres of interest lie.

Our proposed theory also relates to research on organizations’ legitimacy. Indeed, cognitive legitimacy often results from organizations’ ability to conform with pre-existing prototypes (Aldrich & Fiol, 1994; Bitektine, 2011; Suchman, 1995). Our theory does not challenge this fundamental insight, but it does suggest that audiences’ centers of interest may sometimes lie away from cognitively legitimate firms. One possible mechanism driving this phenomena could be that in some contexts audiences’ centers of interest are shaped by other factors. For example, audiences may purposively look for creative or novel entities (Seong & Godart, 2017; Taeuscher, Bouncken, & Pesch, 2020). We believe that future research could greatly enrich our understanding of the relationship between legitimacy and valuation by specifying how audiences’ centers of interest are influenced by multiple sources of legitimacy as a function of contextual factors.

Limitations and conclusion

This article has some limitations. First, we focus on audiences’ valuations but the behaviour of organizations also shape categories, as our discussion of category formation suggests, and it may differ significantly from that of audiences, notably with respect to ambiguous categories (Granqvist, Grodal, & Woolley, 2013; Montauti & Wezel, 2016; Pontikes & Barnett, 2015). We thus left aside the question of how audiences’ centers of interest interact with organizations’ strategic decisions for future research, recognizing that our framework could be further enriched by integrating this dimension. We also implicitly assume that audiences’ centers of interest are narrow when they align with pre-existing prototypes, i.e. that prototypes themselves are defined by a narrow combination of features. This might not be the case in ambiguous categories, which are loosely associated with a broad range of features.

We believe that audiences interested in instances of ambiguous categories can be accommodated in our model as behaving in line with the expectations of the goal-based model of valuation (Pontikes & Kim, 2017). Indeed, audiences with an inclination for ambiguity generally purposively look for ambiguous entities such as VCs and market makers (Pontikes, 2012). Another important limitation is that we consider audiences who have only one center of interest. We believe this situation to be the most frequent as in a given market context, one generally has an interest in a specific kind of entities, which does not imply that one does not have multiple centers of interest out of this particular market context (e.g. one may be generally interested in nanotech and fintech, in online and traditional banks, in RPG games and platformers). In any case, one way to account for the potential variety of individuals' centers of interest is to consider that one can belong to multiple audiences, each audience having a single center of interest.

To conclude, we integrate all three models of valuation and provide a framework based on the analysis of audiences' centers of interest to predict audiences' valuations. We revisit how audiences' heterogeneity impacts the relationship between typicality and valuation and discuss how audiences' centers of interest impact category formation. We hope that our framework adds clarity and structure to the multiple perspectives on valuation that have been burgeoning in organization theory and will help researchers get a better understanding of which models of valuation best describe the behaviour of audiences in their field of enquiry. Finally, we hope that by shifting the discussion toward audiences' centers of interest, we open the door to a further integration of multiple branches of research on audiences' valuation to shade an even greater light on this core driver of organizations' success.

REFERENCES

Aldrich, H. E., & Fiol, C. M. 1994. Fools Rush in? The Institutional Context of Industry Creation. ***The Academy of Management Review***, 19(4): 645.

Askin, N., & Mauskapf, M. 2017. What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. ***American Sociological Review***, 82(5): 910–944.

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

Barsalou, L. W. 1985. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. ***Journal of Experimental Psychology. Learning, Memory, and Cognition***, 11(4): 629–654.

Barsalou, L. W. 1991. Deriving Categories to Achieve Goals. In G. H. Bower (Ed.), ***Psychology of Learning and Motivation***, vol. 27: 1–64. Academic Press.

Beatu, I., & Shultz, T. 2010. Development of Prototype Abstraction and Exemplar Memorization. ***Proceedings of the Annual Meeting of the Cognitive Science Society***, 32(32). https://escholarship.org/uc/item/7pm6n0cj.

Bitektine, A. 2011. Toward a Theory of Social Judgments of Organizations: The Case of Legitimacy, Reputation, and Status. ***Academy of Management Review***, 36(1): 151–179.

Cantor, N., Mischel, W., & Schwartz, J. C. 1982. A prototype analysis of psychological situations. ***Cognitive Psychology***, 14(1): 45–77.

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

Cohen, J. B., & Basu, K. 1987. Alternative Models of Categorization: Toward a Contingent Processing Framework. ***Journal of Consumer Research***, 13(4): 455.

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

Durand, R., & Kremp, P.-A. 2016. Classical Deviation: Organizational and Individual Status as Antecedents of Conformity. ***Academy of Management Journal***, 59(1): 65–89.

Durand, R., & Paolella, L. 2013. Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory. ***Journal of Management Studies***, 50(6): 1100–1123.

Durand, R., & Thornton, P. H. 2018. Categorizing Institutional Logics, Institutionalizing Categories: A Review of Two Literatures. ***Academy of Management Annals***, 12(2): 631–658.

Glaser, V. L., Krikorian Atkinson, M., & Fiss, P. C. 2019. Goal-Based Categorization: Dynamic Classification in the Display Advertising Industry. ***Organization Studies***, 017084061988336.

Goldberg, A., Hannan, M. T., & Kovacs, B. 2016. What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. ***American Sociological Review***, (2): 215.

Granqvist, N., Grodal, S., & Woolley, J. L. 2013. Hedging Your Bets: Explaining Executives' Market Labeling Strategies in Nanotechnology. ***Organization Science***, 24(2): 395–413.

Granqvist, N., & Ritvala, T. 2016. Beyond Prototypes: Drivers of Market Categorization in Functional Foods and Nanotechnology. ***Journal of Management Studies***, 53(2): 210–237.

Grodal, S., Gotsopoulos, A., & Suarez, F. F. 2014. The Coevolution of Technologies and Categories During Industry Emergence. ***Academy of Management Review***, 40(3): 423–445.
Grodal, S., & Kahl, S. J. 2017. The Discursive Perspective of Market Categorization: Interaction, Power, and Context. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 151–184. Emerald Publishing Limited.
Haans, R. F. J. 2019. What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories. ***Strategic Management Journal***, 40(1): 3–27.
Hannan, M. T., Mens, G. L., Hsu, G., Kovács, B., Negro, G., et al. 2019. ***Concepts and Categories: Foundations for Sociological and Cultural Analysis*** (1 edition). New York: Columbia University Press.
Hawn, O., & Ioannou, I. 2016. Mind the gap: The interplay between external and internal actions in the case of corporate social responsibility. ***Strategic Management Journal***, 37(13): 2569–2588.
Homa, D., Sterling, S., & Trepel, L. 1981. Limitations of exemplar-based generalization and the abstraction of categorical information. ***Journal of Experimental Psychology: Human Learning and Memory***, 7(6): 418–439.
Hsu, G. 2006. Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production. ***Administrative Science Quarterly***, 51(3): 420–450.
Hsu, G., Koçak, Ö., & Hannan, M. T. 2009. Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests. ***American Sociological Review***, 74(1): 150–169.
Jones, C., & Massa, F. G. 2013. From Novel Practice to Consecrated Exemplar: Unity Temple as a Case of Institutional Evangelizing. ***Organization Studies***, 34(8): 1099–1136.
Kennedy, M. T. 2008. Getting Counted: Markets, Media, and Reality. ***American Sociological Review***, 73(2): 270–295.
Kennedy, M. T., Lo, J. Y.-C., & Lounsbury, M. 2010. Category currency: The changing value of conformity as a function of ongoing meaning construction. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 369–397. Emerald Group Publishing Limited.
Khaire, M., & Wadhwani, R. D. 2010. Changing Landscapes: The Construction of Meaning and Value in a New Market Category—Modern Indian Art. ***Academy of Management Journal***, 53(6): 1281–1304.
Kovács, B., & Hannan, M. T. 2010. The consequences of category spanning depend on contrast. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 175–201. Emerald Group Publishing Limited.
Kovács, B., & Hannan, M. T. 2015. Conceptual Spaces and the Consequences of Category Spanning. ***Sociological Science***, 2: 252–286.
Leung, M. D., & Sharkey, A. J. 2014. Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount. ***Organization Science***, 25(1): 171–184.
Lo, J. Y.-C., Fiss, P. C., Rhee, E. Y., & Kennedy, M. T. 2019. Category Viability: Balanced Levels of Coherence and Distinctiveness. ***Academy of Management Review***. https://doi.org/10.5465/amr.2017.0011.
Mervis, C. B., & Rosch, E. 1981. Categorization of Natural Objects. ***Annual Review of Psychology***, 32(1): 89–115.
Montauti, M., & Wezel, F. C. 2016. Charting the Territory: Recombination as a Source of Uncertainty for Potential Entrants. ***Organization Science***, 27(4): 954–971.

Murphy, G. L. 2016. Is there an exemplar theory of concepts? ***Psychonomic Bulletin & Review***, 23(4): 1035–1042.
Navis, C., & Glynn, M. A. 2010. How New Market Categories Emerge: Temporal Dynamics of Legitimacy, Identity, and Entrepreneurship in Satellite Radio, 1990–2005. ***Administrative Science Quarterly***, 55(3): 439–471.
Navis, C., & Glynn, M. A. 2011. Legitimate distinctiveness and the entrepreneurial identity: Influence on investor judgments of new venture plausibility. ***Academy of Management Review***, 36(3): 479–499.
Negro, G., Hannan, M. T., & Fassiotto, M. 2015. Category Signaling and Reputation. ***Organization Science***, 26(2): 584–600.
Negro, G., & Leung, M. D. 2013. "Actual" and Perceptual Effects of Category Spanning. ***Organization Science***, 24(3): 684–696.
Nigam, A., & Ocasio, W. 2010. Event Attention, Environmental Sensemaking, and Change in Institutional Logics: An Inductive Analysis of the Effects of Public Attention to Clinton's Health Care Reform Initiative. ***Organization Science***, 21(4): 823–841.
Paolella, L., & Durand, R. 2016. Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market. ***Academy of Management Journal***, 59(1): 330–351.
Paolella, L., & Sharkey, A. 2017. Forging Consensus: An Integrated View of How Categories Shape the Perception of Organizational Identity. In R. Durand, N. Granqvist, & A. Tyllström (Eds.), ***Research in the Sociology of Organizations***, vol. 51: 327–353. Emerald Publishing Limited.
Pontikes, E. G. 2012. Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations. ***Administrative Science Quarterly***, 57(1): 81–118.
Pontikes, E. G., & Barnett, W. P. 2015. The persistence of lenient market categories. ***Organization Science***, (5): 1415.
Pontikes, E. G., & Hannan, M. T. 2014. An Ecology of Social Categories. ***Sociological Science***, 1: 311–343.
Pontikes, E. G., & Kim, R. 2017. Strategic Categorization. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 71–111. Emerald Publishing Limited.
Rao, H., Monin, P., & Durand, R. 2005. Border Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy. ***American Sociological Review***, 70(6): 968–991.
Reed, S. K. 1972. Pattern recognition and categorization. ***Cognitive Psychology***, 3(3): 382–407.
Rosch, E. 1975. Cognitive representations of semantic categories. ***Journal of Experimental Psychology: General***, 104(3): 192–233.
Rosch, E., & Mervis, C. B. 1975. Family resemblances: Studies in the internal structure of categories. ***Cognitive Psychology***, 7(4): 573–605.
Seong, S., & Godart, F. C. 2017. Influencing the Influencers: Diversification, Semantic Strategies, and Creativity Evaluations. ***Academy of Management Journal***, 61(3): 966–993.
Sharkey, A. J. 2014. Categories and Organizational Status: The Role of Industry Status in the Response to Organizational Deviance. ***American Journal of Sociology***, 119(5): 1380–1433.
Smith, E. B. 2011. Identities as lenses: How organizational identity affects audiences' evaluation of organizational performance. ***Administrative Science Quarterly***, 56(1): 61–94.

Suchman, M. C. 1995. Managing legitimacy: Strategic and institutional approaches. ***Academy of Management Review***, 20(3): 571–610.
Syakhroza, M. A., Paolella, L., & Munir, K. 2018. Holier than Thou? Identity Buffers and Adoption of Controversial Practices in the Islamic Banking Category. ***Academy of Management Journal***, 62(4): 1252–1277.
Taeuscher, K., Bouncken, R. B., & Pesch, R. 2020. Gaining Legitimacy by Being Different: Optimal Distinctiveness in Crowdfunding Platforms. ***Academy of Management Journal***. https://doi.org/10.5465/amj.2018.0620.
Zhao, E. Y., Ishihara, M., Jennings, P. D., & Lounsbury, M. 2018. Optimal Distinctiveness in the Console Video Game Industry: An Exemplar-Based Model of Proto-Category Evolution. ***Organization Science***, 29(4): 588–611.
Zuckerman, E. W. 2016. Optimal Distinctiveness Revisited. ***The Oxford Handbook of Organizational Identity***. https://doi.org/10.1093/oxfordhb/9780199689576.013.22.
Zuckerman, E. W. 2017. The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 31–68. Emerald Publishing Limited.

TABLES

Table 1. Main features of all three models of valuation

	Categories used in valuation	Basis for categorization	Mechanisms influencing valuation	Examples of categories
Prototype-based model of valuation	Pre-existing categories shared among audiences	The category's prototype, an abstract representation of the most representative member of a category	Similarity to prototypes leads to easier categorization and greater intrinsic appeal resulting in more positive valuation	Movie genres, types of restaurant, industry categories
Goal-based model of valuation	Idiosyncratic, audience-specific categories	The category's ideal, defined as the best combination of features to achieve the goal defining the category	Similarity to ideals suggests a good tool to achieve the goal defining the category resulting in more positive valuation	Law firms that will meet all my legal needs, films to watch after a break-up, things to take on a trip
Exemplar-based model of valuation	Categories defined in terms of a salient exemplar	The features of the salient exemplar defining the category	Similarity to successful exemplars facilitate valuation (possibility to use exemplar as a yardstick) resulting in more positive valuation	The Uber of X, the Mozart of X, GTA-like

FIGURES

Figure 1. Determinants of audiences' valuations given their interests

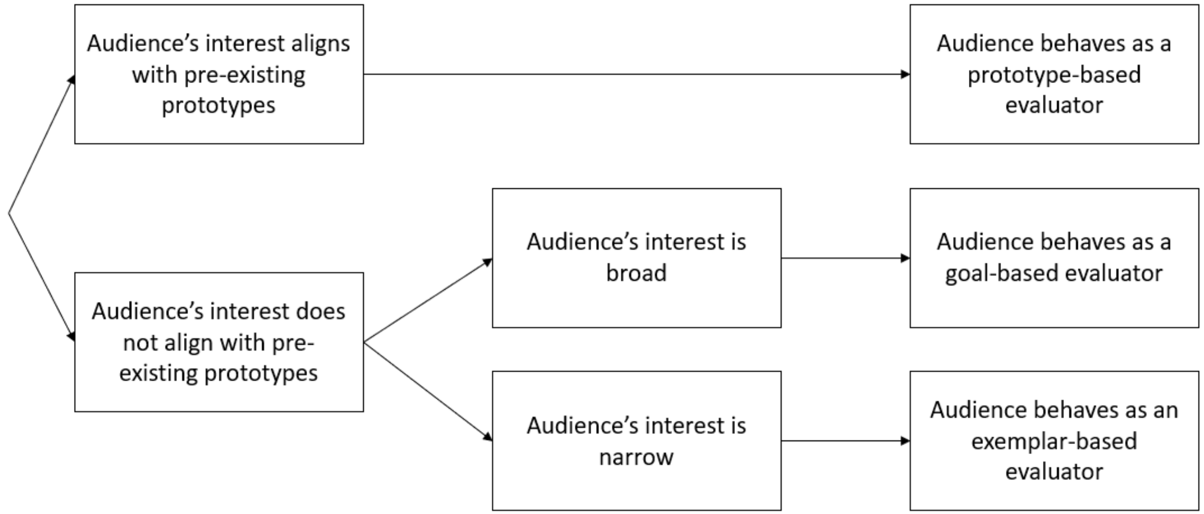


Figure 2a. Illustration of the case where audiences all behave as prototype-based evaluators (Proposition 2a)

Panel A: Categories' association with features

	Feature 1	Feature 2	Feature 3	Feature 4
Category 1	0,8	0,2	0	0
Category 2	0	0	0,2	0,8

The redder the cell, the stronger the association of the category with the feature. The prototype of a category is represented by its association with features. For example, the prototypical member of category 1 has an association of 0,8 with feature 1 and of 0,2 with feature 2.

Panel B: Presentation of the organizations used in this illustration

Organizations are associated with features. The association of an organization with a feature is a digit with at most one decimal ranging between 0 and 1. The sum of each organization's associations with features is 1. We consider all possible patterns of organizations' associations with features given these constraints, which represents a total of 286 organizations.

The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2. Thus, the more an organization is similar to a pre-existing prototype, the higher its typicality. The similarity of an organization to a category's prototype is computed as:

$$2 - \sum_{f \in \left\{\begin{matrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{matrix}\right\}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Category's\ association\ with\ f) \end{matrix}$$

Panel C: Examples of organizations along with their association with features and their level of typicality

	Feature 1	Feature 2	Feature 3	Feature 4	Typicality
Organization 1	0,00	0,00	0,00	1,00	1,60
Organization 3	0,00	0,00	0,20	0,80	2,00
Organization 34	0,00	0,30	0,30	0,40	1,20
Organization 142	0,20	0,20	0,30	0,30	1,00
Organization 11	0,00	0,00	1,00	0,00	0,40
Organization 66	0,00	1,00	0,00	0,00	0,40
Organization 191	0,30	0,30	0,20	0,20	1,00
Organization 224	0,40	0,30	0,30	0,00	1,20
Organization 282	0,80	0,20	0,00	0,00	2,00
Organization 286	1,00	0,00	0,00	0,00	1,60

The redder the cell, the stronger the association between the organization and a feature and the higher the organization's typicality.

Panel D: Audiences' centres of interest

	Feature 1	Feature 2	Feature 3	Feature 4
Audience 1	0,90	0,10	0,00	0,00
Audience 2	0,80	0,10	0,10	0,00
Audience 3	0,85	0,15	0,00	0,00
Audience 4	0,00	0,00	0,10	0,90
Audience 5	0,00	0,10	0,10	0,80
Audience 6	0,00	0,00	0,15	0,85

Panel A presents each audience's center of interest. The redder the cell, the stronger the association between the audience's centre of interest and a feature.

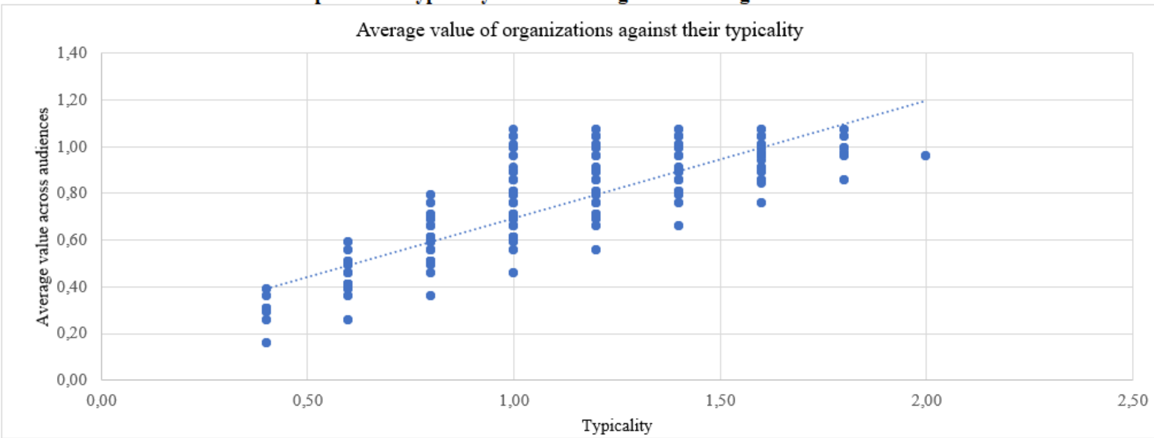
Panel E: Computation of organizations' value in the eye of audiences

An organization's value in the eye of an audience is equal to its alignment with the audience's centre of interest.

The alignment of an organization with an audience's center of interest is computed as:

$$2 - \sum_{f \in \left\{\begin{matrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{matrix}\right\}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Audience's\ center\ of\ interest's\ association\ with\ f) \end{matrix}$$

Panel F: Relationship between typicality and the average value of organizations across all audiences



The average value of an organization is the average of its value in the eyes of each audiences taken over all six audiences (Panel E).
The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2 (Panel C).
The trend line is produced through an ordinary-least square regression.

Figure 2b. Illustration of the case where audiences all behave as exemplar-based evaluators (Proposition 2b)

Panel A: Categories' association with features

	Feature 1	Feature 2	Feature 3	Feature 4
Category 1	0,8	0,2	0	0
Category 2	0	0	0,2	0,8

The redder the cell, the stronger the association of the category with the feature. The prototype of a category is represented by its association with features. For example, the prototypical member of category 1 has an association of 0,8 with feature 1 and of 0,2 with feature 2.

Panel B: Presentation of the organizations used in this illustration

Organizations are associated with features. The association of an organization with a feature is a digit with at most one decimal ranging between 0 and 1. The sum of each organization's associations with features is 1. We consider all possible patterns of organizations' associations with features given these constraints, which represents a total of 286 organizations.

The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2. Thus, the more an organization is similar to a pre-existing prototype, the higher its typicality. The similarity of an organization to a category's prototype is computed as:

$$2 - \sum_{f \in \begin{Bmatrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{Bmatrix}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Category's\ association\ with\ f) \end{matrix}$$

Panel C: Examples of organizations along with their association with features and their level of typicality

	Feature 1	Feature 2	Feature 3	Feature 4	Typicality
Organization 1	0,00	0,00	0,00	1,00	1,60
Organization 3	0,00	0,00	0,20	0,80	2,00
Organization 34	0,00	0,30	0,30	0,40	1,20
Organization 142	0,20	0,20	0,30	0,30	1,00
Organization 11	0,00	0,00	1,00	0,00	0,40
Organization 66	0,00	1,00	0,00	0,00	0,40
Organization 191	0,30	0,30	0,20	0,20	1,00
Organization 224	0,40	0,30	0,30	0,00	1,20
Organization 282	0,80	0,20	0,00	0,00	2,00
Organization 286	1,00	0,00	0,00	0,00	1,60

The redder the cell, the stronger the association between the organization and a feature and the higher the organization's typicality.

Panel D: Audiences' centres of interest

	Feature 1	Feature 2	Feature 3	Feature 4
Audience 1	0,10	0,90	0,00	0,00
Audience 2	0,10	0,80	0,10	0,00
Audience 3	0,15	0,80	0,00	0,00
Audience 4	0,00	0,00	0,90	0,10
Audience 5	0,00	0,10	0,80	0,10
Audience 6	0,00	0,00	0,85	0,15

Panel A presents each audience's center of interest. The redder the cell, the stronger the association between the audience's center of interest and a feature. For Audiences 1 to 3 (1 to 6) we reversed the roles of Features 1 and 2 (3 and 4) relative to Figure 2a.

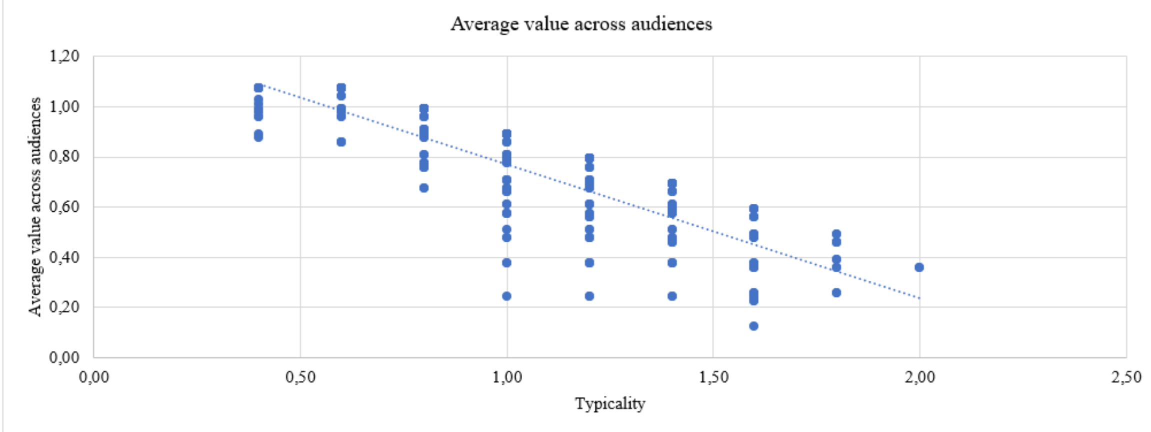
Panel E: Computation of organizations' value in the eye of audiences

An organization's value in the eye of an audience is equal to its alignment with the audience's centre of interest.

The alignment of an organization with an audience's center of interest is computed as:

$$2 - \sum_{f \in \begin{Bmatrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{Bmatrix}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Audience's\ center\ of\ interest's\ association\ with\ f) \end{matrix}$$

Panel F: Relationship between typicality and the average value of organizations across all audiences



The average value of an organization is the average of its value in the eyes of each audiences taken over all six audiences (Panel E).
The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2 (Panel C).
The trend line is produced through an ordinary-least square regression.

Figure 2c. Illustration of a case with both prototype-based and exemplar-based evaluators (Proposition 2c)

Panel A: Categories' association with features

	Feature 1	Feature 2	Feature 3	Feature 4
Category 1	0,8	0,2	0	0
Category 2	0	0	0,2	0,8

The redder the cell, the stronger the association of the category with the feature. The prototype of a category is represented by its association with features. For example, the prototypical member of category 1 has an association of 0,8 with feature 1 and of 0,2 with feature 2.

Panel B: Presentation of the organizations used in this illustration

Organizations are associated with features. The association of an organization with a feature is a digit with at most one decimal ranging between 0 and 1. The sum of each organization's associations with features is 1. We consider all possible patterns of organizations' associations with features given these constraints, which represents a total of 286 organizations.

The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2. Thus, the more an organization is similar to a pre-existing prototype, the higher its typicality. The similarity of an organization to a category's prototype is computed as:

$$2-\sum_{f\in\begin{Bmatrix}Feature\ 1;\\ Feature\ 2;\\ Feature\ 3;\\ Feature\ 4\end{Bmatrix}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f\\ -Category's\ association\ with\ f)\end{matrix}$$

Panel C: Examples of organizations along with their association with features and their level of typicality

	Feature 1	Feature 2	Feature 3	Feature 4	Typicality
Organization 1	0,00	0,00	0,00	1,00	1,60
Organization 3	0,00	0,00	0,20	0,80	2,00
Organization 34	0,00	0,30	0,30	0,40	1,20
Organization 142	0,20	0,20	0,30	0,30	1,00
Organization 11	0,00	0,00	1,00	0,00	0,40
Organization 66	0,00	1,00	0,00	0,00	0,40
Organization 191	0,30	0,30	0,20	0,20	1,00
Organization 224	0,40	0,30	0,30	0,00	1,20
Organization 282	0,80	0,20	0,00	0,00	2,00
Organization 286	1,00	0,00	0,00	0,00	1,60

The redder the cell, the stronger the association between the organization and a feature and the higher the organization's typicality.

Panel D: Audiences' centres of interest

	Feature 1	Feature 2	Feature 3	Feature 4
Audience 1	0,9	0,1	0	0
Audience 2	0,8	0,1	0,1	0
Audience 3	0,15	0,8	0	0
Audience 4	0	0	0,9	0,1
Audience 5	0	0,1	0,1	0,8
Audience 6	0	0	0,15	0,85

Panel A presents each audience's center of interest. The redder the cell, the stronger the association between the audience's center of interest and a feature. For Audiences 3 (4) we reversed the roles of Features 1 and 2 (3 and 4) relative to Figure 2a.

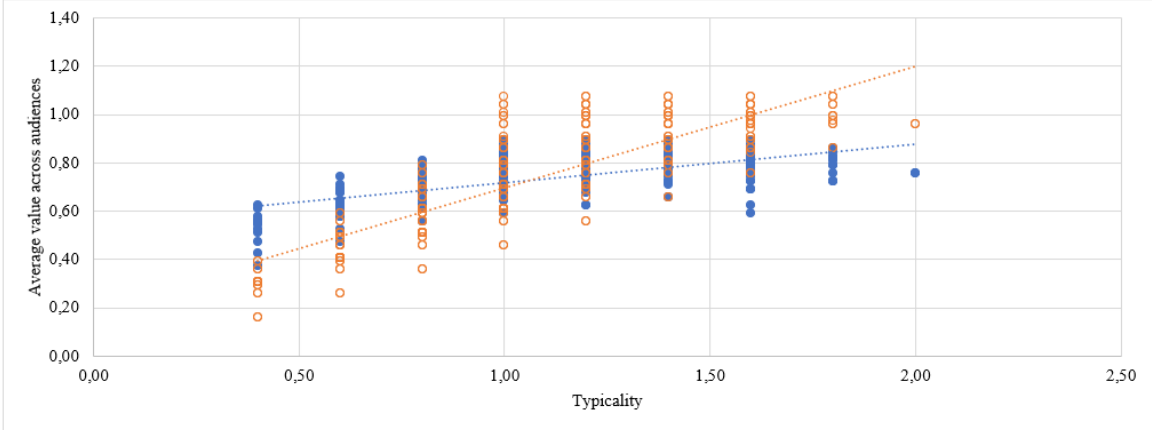
Panel E: Computation of organizations' value in the eye of audiences

An organization's value in the eye of an audience is equal to its alignment with the audience's centre of interest.

The alignment of an organization with an audience's center of interest is computed as:

$$2-\sum_{f\in\begin{Bmatrix}Feature\ 1;\\ Feature\ 2;\\ Feature\ 3;\\ Feature\ 4\end{Bmatrix}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f\\ -Audience's\ center\ of\ interest's\ association\ with\ f)\end{matrix}$$

Panel F: Relationship between typicality and the average value of organizations across all audiences



The average value of an organization is the average of its value in the eyes of each audiences taken over all six audiences (Panel E). The blue dots present organizations' average value given audiences' centres of interest from Panel D. The orange dots present organizations' average value given audiences centres of interest from Panel D in Figure 2a (case with only prototype-based evaluators).

The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2 (Panel C).

The trend line is produced through an ordinary-least square regression.

Figure 3a. Illustration of the case where audiences all behave as goal-based evaluators (Proposition 3a)

Panel A: Categories' association with features

	Feature 1	Feature 2	Feature 3	Feature 4
Category 1	0,8	0,2	0	0
Category 2	0	0	0,2	0,8

The redder the cell, the stronger the association of the category with the feature. The prototype of a category is represented by its association with features. For example, the prototypical member of category 1 has an association of 0,8 with feature 1 and of 0,2 with feature 2.

Panel B: Presentation of the organizations used in this illustration

Organizations are associated with features. The association of an organization with a feature is a digit with at most one decimal ranging between 0 and 1. The sum of each organization's associations with features is 1. We consider all possible patterns of organizations' associations with features given these constraints, which represents a total of 286 organizations.

The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2. Thus, the more an organization is similar to a pre-existing prototype, the higher its typicality. The similarity of an organization to a category's prototype is computed as:

$$2 - \sum_{f \in \left\{\begin{matrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{matrix}\right\}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Category's\ association\ with\ f) \end{matrix}$$

Panel C: Examples of organizations along with their association with features and their level of typicality

	Feature 1	Feature 2	Feature 3	Feature 4	Typicality
Organization 1	0,00	0,00	0,00	1,00	1,60
Organization 3	0,00	0,00	0,20	0,80	2,00
Organization 34	0,00	0,30	0,30	0,40	1,20
Organization 142	0,20	0,20	0,30	0,30	1,00
Organization 11	0,00	0,00	1,00	0,00	0,40
Organization 66	0,00	1,00	0,00	0,00	0,40
Organization 191	0,30	0,30	0,20	0,20	1,00
Organization 224	0,40	0,30	0,30	0,00	1,20
Organization 282	0,80	0,20	0,00	0,00	2,00
Organization 286	1,00	0,00	0,00	0,00	1,60

The redder the cell, the stronger the association between the organization and a feature and the higher the organization's typicality.

Panel D: Audiences' centres of interest

	Feature 1	Feature 2	Feature 3	Feature 4
Audience 1	0,25	0,25	0,25	0,25
Audience 2	0,3	0,25	0,25	0,2
Audience 3	0,2	0,2	0,3	0,3
Audience 4	0,2	0,25	0,3	0,25
Audience 5	0,3	0,25	0,2	0,25
Audience 6	0,2	0,2	0,3	0,25

Panel A presents each audience's center of interest. The redder the cell, the stronger the association between the audience's center of interest and a feature.

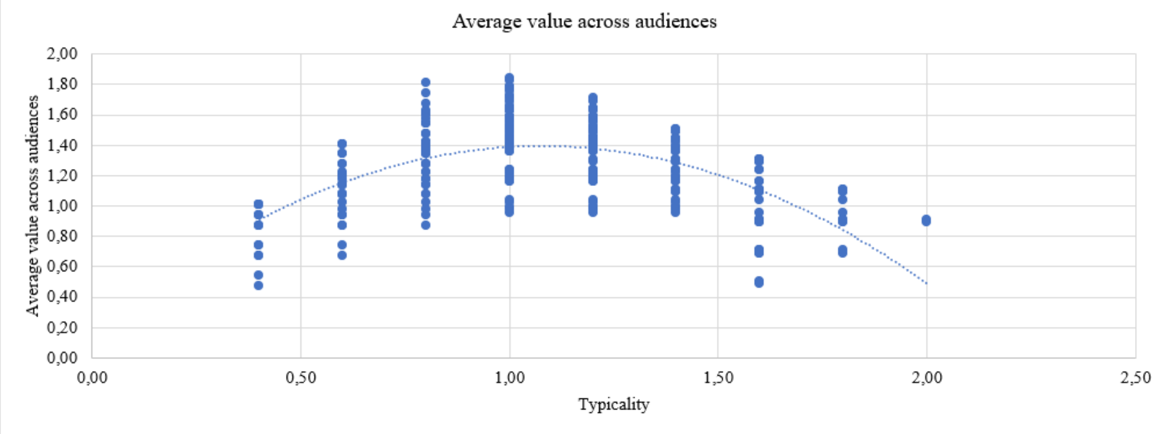
Panel E: Computation of organizations' value in the eye of audiences

An organization's value in the eye of an audience is equal to its alignment with the audience's centre of interest.

The alignment of an organization with an audience's center of interest is computed as:

$$2 - \sum_{f \in \left\{\begin{matrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{matrix}\right\}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Audience's\ center\ of\ interest's\ association\ with\ f) \end{matrix}$$

Panel F: Relationship between typicality and the average value of organizations across all audiences



The average value of an organization is the average of its value in the eyes of each audiences taken over all six audiences (Panel E).
The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2 (Panel C).
The trend line is produced through an ordinary-least square regression.

Figure 3b. Illustration of the case with both prototype-based and goal-based evaluators (Proposition 3b)

Panel A: Categories' association with features

	Feature 1	Feature 2	Feature 3	Feature 4
Category 1	0,8	0,2	0	0
Category 2	0	0	0,2	0,8

The redder the cell, the stronger the association of the category with the feature. The prototype of a category is represented by its association with features. For example, the prototypical member of category 1 has an association of 0,8 with feature 1 and of 0,2 with feature 2.

Panel B: Presentation of the organizations used in this illustration

Organizations are associated with features. The association of an organization with a feature is a digit with at most one decimal ranging between 0 and 1. The sum of each organization's associations with features is 1. We consider all possible patterns of organizations' associations with features given these constraints, which represents a total of 286 organizations.

The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2. Thus, the more an organization is similar to a pre-existing prototype, the higher its typicality. The similarity of an organization to a category's prototype is computed as:

$$2 - \sum_{f \in \begin{Bmatrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{Bmatrix}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Category's\ association\ with\ f) \end{matrix}$$

Panel C: Examples of organizations along with their association with features and their level of typicality

	Feature 1	Feature 2	Feature 3	Feature 4	Typicality
Organization 1	0,00	0,00	0,00	1,00	1,60
Organization 3	0,00	0,00	0,20	0,80	2,00
Organization 34	0,00	0,30	0,30	0,40	1,20
Organization 142	0,20	0,20	0,30	0,30	1,00
Organization 11	0,00	0,00	1,00	0,00	0,40
Organization 66	0,00	1,00	0,00	0,00	0,40
Organization 191	0,30	0,30	0,20	0,20	1,00
Organization 224	0,40	0,30	0,30	0,00	1,20
Organization 282	0,80	0,20	0,00	0,00	2,00
Organization 286	1,00	0,00	0,00	0,00	1,60

The redder the cell, the stronger the association between the organization and a feature and the higher the organization's typicality.

Panel D: Audiences' centres of interest

	Feature 1	Feature 2	Feature 3	Feature 4
Audience 1	0,90	0,10	0,00	0,00
Audience 2	0,80	0,10	0,10	0,00
Audience 3	0,2	0,2	0,3	0,3
Audience 4	0,2	0,25	0,3	0,25
Audience 5	0,00	0,10	0,10	0,80
Audience 6	0,00	0,00	0,15	0,85

Panel A presents each audience's center of interest. The redder the cell, the stronger the association between the audience's center of interest and a feature. Audiences 3 and 4 have the same centres of interest than in Figure 3a while other audiences have the same centres of interest than in Figure 2a

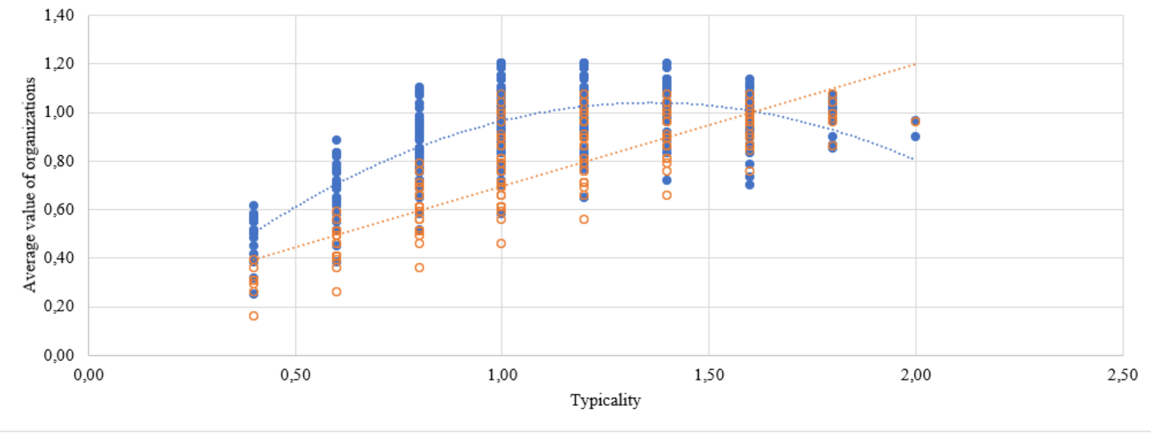
Panel E: Computation of organizations' value in the eye of audiences

An organization's value in the eye of an audience is equal to its alignment with the audience's centre of interest.

The alignment of an organization with an audience's center of interest is computed as:

$$2 - \sum_{f \in \begin{Bmatrix} Feature\ 1; \\ Feature\ 2; \\ Feature\ 3; \\ Feature\ 4 \end{Bmatrix}} \begin{matrix} ABSOLUTE\ VALUE(Organization's\ association\ with\ f \\ -Audience's\ center\ of\ interest's\ association\ with\ f) \end{matrix}$$

Panel F: Relationship between typicality and the average value of organizations across all audiences



The average value of an organization is the average of its value in the eyes of each audiences taken over all six audiences (Panel E). The blue dots present organizations' average value given audiences' centres of interest from Panel D. The orange dots present organizations' average value given audiences centres of interest from Panel D in Figure 2a (case with only prototype-based evaluators).

The typicality of an organization is the maximum of its similarity to the prototypes of Category 1 and 2 (Panel C).

The trend line is produced through an ordinary-least square regression.

CONCLUSION

Seminal papers on market categories showed their stabilizing roles as they help audiences converge on their valuations of market offerings (Negro, Koçak, & Hsu, 2010; Zuckerman, 1999). However, recent research emphasizes how market categories may contribute to create variability rather than stability in audiences' valuations. Within the framework of prototype-based categories, different audiences can value typical entities differently (Goldberg, Hannan, & Kovacs, 2016; Pontikes, 2012; Smith, 2011). The meaning of existing categories may also change or new categories form, in turn modifying the values of their members (Delmestri & Greenwood, 2016; Kennedy, Lo, & Lounsbury, 2010). Finally, audiences may use different models of valuation, which do not rely on prototypes but on ideals or salient exemplars, which can in turn create variability in their valuations (Durand & Thornton, 2018; Zhao, Ishihara, Jennings, & Lounsbury, 2018).

This dissertation sought to embrace the inherent variability of audiences' valuations and asked: why are audiences' valuations so variable? The first chapter of this dissertation shows that the greater stability of the value of typical entities is contingent on categorical ambiguity. Hence, even typical entities may experience variability in their valuations if they belong to ambiguous categories. The second chapter of this dissertation studies the impact of organizations' attractiveness on audiences' valuations alongside that of typicality and shows that this impact is substantial, at least in the IPO setting, and depends on audiences' sentiment at a given point in time. Hence, although audiences may rely on pre-existing and stable categories to structure their valuations, they are also influenced by temporary attractions toward certain features associated with success which thus induce temporary variations in audiences' valuations. Finally, the third chapter of this dissertation developed a theory integrating the different models of valuation used by audiences. It notably proposes that the relationship between typicality and valuation varies as a function of audiences' uses of different models of valuation.

Overall, this dissertation provides three related answers to the question "Why are audiences' valuations so variable?". First, it suggests that audiences' valuations can remain variable despite the stabilizing role of market categories if there exist ambiguous categories among them (Essay 1). Second, it suggests that even in the presence of relatively stable market categories, temporary attractions toward certain features continue to influence audiences' valuations, creating variability over time (Essay 2). Third, it proposes that the co-existence of prototype-based, exemplar-based and goal-based evaluators with potentially incongruent centers of interest is an additional source of variability in audiences' valuations both from one audience to another and over time (Essay 3).

1. Contributions

The contributions of this dissertation are threefold. First, it enriches different theoretical perspectives on market categories and audiences' valuations. Second, it proposes new ways of locating firms in semantic spaces and new perspectives on audiences' valuations inspired by these methods and thus contributes to computational approaches to culture and organizations, a novel and quickly expanding field which studies culture and its impact on organizations using NLP and machine learning (see, e.g., Corritore, Goldberg, & Srivastava, 2019; Hannigan et al., 2019; Srivastava, Goldberg, Manian, & Potts, 2017). Third, it has various theoretical and practical implications of central interest to practitioners.

1.1. Contributions to the study of market categories and audiences' valuations

This dissertation contributes to the study of prototype-based market categories as it explores their impact on the variability of audiences' valuations. It also contributes to studies of goal-based and exemplar-based categories and tries to integrate these three different perspectives in a coherent framework. Finally, it contributes to the literature on optimal distinctiveness (Zhao, Fisher, Lounsbury, & Miller, 2017; Zuckerman, 2016), which emphasizes organizations' needs to both 'blend in' -i.e. be typical- and stand out.

1.1.1. Contribution to the literature on prototype-based categories

This dissertation makes several contributions to the literature on prototype-based market categories. First, it shows that audiences use several bases to define organizations and their products beyond pre-existing and well-established categories. Audiences' definitions of the entities they encounter are as much influenced by pre-existing categories as by temporary attractions toward certain features (Essay 2), by audiences' ideals or by salient exemplars (Essay 3). Extant literature presents audiences' valuations as primarily influenced by their understanding of whether organizations are typical instances of stable categories -i.e. of what they are *in general*. The second essay of this dissertation provides evidence that audience' valuations are also influenced by audiences' understanding of whether organizations have attractive features at a given point in time -i.e. of whether they are 'hot' *right now*. Hence, audiences' valuations rest both on a judgment regarding the 'intemporal' essence of an organization (Bitektine, 2011; Hannan et al., 2019) and a judgment regarding the temporal correspondence between the organization's features and recent trends. Essay 3 discusses how audiences may use ideals -i.e. combinations of features identified as good tools to achieve their goals or relevant solutions to their needs- or salient exemplars rather than prototypes to structure their valuations.

Second, the first essay of this dissertation contributes to the expanding literature on prototype-based categories studying ambiguity and related constructs such as contrast (Kovács & Hannan, 2010), leniency (Pontikes, 2012) or coherence (Lo, Fiss, Rhee, & Kennedy, 2019). This literature notably identifies ambiguity as influencing organizations' decisions to enter or leave a market category (Montauti & Wezel, 2016; Pontikes & Barnett,

2015) and as attenuating the positive effect of typicality on the valuation of certain products (Kovács & Hannan, 2010). The first essay of this thesis expands this literature by uncovering the role of categorical ambiguity in shaping the relationship between typicality and the variability of audiences' valuations. In unambiguous categories, the information encoded in prototypes is highly relevant to value typical entities. All audiences thus share a common and reliable source of information to value typical entities and are thus more likely to converge on comparable assessments of their values (Hsu, Roberts, & Swaminathan, 2012; Zuckerman, 2004). However, in ambiguous categories, the information encoded is prototype is less relevant to the valuation of typical entities. It is thus more likely that poorly informed audiences will overestimate or underestimate their value, producing variability in audiences' valuations.

1.1.2. Contributions to the literature on goal-based and exemplar-based categories

This dissertation further contributes to recent research streams dedicated to the study of goal-based and exemplar-based categories (Durand & Thornton, 2018; Zhao et al., 2018). First, this dissertation contributes more specifically to the burgeoning literature on goal-based categories (Durand & Boulongne, 2017; Durand & Paolella, 2013; Glaser, Krikorian Atkinson, & Fiss, 2019). Unlike prototype-based categories, which are shared among audiences, audiences idiosyncratically derive goal-based categories as a function of their current goals (Barsalou, 1991). Research on goal-based categories mostly showed how atypical entities can be devalued if audiences' goals lead them to purposively look for atypical entities (Paolella & Durand, 2016; Paolella & Sharkey, 2017). Essay 2 of this dissertation expands this strand of research and shows that audiences' dominant sentiment at a given point in time, and thus notably their propensity to purposively seek attractive entities or to shun typical ones, is an important determinant of the relationship between typicality and valuation on the one hand and attractiveness and valuation on the other hand. Essay 3 of this dissertation further contributes to the study of the influence of goal-based categorization on category formation (Granqvist & Ritvala, 2016). Essay 3 argues that a high proportion of goal-based evaluators is generally detrimental to the emergence of a shared category system. Indeed, when all audiences are goal-based evaluators, organizations have a harder time converging on a narrow set of features which audiences would value more positively and that could serve as a basis to construct category prototypes.

This dissertation also relates to exemplar-based categorization and its impact on audiences' valuations of organizations (Barlow, Verhaal, & Angus, 2019; Zhao et al., 2018). In essay 2, attractiveness is thought of in terms of similarity to IPOs which have known high-level of first-day returns. It is thus a measure of an issuing firm's similarity to salient exemplars. As such, essay 2 shows that exemplar-based categorization plays an important role in the valorisation of IPOs, above and beyond that played by prototype-based categorization. Essay 3 develops the model of category formation based on exemplars proposed by Zhao and colleagues and shows that a high proportion of exemplar-based evaluators among audiences favour the formation of new categories. Indeed, organizations have an easier time converging on a narrow set of features that can serve as a basis to construct category prototypes when they can rely on salient exemplars to gauge audiences' centers of interest (Zhao et al., 2018).

1.1.3. Contributions to the literature on optimal distinctiveness

Beyond the category literature specifically, this dissertation contributes to the quickly expanding research on optimal distinctiveness which emphasizes that organizations must find the optimal balance between 'blending in' and 'standing out' relative to their peers (Zhao et al., 2017; Zuckerman, 2016). Blending in provides cognitive legitimacy in the eyes of audiences by clarifying one's membership into an established category (Zuckerman, 1999) but it also draws organizations closer to the crowd of their peers (Navis & Glynn, 2011). Organizations that are too typical of a category thus lack distinctive features that would attract audiences' attention (Deephouse, 1999). Standing out by adopting distinctive features prevent organizations from being overlooked. As a result, organizations that have a moderate level of typicality tend to enjoy superior valuations or performance. For example, moderately typical songs are more likely to rise up the Billboard's 100 charts as they both demonstrate that they belong to existing genres and attract attention (Askin & Mauskapf, 2017). Since the need to 'stand out' depends on how crowded the category center is, it is less pronounced or even disappears in heterogeneous categories (Haans, 2019).

Essay 1 relates to the issue of optimal distinctiveness by suggesting that typical firms tend to experience less volatility. Under a perspective that would see the volatility of a firm's stock price as a measure of risk (see for example Bansal & Clelland, 2004; Harrison, Thurgood, Boivie, & Pfarrer, 2019) and thus of the potential returns that may come to investors who invest in it, firms may want to avoid appearing as too risky in the eyes of investors while also offering good prospects for future returns, i.e. to ensure moderate levels of volatility. Essay 1 suggests two ways of achieving this objective. Firms may either adopt a moderate level of typicality in an unambiguous industry category or become typical of an ambiguous one. Essay 2 also relates directly to optimal distinctiveness as issuing firms and underwriters have to balance their respective goals when crafting IPO prospectuses : very high first-day returns means a lot of money has been 'left on the table' for the issuing firm (Loughran & Ritter, 2002) while low or negative returns means institutional investors who have been allocated shares in the IPO by the underwriters did not profit from it (at least in the first day) (Goldstein, Irvine, Kandel, & Wiener, 2009; Goldstein, Irvine, & Puckett, 2011). Essay 2 identifies typicality and attractiveness as two dimensions which issuing firms and underwriters can optimally balance so as to achieve levels of returns that satisfy them both.

Finally, essay 3 presents the relationship between typicality and valuation as resulting from the relative proportion of prototype-based, exemplar-based and goal-based evaluators among audiences. Thus, organizations' need to 'blend in' or 'stand out' to achieve superior value appears as contingent on whether audiences tend to behave as prototype-based, goal-based or exemplar-based evaluators. This is an important contribution to the literature on optimal distinctiveness, which so far has largely approached the problem of the optimal distinctiveness of organizations in terms of their positioning relative to their peers without necessarily accounting for audiences' theory of value (Zuckerman, 2017). Essay 3 suggests that as far as organizations' valuation is concerned, whether organizations' need to blend in, stand out, or find a balance between the two ultimately lies in the eyes of audiences.

1.2. Contributions to computational approaches to organizations

In line with the most recent developments in the studies of categories and organizations (Hannan et al., 2019), this dissertation uses advanced Natural Language Processing techniques to locate organizations in semantic spaces and measure typicality (essays 1 and 2), ambiguity (essay 1) and attractiveness (essay 2). Both essays 1 and 2 leverage large corpora of financial documents to train NLP models to represent the meaning of words and entire documents. The position of a firm in the semantic spaces learnt by the NLP models is then that of the documents it produces. Essay 1 uses a word embedding model (Mikolov, Chen, Corrado, & Dean, 2013) to represent the content of annual reports while essay 2 uses a document embedding model (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) to represent the content of IPO prospectuses and annual reports. These two approaches are highly similar in nature and rely on co-occurrences of words to model semantic relations between words, documents and, ultimately, firms.

The method used in essay 1 has the advantage of accounting for semantic relations between words which are otherwise ignored. The method used in essay 2 directly models the content of entire documents by learning to predict the words that they contain. Compared to already existing approaches modelling the position of firms in semantic spaces using annual reports (Hoberg & Phillips, 2010, 2016), the approaches proposed in both essays 1 and 2 have the advantage of locating firms using semantically related words in the same region of the semantic space, even if they do not use the same words.

The theory developed in essay 3 was also inspired by distributional approaches that prevails in NLP studies of meaning (Lenci, 2018), which originated with linguists such as Firth (Firth, 1957) or Harris (Harris, 1954), and now becomes prevalent in research on market categories (Hannan et al., 2019). In this dissertation, I thought of audiences as locating organizations and their products, as well as their centers of interest, in semantic spaces. I then interpreted entitites' alignment with audiences' centers of interest in terms of distances within this space. I believe that this kind of theorization holds great promises as propositions stemming from it can be straightforwardly tested using the rich methodological tools offered by NLP and machine learning in general.

1.3. Implications for practice

This dissertation has several implications for practice. First, it suggests that organizations face audiences which might value them and their products in possibly incongruent ways. When launching a new product, when describing their activities, organizations' members must have a good understanding of the models of valuation that audience members are most likely to use and/or of the ones that the organization wants to prompt. In some contexts, organizations may want to emphasize how they can help their clients achieve their goals. In others, it might be better to establish one's typicality relative to pre-existing and well-known categories. In still others, it might be important to do both or strike a balance between these two objectives.

I firmly believe that these considerations are a key area of concerns for practitioners. For example, during my PhD thesis, I was involved in a research project with a NGO providing coupons to disadvantaged families with infants to buy good quality baby food. A key question faced by this organization was how to best present themselves to prospective beneficiaries to ensure that they would understand the organization's purpose, apply for coupons and use them: should the NGO emphasize its typicality as a socially-oriented organization or its ability to help families achieve their goal of providing good quality food to their children? During the project, we got involved with a French governmental agency in the north of Paris and realized that they faced similar issues with their audiences. The agency was eager to better understand the cognitive mechanisms driving the perceptions of their key audiences and we are in the process of launching a research project related to this issue. These examples show that studying organizations' typicality, alignment with audiences' ideals or similarity to salient exemplars is not just a research exercise: organization members themselves constantly try to understand how to best position themselves in the eyes of their audiences to succeed in creating a meaningful bound with them.

This dissertation also has more precise implications for practice in financial markets. Text-based analysis is developing quickly in finance (Loughran & Mcdonald, 2016) and numerous start-ups and financial institutions are now using textual analysis to construct innovative financial products. As an illustration of this quickly developing field, a recent paper managed to predict stock prices based on information extracted from 8-K documents, a feat which will surely garner attention from both researchers and practitioners (Lee, Surdeanu, MacCartney, & Jurafsky, 2014). Practitioners and researchers alike mainly use NLP techniques to extract content from financial document, such as their sentiment or the topics they discuss. This dissertation suggests that adopting a relational approach to financial documents rather than simple content extraction can help predict financial variables of interest to practitioners, such as volatility (essay 1) or IPO returns (essay 2). I am currently working on another project which uses a similar relational approach on transcripts of quarterly earnings calls and finds that it predicts quarterly earnings surprise. Preliminary discussions with practitioners with expertise in this domain -communication professionals and bankerssuggest widespread interest in better understanding NLP methods and how they could contribute to help them define their communication or predict key variables of interest. In the future, I hope to be able to promote and develop these kinds of approaches with practitioners.

More broadly, this dissertation contributes to bridge the gap between the ever-expanding use of NLP among practitioners and research on organizations. It shows how one might use advanced NLP techniques which are accessible through open Python libraries to produce measures of firms' relatedness and then predict important outcomes. Data scientists working within organizations, managers, analysts or investors may all be interested in adapting this kind of study to their own needs, and I believe that business schools have a strong interest in internalizing this kind of competencies and teaching them to their students, and I hope to be able to do so shortly.

2. Limitations and future research

2.1. Current work limitations

This dissertation has several limitations. In the two empirical essays, typicality is measured using firms' similarity to the prototypical member of their main industry has defined using SIC code. One potential issue with this approach is that investors do not necessarily use categories which exactly correspond to SIC industries. However, it seems reasonable to assume that the proposed measure of typicality still approximates the typicality of firms as perceived by investors. Even if a given SIC industry does not gather all or exactly the peers that investors would associate to a focal firm, it likely gathers most of them as well as firms with related activities. Thus, taking a firm's similarity to the 'average' member of its SIC industry seems a reasonable proxy for its actual typicality.

This dissertation uncovers relationships between constructs such as typicality, attractiveness and ambiguity and audiences' valuations. To do so, it uses innovative NLP techniques to reveal patterns in large datasets which would otherwise be hard to observe. However, the observed relationships warrant further exploration to more firmly establish causality. Nonetheless, the richness of the results and the measurement of important socio-cognitive constructs using semantics abstracted from texts opens incredible opportunity for future research and I firmly believe that this kind of more exploratory studies are of great importance for the study of organizations and to further social sciences in general.

Finally, this dissertation focuses mainly on audiences' perceptions and valuations but uses documents produced by organizations to measure typicality, ambiguity or attractiveness. It thus assumes that audiences are sensible to the meanings conveyed by organizations, which seems reasonable. Yet, it is important to acknowledge that other meanings may factor in audiences' valuations. However, this does not substantially alter the interpretation of the relationships uncovered in this thesis, which suggest that the typicality, ambiguity and attractiveness of organizations, as measured through the meanings they convey, all impact the variability of audiences' valuations.

2.2. Avenues for future research

Most of this dissertation focuses on the antecedents of audiences' valuations at the audience level. However, organizations play a significant role in shaping the categories that audiences' use and constantly seeks to influence how audiences perceive them. Organizations selectively emphasize or downplay their membership in categories based on their assessment of audiences' perceptions (Granqvist, Grodal, & Woolley, 2013). In some contexts, they purposively favour ambiguous categories as they offer them more flexibility and help them avoid scrutiny (Pontikes & Kim, 2017). Competing organizations may nonetheless use common frames or selectively name each other in their press releases to ensure that they are well-positioned in emerging categories (Kennedy, 2008; Navis & Glynn, 2010). Future research may further explore how organizations influence audiences' assessments of their typicality or of their similarity to ideals or exemplars through their public interventions -such as during quarterly earnings calls, annual meetings, conferences, etc.- and how this in turn relates to audiences' valuations. It may also explore how organizations may purposively prompt models of valuation which are more favourable to them, and, in so doing, impact audiences' valuations.

Specific audiences, such as underwriters (essay 2), financial analysts or critics can have an impact on the perceptions of other audiences. These intermediaries are generally recognized as playing a key role in shaping categories. For example, analysts' coverage of publicly listed firms is an important determinant of investors' attention (Zuckerman, 1999). Future research might further refine our understanding of how organizations may impact the categorization processes and valuations of intermediaries -and in turn those of their broader audiences- through their direct interaction with them (e.g. when managers meet financial analysts).

This dissertation adopts an approach to categories which sees organizations as located in a semantic space. Their position in this space is inferred from the words that they use in financial documents. Such an approach puts language and communication at the center of audiences' valuations. As such, the results and arguments developed in this dissertation naturally relates to different linguistic approaches to organizations and most notably discursive approaches, vocabularies and rhetorical perspectives. Under a discursive lens, one places a heavier emphasis on texts and how broad discourses define and shape organizational and individual actions (Hardy & Maguire, 2010; Phillips, Lawrence, & Hardy, 2004). One way to articulate this lens with the approach to categories and organizations developed in this dissertation would be to study simultaneously firms' positions relative to category prototypes *and* the evolving and changing meaning of these categories as they are constructed in the broader discourse. One could for example study how a firm's typicality results both from its attempts at getting closer to the category center *and* from the propensity of the category's meaning to converge toward the position occupied by the firm in the semantic space.

Under a vocabularies lens, the emphasis is placed on three kinds of semantic relationships: category-to-category relationships (as in "a bank is a financial institution"), category-to-example relationships (as in "HSBC is a bank") and example-to-example relationships (as in "Revolut is the Uber of banking") (Loewenstein, Ocasio, & Jones, 2012; Mills, 1940; Ocasio, Loewenstein, & Nigam, 2015). The study of vocabulary thus offers interesting opportunities to explore how prototype-based valuation (structured by categories) articulates with exemplar-based valuation (structured by exemplars). For example, one way of exploring the implications of essay 3 could be to study whether firms tend to mention category-to-category relationships in their press releases (i.e. to prompt prototype-based valuation), example-to-example relationships (i.e. to prompt exemplar-based valuation) or category-to-example relationships (i.e. to articulate both models of valuation) and the impact this has on investors' valuations.

Finally, under a rhetorical lens, the emphasis lays on the nature of the arguments used by organizations and audiences (Green, Li, & Nohria, 2009; Harmon, Green, & Goodnight, 2015). This dissertation approaches typicality in terms of 'raw' semantic meanings conveyed by firms, as measured through their propensity to use certain words rather than others. However, typicality could also be approached in terms of organizations' claims that they belong (or not) to a given category. Such a perspective, focused on exploring organizations' arguments as they relate to pre-existing categories, could complement some of the insights developed in this thesis. For instance, one could look at claims made by firms and how they influence their perceived typicality or attractiveness in the eyes of audiences. Recent developments in rhetorical approaches to organizations suggest that making claims which states taken-for-granted assumptions has the adverse effect of suggesting that these assumptions are no longer to be taken-for-granted (Harmon, 2019). This could mean that well-established organizations making explicit claims of typicality would paradoxically reduce their typicality in the eyes of audiences. Similarly, new or ‘hot’ organizations making claims of attractiveness could create doubts among audiences regarding their ‘hotness’.

Final words

This thesis asked “Why are audiences’ valuations of organizations so variable?”, adopting the lens of market categories, and sought to answer it using natural language processing techniques. I would like to conclude by emphasizing that the choice of these methods is not incidental. NLP techniques offer an incredible opportunities to reunite distributional approaches to words’ meaning and the study of market categories, which were both inspired by the view that meaning is use (Wittgenstein, 1953). They allow scholars to study how organizations’ use words and thus how they shape, willingly or unwillingly, what they mean – both in each text that they produce and in general, in the eyes of audiences. I further believe that these methods are theoretical tools, offering a unique view on meaning as being relational and distributional, which challenges our day-to-day, naive assumptions about language, such as its being referential or its being structured by well-defined categories of words. Although these ideas are not new per se in the study of organizations (Cornelissen, Durand, Fiss, Lammers, & Vaara, 2015), the quickly expanding use of NLP in organization studies -as exemplified by the success of topic modelling (Hannigan et al., 2019)- will probably lead to their diffusion at a much wider scale and thus to a greater awareness among scholars of their theoretical implications. Notably, beyond their correspondence with extant theorizing on market categories, distributional approaches to meaning have important consequences for the study of audiences’ valuations as it relates to organizational wrong-doings or purposeful actions (Hollensbe, Wookey, Hickey, George, & Nichols, 2014). Indeed, if words are not referential and their meaning derives from how they are used, then what does it mean that an audience finds that an organization is ‘good’ or ‘bad’, or ‘doing the right thing’? Does it even have a sense to ask such questions? These considerations open a fascinating area of future research for the study of audiences’ valuations of organizations.

REFERENCES

Askin, N., & Mauskapf, M. 2017. What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. ***American Sociological Review***, 82(5): 910–944.

Bansal, P., & Clelland, I. 2004. Talking trash: Legitimacy, impression management, and unsystematic risk in the context of natural environment. ***Academy of Management Journal***, (1): 93.

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

Barsalou, L. W. 1991. Deriving Categories to Achieve Goals. In G. H. Bower (Ed.), ***Psychology of Learning and Motivation***, vol. 27: 1–64. Academic Press.

Bitektine, A. 2011. Toward a Theory of Social Judgments of Organizations: The Case of Legitimacy, Reputation, and Status. ***Academy of Management Review***, 36(1): 151–179.

Cornelissen, J. P., Durand, R., Fiss, P. C., Lammers, J. C., & Vaara, E. 2015. Putting Communication Front and Center in Institutional Theory and Analysis. ***Academy of Management Review***, 40(1): 10–27.

Corritore, M., Goldberg, A., & Srivastava, S. B. 2019. Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance. ***Administrative Science Quarterly***, 0001839219844175.

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

Delmestri, G., & Greenwood, R. 2016. How Cinderella Became a Queen: Theorizing Radical Status Change. ***Administrative Science Quarterly***, 61(4): 507–550.

Durand, R., & Boulongne, R. 2017. Advancing category research: Theoretical mapping and under-researched areas. ***The Sage Handbook of Organizational Institutionalism***. London.

Durand, R., & Paolella, L. 2013. Category Stretching: Reorienting Research on Categories in Strategy, Entrepreneurship, and Organization Theory. ***Journal of Management Studies***, 50(6): 1100–1123.

Durand, R., & Thornton, P. H. 2018. Categorizing Institutional Logics, Institutionalizing Categories: A Review of Two Literatures. ***Academy of Management Annals***, 12(2): 631–658.

Firth, J. R. 1957. A synopsis of linguistic theory 1930-1955. ***Selected Papers of J. R. Firth***.

Glaser, V. L., Krikorian Atkinson, M., & Fiss, P. C. 2019. Goal-Based Categorization: Dynamic Classification in the Display Advertising Industry. ***Organization Studies***, 017084061988336.

Goldberg, A., Hannan, M. T., & Kovacs, B. 2016. What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. ***American Sociological Review***, (2): 215.

Goldstein, M. A., Irvine, P., Kandel, E., & Wiener, Z. 2009. Brokerage Commissions and Institutional Trading Patterns. ***Review of Financial Studies***, 22(12): 5175–5212.

Goldstein, M. A., Irvine, P., & Puckett, A. 2011. Purchasing IPOs with Commissions. ***Journal of Financial and Quantitative Analysis***, 46(05): 1193–1225.

Granqvist, N., Grodal, S., & Woolley, J. L. 2013. Hedging Your Bets: Explaining Executives' Market Labeling Strategies in Nanotechnology. ***Organization Science***, 24(2): 395–413.

Granqvist, N., & Ritvala, T. 2016. Beyond Prototypes: Drivers of Market Categorization in Functional Foods and Nanotechnology. ***Journal of Management Studies***, 53(2): 210–237.

Green, S. E., Li, Y., & Nohria, N. 2009. Suspended in self-spun webs of significance: A rhetorical model of institutionalization and institutionally embedded agency. ***Academy of Management Journal***, 52(1): 11–36.

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

Hannan, M. T., Mens, G. L., Hsu, G., Kovács, B., Negro, G., et al. 2019. ***Concepts and Categories: Foundations for Sociological and Cultural Analysis*** (1 edition). New York: Columbia University Press.
Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., et al. 2019. Topic Modeling in Management Research: Rendering New Theory from Textual Data. ***Academy of Management Annals***, 13(2): 586–632.
Hardy, C., & Maguire, S. 2010. Discourse, field-configuring events, and change in organizations and institutional fields: Narratives of DDT and the Stockholm Convention. ***Academy of Management Journal***, 53(6): 1365–1392.
Harmon, D. J. 2019. When the Fed Speaks: Arguments, Emotions, and the Microfoundations of Institutions. ***Administrative Science Quarterly***, 64(3): 542–575.
Harmon, D. J., Green, Jr., Sandy E., & Goodnight, G. T. 2015. A Model of Rhetorical Legitimation: The Structure of Communication and Cognition Underlying Institutional Maintenance and Change. ***Academy of Management Review***, 40(1): 76–95.
Harris, Z. S. 1954. Distributional Structure. ***WORD***, 10(2–3): 146–162.
Harrison, J., Thurgood, G. R., Boivie, S., & Pfarrer, M. 2019. Perception Is Reality: How CEOs' Observed Personality Influences Market Perceptions of Firm Risk and Shareholder Returns. ***Academy of Management Journal***. https://doi.org/10.5465/amj.2018.0626.
Hoberg, G., & Phillips, G. 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. ***Review of Financial Studies***, 23(10): 3773–3811.
Hoberg, G., & Phillips, G. 2016. Text-Based Network Industries and Endogenous Product Differentiation. ***Journal of Political Economy***, 124(5): 1423–1465.
Hollensbe, E., Wookey, C., Hickey, L., George, G., & Nichols, C. V. 2014. Organizations with Purpose. ***Academy of Management Journal***, 57(5): 1227–1234.
Hsu, G., Roberts, P. W., & Swaminathan, A. 2012. Evaluative Schemas and the Mediating Role of Critics. ***Organization Science***, 23(1): 83–97.
Kennedy, M. T. 2008. Getting Counted: Markets, Media, and Reality. ***American Sociological Review***, 73(2): 270–295.
Kennedy, M. T., Lo, J. Y.-C., & Lounsbury, M. 2010. Category currency: The changing value of conformity as a function of ongoing meaning construction. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 369–397. Emerald Group Publishing Limited.
Kovács, B., & Hannan, M. T. 2010. The consequences of category spanning depend on contrast. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 175–201. Emerald Group Publishing Limited.
Lee, H., Surdeanu, M., MacCartney, B., & Jurafsky, D. 2014. On the importance of text analysis for stock price prediction. ***Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014***, 1170–1175. Presented at the 9th International Conference on Language Resources and Evaluation, LREC 2014, European Language Resources Association (ELRA).
Lenci, A. 2018. Distributional Models of Word Meaning. ***Annual Review of Linguistics***, 4(1): 151–171.
Lo, J. Y.-C., Fiss, P. C., Rhee, E. Y., & Kennedy, M. T. 2019. Category Viability: Balanced Levels of Coherence and Distinctiveness. ***Academy of Management Review***. https://doi.org/10.5465/amr.2017.0011.
Loewenstein, J., Ocasio, W., & Jones, C. 2012. Vocabularies and Vocabulary Structure: A New Approach Linking Categories, Practices, and Institutions. ***The Academy of Management Annals***, 6(1): 41–86.
Loughran, T., & Mcdonald, B. 2016. Textual Analysis in Accounting and Finance: A Survey. ***Journal of Accounting Research***, 54(4): 1187–1230.
Loughran, T., & Ritter, J. R. 2002. Why Don't Issuers Get Upset about Leaving Money on the Table in IPOs? ***The Review of Financial Studies***, 15(2): 413–443.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. 2013. Efficient Estimation of Word Representations in Vector Space. ***ArXiv:1301.3781 [Cs]***. http://arxiv.org/abs/1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), ***Advances in Neural Information Processing Systems 26***: 3111–3119. Curran Associates, Inc.

Mills, C. W. 1940. Situated Actions and Vocabularies of Motive. ***American Sociological Review***, 5(6): 904.

Montauti, M., & Wezel, F. C. 2016. Charting the Territory: Recombination as a Source of Uncertainty for Potential Entrants. ***Organization Science***, 27(4): 954–971.

Navis, C., & Glynn, M. A. 2010. How New Market Categories Emerge: Temporal Dynamics of Legitimacy, Identity, and Entrepreneurship in Satellite Radio, 1990–2005. ***Administrative Science Quarterly***, 55(3): 439–471.

Navis, C., & Glynn, M. A. 2011. Legitimate distinctiveness and the entrepreneurial identity: Influence on investor judgments of new venture plausibility. ***Academy of Management Review***, 36(3): 479–499.

Negro, G., Koçak, Ö., & Hsu, G. 2010. Research on categories in the sociology of organizations. In G. Hsu, G. Negro, & Ö. Koçak (Eds.), ***Research in the Sociology of Organizations***, vol. 31: 3–35. Emerald Group Publishing Limited.

Ocasio, W., Loewenstein, J., & Nigam, A. 2015. How Streams of Communication Reproduce and Change Institutional Logics: The Role of Categories. ***Academy of Management Review***, 40(1): 28–48.

Paolella, L., & Durand, R. 2016. Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market. ***Academy of Management Journal***, 59(1): 330–351.

Paolella, L., & Sharkey, A. 2017. Forging Consensus: An Integrated View of How Categories Shape the Perception of Organizational Identity. In R. Durand, N. Granqvist, & A. Tyllström (Eds.), ***Research in the Sociology of Organizations***, vol. 51: 327–353. Emerald Publishing Limited.

Phillips, N., Lawrence, T. B., & Hardy, C. 2004. Discourse and Institutions. ***The Academy of Management Review***, 29(4): 635–652.

Pontikes, E. G. 2012. Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences' Evaluations. ***Administrative Science Quarterly***, 57(1): 81–118.

Pontikes, E. G., & Barnett, W. P. 2015. The persistence of lenient market categories. ***Organization Science***, (5): 1415.

Pontikes, E. G., & Kim, R. 2017. Strategic Categorization. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 71–111. Emerald Publishing Limited.

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

Srivastava, S. B., Goldberg, A., Manian, V. G., & Potts, C. 2017. Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations. ***Management Science***. https://doi.org/10.1287/mnsc.2016.2671.

Wittgenstein, L. 1953. ***Philosophical Investigations***. Paris: Blackwell Publishers.

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

Zhao, E. Y., Ishihara, M., Jennings, P. D., & Lounsbury, M. 2018. Optimal Distinctiveness in the Console Video Game Industry: An Exemplar-Based Model of Proto-Category Evolution. ***Organization Science***, 29(4): 588–611.

Zuckerman, E. W. 1999. The Categorical Imperative: Securities Analysts and the Illegitimacy Discount. ***American Journal of Sociology***, 104(5): 1398–1438.

Zuckerman, E. W. 2004. Structural Incoherence and Stock Market Activity. ***American Sociological Review***, 69(3): 405–432.

Zuckerman, E. W. 2016. Optimal Distinctiveness Revisited. ***The Oxford Handbook of Organizational Identity***. https://doi.org/10.1093/oxfordhb/9780199689576.013.22.

Zuckerman, E. W. 2017. The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool. ***From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads***, vol. 51: 31–68. Emerald Publishing Limited.

Titre : Expliquer la Variabilité des Evaluations des Audiences : une Approche Basée sur les Catégories de Marché et le Traitement du Langage Automatisé

Mots clés : Catégorisation, Evaluation, Traitement automatique du langage, Entreprises cotées en bourse

Résumé : Cette thèse cherche à établir si les différents processus de catégorisation influençant les évaluations des audiences sur les marchés conduisent à une stabilisation ou à une plus grande variabilité de leurs évaluations. Bien que les travaux de recherche fondateurs portant sur la catégorisation aient insisté sur le rôle stabilisateur des catégories sur les marchés, la recherche récente suggère que les évaluations des audiences peuvent varier substantiellement, même sur des marchés dotés de catégories pré-existantes bien établies. Cette variabilité résulte notamment des préférences hétérogènes des audiences pour les offres typiques, de changements dans les significations associées aux catégories ou de l'utilisation par les audiences de plusieurs modes d'évaluation. En se basant sur ces nouveaux résultats, cette thèse cherche pourquoi les évaluations des audiences sont si variables et explore en détail le rôle joué par les catégories de marché dans cette variabilité. Cette thèse propose que i) les catégories ambigües, ii) l'influence d'attractions temporaires parmis les audiences aux côtés des catégories plus stables et iii) la co-existence de plusieurs types d'évaluateurs contribuent à produire de la variabilité dans les évaluations des audiences. Les deux premiers essais empiriques utilisent des données sur des entreprises cotées en bourse aux Etats-Unis. Dans ces essais, la similarité des entreprises aux prototypes des catégories existantes ou l'attraction temporaire des audiences vers certains attributs sont mesurés à l'aide de contenus sémantiques extraits d'un large corpus de rapports annuels et de prospectus d'entrée en bourse. Le troisième essai est un modèle théorique. Cette thèse contribue à la littérature sur le rôle des catégories sur les marchés, à la recherche émergente sur le niveau de distinction optimal et aux approches computationelles de l'étude des organisations.

Title : Explaining the Variability of Audiences' Valuations: An Approach Based on Market Categories and Natural Language Processing

Keywords : Categorization, Valuation, Natural Language Processing, Publicly listed firms

Abstract : This dissertation examines whether the different categorization processes shaping audiences' valuations in markets bring stability or variability to audiences' valuations. While seminal research on categorization emphasized the stabilizing role of market categories, recent research suggests that audiences' valuations can vary substantially even in markets which are well-structured by pre-existing categories. This variability notably results from audiences' heterogeneous preferences for typical offerings, from shifts in categories' meanings or from audiences' reliance on multiple models of valuation. Taking stock of these new results, this dissertation asks why audiences' valuations are so variable and explores in more details the role that market categories play in this phenomenon. This dissertation proposes that i) ambiguous categories, ii) the influence of temporary attractions among audiences alongside more stable categories and iii) the co-existence of different types of evaluators all contribute to produce variability in audiences' valuations. The first two empirical essays use data from publicly listed firms in the U.S. In these essays, firms' similarity to existing category prototypes or audiences' temporary attractions toward certain features are measured using semantics extracted from large corpora of annual reports and IPO prospectuses. The third essay is a theoretical model. This dissertation contributes to the literature on market categories, to the burgeoning research on optimal distinctiveness and to computational approaches to the study of organizations.

Institut Polytechnique de Paris
91120 Palaiseau, France