Who needs a shrink when you have *Businessweek?*

Using content analysis to get inside the heads of entrepreneurs, VCs and other market participants

Timothy G. Pollock and Kisha Lashley

Introduction

A great deal of entrepreneurship research involves trying to access the minds of entrepreneurs, venture capitalists (VCs), angels and the general market. The goals are varied. Some scholars are interested in whether and how entrepreneurs recognize and decide to pursue opportunities (Baron and Ensley 2006; Grégoire et al. 2010). Others are interested in their perceptions of risk (Carpenter et al. 2003), the strategies they pursue (Mishina et al. 2004; Wasserman 2006; Zachary et al. 2011b) or how they view the resources available to them (Baker and Nelson 2005). Still other scholars seek to understand how VCs decide which firms to fund (Kirsch et al. 2009; Zacharakis and Shepherd 2005) or how they manage uncertainty (Pollock et al. 2009; Wasserman 2003). Some focus on how markets make sense of and value firms when they do not possess much information (Nelson 2003; Pollock and Rindova 2003; Pollock et al. 2008), how entrepreneurial firms attempt to influence investors' perceptions (Lounsbury and Glynn 2001; Martens et al. 2007) or build intangible assets such as reputation and celebrity (Rindova et al. 2006, 2007).

Gaining access to this kind of information is difficult. Although a variety of approaches including surveys, policy capturing and qualitative research using interviews and participant observation have been employed with some success, these methods all share some common limitations. One limitation is that they are intrusive; that is, they involve direct interaction with the subject, which poses certain threats to internal validity (Shadish et al. 2002). Another limitation is that they are costly in terms of money, time and effort, thereby limiting the size of the samples used. They are also bound by the data collection's time period, making meaningful longitudinal studies difficult, if not impossible. However, there is a method that can provide access to individual and collective cognitions that is unobtrusive, relatively inexpensive and is not bound by time: the content analysis of texts (Duriau et al. 2007; Pollach 2012; Short and Palmer 2008).

This chapter's purpose is to identify when and how content analysis can be applied to questions interesting to entrepreneurship scholars. We first describe the basic concepts and processes that fall under the content-analysis umbrella, and then discuss the different texts entrepreneurship scholars have analysed, providing examples of the types of studies conducted and questions pursued with content analysis. Finally, we identify a set of broad research questions and topics that entrepreneurship scholars can pursue using content analysis. We focus primarily on the use of content analysis in quantitative research. We do not discuss inductive, qualitative research – computer-aided or otherwise. While these methods have a long tradition in entrepreneurship research, they also suffer from the limitations of intrusiveness, cost and being bounded in time.

What is content analysis?

Content analysis is an umbrella term used to describe a range of techniques for analyzing texts, images and other symbolic data (Duriau et al. 2007; Krippendorff 2012; Pollach 2012). 'The key assumption [of content analysis] is that the analysis of texts lets the researcher understand other people's cognitive schemas' (Duriau et al. 2007: 6). The tools employed assess natural language in a variety of ways, including employing word counts based on dictionaries of terms to identify the presence or frequency of particular constructs, holistic assessments of blocks of text and analysing the juxtaposition of different words and phrases to gain insights into individuals' perceptions, feelings and beliefs. Content analysis can be conducted manually by trained coders, or it can be automated – a class of techniques known as 'computer-aided text analysis', or CATA.

Benefits of content analysis

Because these tools can be applied unobtrusively, they are not subject to validity threats associated with retrospective sensemaking, testing (i.e. influencing the results by collecting data from a respondent over multiple periods), instrumentation (i.e. changes in how the data is collected, including the effects of the researcher) and other threats to internal validity (Shadish et al. 2002). Content analysis is also inexpensive relative to other, more intrusive methods. It can be conducted on texts that are freely available or have been collected for other purposes; it does not require identifying or gaining the consent of the individual or organization being studied (or your university's human subjects committee!); and you do not have to create and validate instruments to collect data from respondents. Textual data can also be collected retrospectively, thereby facilitating the creation of large samples and longitudinal research designs. Content analysis is also a replicable means for gaining access to (sometimes unconscious) cognitive structures, such as managerial attributes (Barr et al. 1992), schemas (Fiol 1995) and causal reasoning (Bettman and Weitz 1983). Content analysis is also extremely flexible, and can be used in conjunction with other sources of data in quantitative analysis (Duriau et al. 2007).

Selecting the unit of analysis

The type of content analysis you employ is a function of the unit of text you choose to analyse. Texts can be analysed at the word, phrase, sentence and paragraph levels, or at the level of the full text. For example, if you are interested in identifying a particular temporal orientation, such as past or future, you would likely choose the word as your unit of analysis and code the frequency with which words employing the past or future tense appeared in the text. If you are interested

in a more complex construct, such as the extent to which a CEO is focused on the alignment of the firm's strategy with shareholder interests in maximizing profits, you might employ the phrase or sentence as the unit of analysis (Wade et al. 1997). If you are conducting more thematic forms of analysis that require assessing the logical structure of narratives or the types of arguments used, then the paragraph or full text may be a more useful unit of analysis (Martens et al. 2007). The smaller the unit of analysis, the easier it is to use CATA to analyse the data. Larger units of analysis are more likely to require using trained coders, which can be more costly and limit the size of the text corpus analysed, but provides the ability to assess more complex constructs and cognitive processes.

It is important, however, to distinguish between the unit of analysis that you will be content analysing and the level of aggregation at which you collect and use the content-analysed data. For example, when assessing the relative positive emotional content of firms' media coverage, Pfarrer et al. (2010) coded the frequencies of positive and negative affect words but aggregated their data at the overall level of the full text, and then assessed the overall affectivity of each firm's media coverage across all the articles about the firm in a given year.

Developing concept categories

Perhaps the biggest challenge in using content analysis is developing dictionaries that accurately capture the concept categories you want to study. This is particularly critical when employing CATA, because the software programmes used cannot make judgement calls. They employ whatever decision rules you establish and search for the words you identify are indicative of a particular concept. The challenge lies in identifying words and phrases that reliably capture instances of the concept of interest (i.e. minimizing Type II errors) while avoiding 'false positives' or coding the concept as present when it is not (i.e. avoiding Type I errors) (Wade et al. 1997). The three most common strategies for developing concept categories are (1) employing validated dictionaries of previously defined concepts; (2) starting with a partially defined set of categories and then adding, deleting and modifying the concept categories and associated dictionaries as the analysis proceeds; and (3) starting with no pre-defined concepts or terms, instead letting them emerge from the text and/or interaction between the text and theory inductively.

A number of content analysis software packages include pre-defined and validated dictionaries of terms associated with different constructs. Two of the more popular programmes used in strategy and entrepreneurship research are Linguistic Inquiry Word Count (LIWC) (Bednar 2012; Pennebaker et al. 2003; Pfarrer et al. 2010) and DICTION (Hart 2000; Short and Palmer 2008). The advantages of using pre-defined concepts and dictionaries are that it saves an enormous amount of time, and the validity and reliability of the dictionaries have already been established using a variety of texts. The potential challenge, of course, is that you are constrained in the types of constructs you can consider. However, these programmes are flexible; they allow you to modify the dictionaries or add new, custom dictionaries to capture constructs of interest to you.

More challenging, but with greater potential reward, is developing your own custom categories and dictionaries. This process can be time consuming and involves a fair bit of trial and error. Researchers often start with a set of categories at least partially derived from the literature or phenomenon (although they may be inducted from the text itself). The next step involves identifying words that are likely to be associated with the categories. These words may be taken from a thesaurus; result from brainstorming sessions; be derived from fuzzy searches of the text corpus using if/then statements, or from searches of certain words within particular ranges of other words; or borrowed from previously created dictionaries. Once the words are

identified, the dictionary's validity must be established, typically by calculating inter-rater reliabilities between a computer-coded sample of text and the same text coded manually (Wade et al. 1997). The dictionary is validated when it reliably captures most instances of words associated with the construct of interest while resulting in few false positives. Wade and colleagues (1997) suggest that dictionaries that identify 80 per cent or more of the instances of a construct with 5 per cent or fewer false positives are valid and reliable. Short and colleagues (2009) also provide an in-depth description of the process for developing custom dictionaries and assessing validity. Cluster analyses can also be employed to establish construct validity (Porac et al. 2002).

Texts and methods used in entrepreneurship research

To engage in content analysis you first need text to analyse. In this section we describe some of the different texts that entrepreneurship researchers have content analysed, the constructs studied and the measures generated. They include publicly available government filings, media coverage, press releases, business plans, websites and award applications.

Government filings

Regular government filings offer rich sources of text data that are freely available online (e.g. www.sec.gov/edgar, accessed 1 May 2014) and already in machine-readable formats. The repetitive nature of many required filings, such as annual reports and proxy statements, facilitate collecting data across years.[1] Within the entrepreneurial realm, another valuable source of text data is the initial public offering (IPO) prospectus that firms must file when the company plans to go public. While strategy scholars have made significant use of these documents (see Duriau et al. 2007 and Short and Palmer 2008 for reviews), entrepreneurship scholars have not taken similar advantage. However, there have been some entrepreneurship studies that used text data from offering prospectuses and letters to shareholders to explore different questions.

Offering prospectuses

Offering prospectuses are both the richest and most under-used source of text data on entrepreneurial firms. While numerous studies have been conducted using other data coded from offering prospectuses, few have made use of the narratives provided.

The most frequently used narrative portion of the IPO prospectus is the 'risk factors' section, where firms must describe all the risks they face and everything that could possibly go wrong during and after the offering. Scholars have used counts of the number of risk factors listed as proxies for the overall riskiness of the IPO firm (e.g. Carpenter et al. 2003; Pollock 2004; Welbourne and Andrews 1996). Certain risk factors that are 'boiler-plate' risks, like 'no previous public market for the stock', are sometimes excluded (Pollock 2004). In their study of VCs' and executives' perceptions of risk, Carpenter and colleagues (2003) were interested in IPO firms' global strategic intent; that is, the extent to which they intended to pursue international markets in the future. They identified five different risk factors specifically associated with internationalization and calculated an index based on the number of these risk factors present in an offering divided by five.

Martens et al. (2007) conducted the only study we are aware of that content analyzed IPO prospectus narratives. They combined qualitative and quantitative techniques to study the effects of storytelling on a firm's ability to secure capital. In the first stage of their analysis they used what they called *latent* content analysis – an interpretive reading of the symbolism underlying

the text data – to establish the firm's identity as presented in the prospectus. This involved one author reading the text and making decisions about the underlying message. The other authors then examined the output from the first author for common storylines and inducted a set of six identity categories. Once the identity categories were established, two coders went through the business sections of all the prospectuses and coded them into one of the categories.

In the second stage of the analysis – which they called *manifest* content analysis – they coded and counted specific constructs. The authors identified 76 different risks in the risk–factors section and 17 different strategic actions firms intended to take, and created a 'story net' for each firm that linked the risks with the strategic actions intended to address those risks. They then counted the number of linkages to develop a density measure of the ties reflecting the degree of story elaboration.

They also created two measures that captured the familiarity and unfamiliarity of each strategic action. They determined the frequency with which each strategic action was mentioned by other IPO firms in the same industry the prior year, and then summed the number of familiar (used by 50 per cent or more of IPO firms) and unfamiliar (used by less than 50 percent of IPO firms) strategic actions described.

These measures were used to test hypotheses about the effectiveness of different narrative strategies on resource acquisition. They found that influential narratives (1) construct unambiguous identities for entrepreneurial firms, (2) elaborate how the proposed means of exploitation will attenuate risk (without providing overly complex explanations) and (3) invoke familiar elements to contextually ground those that are less familiar.

Letters to shareholders

The CEO's letter to shareholders published in annual reports is probably the most content-analysed public document in strategy research (see Duriau et al. 2007 and Short and Palmer 2008 for extensive reviews). They are useful because they occur each year and follow a relatively consistent format: offering explanations of the firm's performance over the past year, and calling attention to any new initiatives or plans for maintaining or improving performance. They are also the one document where CEOs speak directly to shareholders. Even though the CEOs may not have written every word, they are intimately involved in the creation of the letter, and it is reasonable to assume they reflect the CEOs' perceptions and beliefs (Fiol 1995). As such, letters to shareholders are a reliable tool for capturing the cognitions of CEOs and other top executives (Zachary et al. 2008b). Despite their availability and utility, entrepreneurship researchers have made limited use of these documents.

There are, however, a few notable examples of entrepreneurship studies that content analysed letters to shareholders. Zachary et al. (2011b) used content analysis to develop a measure of the market orientation of family and non-family businesses. Using a five-year sample of firms that belonged to the Standard and Poor's 500 (S&P 500) – a stock market index based on the leading 500 US publicly traded companies – they classified family firms as those having a founder and/or direct family members of the founder on the top management team or board of directors.

Analysing 1,120 shareholder letters from 224 businesses, they developed and manually validated custom dictionaries for five dimensions of market orientation: three core components (customer orientation, competitor orientation and interfunctional coordination) and two decision components (long-term focus and profitability). Using DICTION 5.0, they first counted the frequency with which words representing each of the five dimensions appeared in the letters to shareholders. Then, they calculated the five-year average number of words for each dimension (in order to smooth out differences in the annual data). Finally, to calculate the overall market

orientation of each firm, they summed the average number of words in each category. Using ANOVAs for their analysis, they found that family businesses had lower market orientations than non-family businesses and put less emphasis on profitability.

Shareholder letters are also useful for exploring forms of organizational identity, such as entrepreneurial orientation (EO). EO is defined as the processes, practices and decision-making activities of entrepreneurial firms (Lumpkin and Dess 1996). Short and colleagues (2009) content analysed 1,278 shareholder letters issued by 426 firms that were in the S&P 500 from 2001 through 2003 to investigate the differences in EO between family and non-family firms.

Their dictionary development was theory driven. They first developed word lists for each dimension of EO and then applied the validated lists to the shareholder letters. They used *The synonym finder* (Rodale 1978) to develop their dictionaries, and validated them by having multiple authors assess the word lists independently, retaining words they agreed were related to the theoretical constructs.

They then used DICTION 5.0 to apply their custom dictionaries to the shareholder letters; employed t-tests to assess whether family firms made significant use of language associated with these dimensions; and used multivariate analysis of variance to test whether the dimensions of EO were more evident in family or non-family firms. They found that although family firms used language consistent with all the dimensions of EO in their shareholder letters, they used language associated with autonomy, proactiveness and risk taking less often than non-family firms.

Media coverage

Media coverage offers another rich, and relatively under-used, source of text data for entrepreneurship scholars. Although media accounts can be a source of direct quotations from entrepreneurs that provide insights into their thinking, they have greater utility for establishing how *others* perceive entrepreneurial firms. Thus, they are useful for answering questions about the social construction of entrepreneurial markets (Kennedy 2008; Pollock et al. 2008), firm value (Pollock and Rindova 2003) and measuring social approval assets, such as legitimacy, reputation and celebrity (Deephouse 2000; Kennedy et al. 2012; Pfarrer et al. 2010).

The two most common constructs derived from media coverage have been the volume of media coverage firms receive, and the positive or negative tenor of that coverage. Since the volume of coverage does not require any content analysis beyond verifying that a particular firm is discussed in an article, we focus our discussion on how the tenor of media coverage has been measured and used.

Pollock and Rindova (2003) studied how the volume and tenor of media coverage firms received prior to their IPOs affected the amount of under pricing (the percentage change in stock price on the first day of trading) and turnover (the percentage of shares offered that were traded) on the day they went public. They used a trained coder to content analyse 514 media articles published prior to the IPOs of 225 firms that went public in 1992. Each paragraph in an article specifically mentioning the firm was coded as positive, negative or neutral in tenor. The full text was coded as positive if it contained predominantly positive or a mix of positive and neutral statements, negative if it contained predominantly negative or negative and neutral statements, and neutral if it contained predominantly neutral, or an equal mix of positive and negative statements. They then calculated the Janis-Fadner coefficient of imbalance (Deephouse 2000; Janis and Fadner 1965), which reflects the overall positive or negative tenor of the coverage, and included linear and squared values for the tenor (and volume) of media coverage in their models.

Consistent with their hypotheses, they found that the volume of media coverage has a negative but diminishing effect on under-pricing and a positive but diminishing effect on turnover. However, they also found that tenor experienced threshold effects and influenced under pricing and turnover in unexpected ways. Positive tenor had no effect on under pricing up to a threshold level, but after this 'tipping point' it had a positive, non-linear relationship with under pricing. Positive tenor also had no effect on turnover up to a threshold level, but thereafter had a significant and negative, non-linear relationship with turnover. They argued these findings provided evidence that the volume of media coverage influenced investor interest and attention, while the tenor of media coverage influenced investor preferences.

In a second study, Pollock and colleagues (2008) studied availability and information cascades within and between the media and investor communities by exploring the relationships between the volume and tenor of media coverage and daily changes in stock price and trading volume for the 60 days following firms' IPOs. They calculated both daily and cumulative volume and tenor measures of media coverage, using 514 pre-IPO (12 months prior to the IPO) and 401 post-IPO (60 days after the IPO) media articles for 225 firms, generating 13,500 firm–day observations.

The cumulative measures were calculated using the same approach employed in Pollock and Rindova (2003), and were based on all pre- and post-IPO coverage up to two days before the current day, so as to keep them distinct from the daily measures. The daily measures were lagged by one day; daily media attention was coded 1 if a firm received media coverage, and daily media evaluations were coded 1 if a firm received coverage that was positive in tenor. They found that both cumulative and recent investor attention and evaluations generally had positive and significant effects on media attention and evaluations. In contrast, investor attention and evaluations are negatively influenced by recent media attention and evaluations, but are positively influenced by cumulative media attention and evaluations.

Although Pollock and colleagues used manual methods to code media tenor, more recent studies (e.g. Bednar 2012; Pfarrer et al. 2010; Zavyalova et al. 2012) have employed LIWC to automate the coding of positive and negative affective language, enabling the use of much larger bodies of text. While this approach makes it more difficult to associate specific instances of positive and negative language with the focal firm, the process is robust and generates useful measures of media tenor (Zavyalova et al. 2012).

Kennedy (2008) used media coverage to explore how new industry categories are created. He collected over 28,000 media articles and press releases on the computer workstation market between 1980 and 1990. Using custom software that he developed, Kennedy coded the average story density (ASD) for a firm by summing the number of other firms mentioned in conjunction with the focal firm and dividing by the total number of articles that mention the focal firm. He used the linear and squared terms for this measure in his models. He also used the volume of coverage a firm received and its prominence in the industry network, measured using the in-degree measure of centrality based on the number of firms linked to the focal firm, as dependent variables. Kennedy found that ASD had a positive, curvilinear relationship with coverage and prominence, suggesting that linkages to some firms are beneficial as they reduce the firm's obscurity, but too many linkages result in crowding and increase the likelihood the firm will be forgotten. He also found that ASD is negatively related to exiting the market.

Press releases

Press releases are another under-used source of text data in entrepreneurship research. They can provide insights into some of the same cognitive constructs and processes as letters to

shareholders, but they are more commonly used as a source of data on firms' strategic actions (Kennedy 2008; Rindova et al. 2007; Zavyalova et al. 2012) and how these actions influence others' perceptions of the firm.

Kennedy (2008) used press releases to calculate average release density (ARD) – a measure analogous to average story density, discussed above – to capture the frequency with which firms mention competitors in their press releases. While the effects of ARD on coverage and prominence are the same, the effects of ARD on exit vary over time. In the industry's early years higher levels of ARD reduce the likelihood of exit. However, as the industry category matures, lower levels of ARD decrease the likelihood of exit, while higher levels of ARD increase the likelihood of exit.

Although not an entrepreneurship study, Zavyalova and colleagues (2012) provide a nice illustration of how press releases can be used to operationalize strategic actions. They used the content of press releases to predict how the media would react to different types of firm responses to product recalls in the toy industry; differentiating between recalls for products offered by the focal firm and recalls by other firms in the industry. Their dependent variable, media tenor, was calculated from a corpus of 37,500 articles and blog postings about the 45 firms in the industry, 21 of which experienced at least one recall. They analysed data on a quarterly basis from 1998 to 2007, yielding 940 firm-quarter observations.

The authors manually coded over 5,500 press releases. Firms' responses to recalls were coded as either ceremonial (i.e. do not directly address the recall, but highlight other positive aspects of the firm) or technical (i.e. directly address the problem of manufacturing and selling defective toys). They found that ceremonial actions had a positive effect on media tenor when the focal firm was not the subject of the recall, but had a negative effect on media tenor when the firm was the subject of the recall. In contrast, technical actions helped attenuate the negative effects of product recalls for the focal firm, but were less effective at attenuating the negative spillover effects of recalls by other toy manufacturers.

Rindova and colleagues (2007) also manually coded press releases in their inductive study of how three new firms, Amazon, Barnes and Noble.com and CDNow, built their reputations. They identified five classes of strategic actions – new service development, customer relations, partnering, market and symbolic actions – and matched them with the co-occurrence of reputation-building elements reflected in their media coverage. They found that although CDNow entered the market first they took the fewest number of actions early on, stepping up their activity only after being threatened. Amazon, on the other hand, took a significant number of actions early on and continued to do so throughout the period of study. Barnes and Noble.com took more moderate levels of actions, assuming that the brand identity of its parent firm would have greater influence on the spin-off. In particular, Amazon took higher levels of customer service, new service development and symbolic actions than the other firms, and also built the strongest reputation.

Business plans

Business plans are another potentially rich, but little-used source of text data (Forbes and Kirsch 2011; Kirsch et al. 2009). While they are often derided for their lack of accuracy and the magical thinking reflected in their financial projections, these documents offer tremendous insights into how entrepreneurs view their firms and markets, make sense of opportunities and assess risks. Even 'bad' business plans from an investor's perspective can provide useful insights into entrepreneurs' perceptions and cognitive processes. Unfortunately, thus far business plans have not been used for this purpose.

But, they have been used to explore how VCs make investment decisions. Business plans are written for those with the ability to provide resources to start-ups. To be effective they must communicate information that facilitates positive assessments (Rosch 1975), and thus have been evaluated for the effectiveness of the cues encoded in their texts.

Kirsch et al. (2009) used content analysis in a study of 722 business plans entrepreneurs submitted to one VC firm to solicit funding. They manually coded the business plans, using the following cues to predict success in receiving VC funding: plan completeness, human capital, educational human capital, absolute size, entrepreneurial experience, professional experience, entrepreneurial prominence, team completeness and preparation, commitment of resources and the amount requested. They concluded that business plans did not play a direct communicative role (i.e. they were not predictive of VC decision making), but that they did play a weak ceremonial role; that is, their presence was indicative of VC decision making influenced by similar information collected through means other than the business plans.

MacMillan and Narasimha (1987) content analysed 82 business plans submitted to five New York VC firms. They had trained research assistants manually code the plans for financial projections and calculated the ratio of the smallest expense to the largest expense on the most recent income statement. In addition, they counted the nouns and adjectives in the plans' executive summaries and assessed the percentage of each plan that was dedicated to its constituent elements of marketing, management, finance and production. They found that plans with overly optimistic revenue and performance projections were less successful at getting funding than more realistic plans. They theorized that VCs have certain notions about what is achievable, and will not seriously consider funding a new firm whose plan falls outside this 'credibility window'. They also found that plans paying either too much or too little attention to any of the functional areas had less success getting funded.

Websites

Scholars have also begun collecting text data on entrepreneurial firms from company websites (McKenny et al. 2012; Zachary et al. 2011a, 2012). Websites offer a potentially rich source of information about private and public companies, and technologies have been developed to 'scrape' or download entire sites quickly and efficiently. Websites provide organizational narratives that convey entrepreneurs' beliefs and values (Zachary et al. 2011b).

Zachary et al. (2011a) used franchisors' recruitment websites to investigate how the franchisors represented their identities. The authors used two samples: a sample of Franchise 500 firms and a random sample of franchisors that did not belong to the Franchise 500 list. They used pre-validated word lists and employed DICTION to examine the differences between high-performing (Franchise 500) and non-high-performing (non-Franchise 500) franchisors in terms of their market orientation, entrepreneurial orientation and charismatic rhetoric. For market orientation, they used word lists developed and validated by Zachary and colleagues (2011b) and Short and colleagues (2009) to measure entrepreneurial orientation. For charismatic leadership, they used language for each of the eight dimensions identified by Shamir and colleagues (1994). Using MANOVA, they found that Franchise 500 firms used more language consistent with market orientation, entrepreneurial orientation and charismatic leadership than non-Franchise 500 firms. They also found that the size of the firm was positively associated with market orientation and charismatic and entrepreneurial rhetoric.

In a subsequent study, Zachary and colleagues (Brigham et al. 2013; Zachary et al. 2012) developed a measure of long-term orientation (LTO) in entrepreneurial firms by content analysing the language in the 'about us' section of the websites of Inc. 500 and S&P 600 Growth firms.

They argued that the 'about us' section is a valuable firm-level source of information because companies use it 'to explain to stakeholders who they are and what makes them who they are' (Zachary et al. 2012: 9).

They conceptualized LTO as having three dimensions: continuity, futurity and perseverance. In developing the LTO construct they examined its construct validity, external validity, reliability, dimensionality and criterion validity using content analysis. To analyse construct validity – the extent to which a measure represents the construct – the researchers began with a theoretical definition for each dimension of the LTO construct, generated a dictionary of words for those dimensions, then used *The synonym finder* (Rodale 1978) to obtain additional words. Three of the authors assessed the accuracy of the words in capturing the intended meaning. They then used language from the 'about us' pages of the entrepreneurial firms to find additional words that fit the construct and repeated the inter-rater reliability procedure. They added the finalized lists to DICTION and content analysed the websites.

For external validity – the generalizability of results across individuals, contexts and time periods – Zachary and colleagues (Brigham et al. 2013; Zachary et al. 2012) drew their sample from different but related business indices and across different time periods. They also content analysed two different sources of firm-level data: the 'about us' section of their websites and letters to shareholders. To assess reliability – the degree of constant stability, dependability and predictability of a measure – the authors used CATA to measure LTO in firms. That way, they were able to ensure consistency of measurement across samples. For construct dimensionality – the relatedness between individuals and their association to one construct – the authors created multiple word lists for each dimension of the construct and examined the correlations among the dimensions, employing confirmatory factor analysis to test the construct's multidimensionality. Finally, for criterion validity (the causal linkage between a measure and one or more external variables), they used a multinomial logit regression to examine the probability of firms using LTO in one industry versus another. In addition, they used structural equation modelling to examine the LTO–revenue–growth relationship.

McKenny and colleagues (2012) used the narratives on company websites to understand the difficult-to-capture construct of espoused goals (the desired economic or noneconomic outcomes of an organization as communicated to stakeholders via organizational narratives) and how firms' identities were reflected in those goals. They were primarily interested in demonstrating that content analysis is an appropriate method for studying such a construct.

Using a sample of 77 companies with available website narratives – 40 of which also had press releases and/or posted news articles – the researchers identified, tabulated and categorized references to organizational goals. Then, based on their alignment with theoretical definitions, they categorized the goals as either normative or utilitarian. Using t-tests, they found that family firms discussed a significant number of utilitarian and normative goals in the 'about us' sections of their websites. To assess the validity of their coding process, they correlated the number of utilitarian and normative goals with firm revenue and found that utilitarian goals were positively correlated with revenue.

Finally, Nicholls (2010) set out to understand the views of prominent paradigm-building actors in the pre-paradigmatic field of social entrepreneurship. He argued social entrepreneurship lacked both an agreed-upon definition of what the term means and an established epistemology. He defined paradigm-building groups as those that are prominent in existing literature, as well as those with significant investments in the field: governments, foundations, fellowship organizations and network organizations. Based on a content analysis of websites of representatives from these groups, Nicholls was able to identify two main sets of discourses: the first positions

the social entrepreneur as a hero, while the other situates social entrepreneurship within communities and networks of action. In the community-focused discourse, he found further tensions between proponents for a business-like social action and those who put forth frameworks of advocacy and social change for social entrepreneurship.

Award applications

A final source of text data are narratives written for award nominations and other competitions. Although written for admittedly self-serving purposes, these documents can still provide insights into entrepreneurs' cognitive processes and perceptions (Barr 1998; Fiol 1995; Porac et al. 2002).

One source of application narratives used in several studies is the *Inc. Magazine*/Ernst and Young Entrepreneur of the Year (EOY) competition nominations. When the Kauffman Foundation became involved in the EOY competition in the 1990s they acquired all of the narratives written by nominated entrepreneurs as part of the application process. The database included all the regional and national finalists over the early and mid-1990s. This database has been used in multiple studies by Mishina and colleagues (Mishina et al. 2004; Porac et al. 2002) and Barringer and colleagues (Barringer and Greening 1998; Barringer et al. 2005).

Porac and colleagues (2002) content analysed the narratives of 54 companies whose founders were candidates for the EOY award to create a cognitive map of the entrepreneurs' growth strategies. They first coded each sentence in the narratives into one of six categories: CEO characteristics, company characteristics, company capabilities, growth strategies, image of the market and other. Next, they coded each of the growth-strategies sentences, iterating between the data and theory to identify six different growth strategies: capital intensive, non-capital intensive, market expansion, product or service expansion, human resources improvements and process improvements. They then conducted a cluster analysis using the presence or absence of each of these strategies in the narrative to identify whether the six strategies were grouped in different ways. They identified five strategic cluster configurations: expansion of products via continuous improvement, market and product expansion, dealing with capacity deficits, anticipatory growth and scattered growth.

In a second study, Mishina and colleagues (2004) used a sample of 112 public manufacturing firms drawn from the EOY sample to explore how the fit between firms' entrepreneurial growth logics and their financial and human resource slack influenced short-term growth rates. In this study they derived two 'dominant' growth logics from the literature: growth by expanding into new markets with existing products, and growth by developing new products for existing markets. They theorized that these two growth logics would vary in their complexity and the novelty of routines involved in their implementation. They first coded the EOY narratives by identifying future-oriented sentences discussing firm growth (as opposed to sentences discussing past growth). They then coded these future-oriented sentences for the presence of each growth logic, creating two dummy variables. They validated their coding by randomly selecting a 20 per cent subsample of firms and coding their letters to shareholders in their annual 10-K statements using the same approach.

They found that firms pursuing growth through new products grew more slowly than firms not pursuing this strategy, but the presence of financial slack positively moderated this relationship. They also found that firms pursuing a market growth strategy grew more quickly when they possessed higher levels of human-resource slack. In a post-hoc analysis they used finer distinctions for different kinds of product and market growth strategies with different levels of complexity, and found some evidence that more complex market growth strategies, such as

They argued that the 'about us' section is a valuable firm-level source of information because companies use it 'to explain to stakeholders who they are and what makes them who they are' (Zachary et al. 2012: 9).

They conceptualized LTO as having three dimensions: continuity, futurity and perseverance. In developing the LTO construct they examined its construct validity, external validity, reliability, dimensionality and criterion validity using content analysis. To analyse construct validity – the extent to which a measure represents the construct – the researchers began with a theoretical definition for each dimension of the LTO construct, generated a dictionary of words for those dimensions, then used *The synonym finder* (Rodale 1978) to obtain additional words. Three of the authors assessed the accuracy of the words in capturing the intended meaning. They then used language from the 'about us' pages of the entrepreneurial firms to find additional words that fit the construct and repeated the inter-rater reliability procedure. They added the finalized lists to DICTION and content analysed the websites.

For external validity – the generalizability of results across individuals, contexts and time periods – Zachary and colleagues (Brigham et al. 2013; Zachary et al. 2012) drew their sample from different but related business indices and across different time periods. They also content analysed two different sources of firm-level data: the 'about us' section of their websites and letters to shareholders. To assess reliability – the degree of constant stability, dependability and predictability of a measure – the authors used CATA to measure LTO in firms. That way, they were able to ensure consistency of measurement across samples. For construct dimensionality – the relatedness between individuals and their association to one construct – the authors created multiple word lists for each dimension of the construct and examined the correlations among the dimensions, employing confirmatory factor analysis to test the construct's multidimensionality. Finally, for criterion validity (the causal linkage between a measure and one or more external variables), they used a multinomial logit regression to examine the probability of firms using LTO in one industry versus another. In addition, they used structural equation modelling to examine the LTO–revenue-growth relationship.

McKenny and colleagues (2012) used the narratives on company websites to understand the difficult-to-capture construct of espoused goals (the desired economic or noneconomic outcomes of an organization as communicated to stakeholders via organizational narratives) and how firms' identities were reflected in those goals. They were primarily interested in demonstrating that content analysis is an appropriate method for studying such a construct.

Using a sample of 77 companies with available website narratives – 40 of which also had press releases and/or posted news articles – the researchers identified, tabulated and categorized references to organizational goals. Then, based on their alignment with theoretical definitions, they categorized the goals as either normative or utilitarian. Using t-tests, they found that family firms discussed a significant number of utilitarian and normative goals in the 'about us' sections of their websites. To assess the validity of their coding process, they correlated the number of utilitarian and normative goals with firm revenue and found that utilitarian goals were positively correlated with revenue.

Finally, Nicholls (2010) set out to understand the views of prominent paradigm-building actors in the pre-paradigmatic field of social entrepreneurship. He argued social entrepreneurship lacked both an agreed-upon definition of what the term means and an established epistemology. He defined paradigm-building groups as those that are prominent in existing literature, as well as those with significant investments in the field: governments, foundations, fellowship organizations and network organizations. Based on a content analysis of websites of representatives from these groups, Nicholls was able to identify two main sets of discourses: the first positions

the social entrepreneur as a hero, while the other situates social entrepreneurship within communities and networks of action. In the community-focused discourse, he found further tensions between proponents for a business-like social action and those who put forth frameworks of advocacy and social change for social entrepreneurship.

Award applications

A final source of text data are narratives written for award nominations and other competitions. Although written for admittedly self-serving purposes, these documents can still provide insights into entrepreneurs' cognitive processes and perceptions (Barr 1998; Fiol 1995; Porac et al. 2002).

One source of application narratives used in several studies is the *Inc. Magazine*/Ernst and Young Entrepreneur of the Year (EOY) competition nominations. When the Kauffman Foundation became involved in the EOY competition in the 1990s they acquired all of the narratives written by nominated entrepreneurs as part of the application process. The database included all the regional and national finalists over the early and mid-1990s. This database has been used in multiple studies by Mishina and colleagues (Mishina et al. 2004; Porac et al. 2002) and Barringer and colleagues (Barringer and Greening 1998; Barringer et al. 2005).

Porac and colleagues (2002) content analysed the narratives of 54 companies whose founders were candidates for the EOY award to create a cognitive map of the entrepreneurs' growth strategies. They first coded each sentence in the narratives into one of six categories: CEO characteristics, company characteristics, company capabilities, growth strategies, image of the market and other. Next, they coded each of the growth-strategies sentences, iterating between the data and theory to identify six different growth strategies: capital intensive, non-capital intensive, market expansion, product or service expansion, human resources improvements and process improvements. They then conducted a cluster analysis using the presence or absence of each of these strategies in the narrative to identify whether the six strategies were grouped in different ways. They identified five strategic cluster configurations: expansion of products via continuous improvement, market and product expansion, dealing with capacity deficits, anticipatory growth and scattered growth.

In a second study, Mishina and colleagues (2004) used a sample of 112 public manufacturing firms drawn from the EOY sample to explore how the fit between firms' entrepreneurial growth logics and their financial and human resource slack influenced short-term growth rates. In this study they derived two 'dominant' growth logics from the literature: growth by expanding into new markets with existing products, and growth by developing new products for existing markets. They theorized that these two growth logics would vary in their complexity and the novelty of routines involved in their implementation. They first coded the EOY narratives by identifying future-oriented sentences discussing firm growth (as opposed to sentences discussing past growth). They then coded these future-oriented sentences for the presence of each growth logic, creating two dummy variables. They validated their coding by randomly selecting a 20 per cent subsample of firms and coding their letters to shareholders in their annual 10-K statements using the same approach.

They found that firms pursuing growth through new products grew more slowly than firms not pursuing this strategy, but the presence of financial slack positively moderated this relationship. They also found that firms pursuing a market growth strategy grew more quickly when they possessed higher levels of human-resource slack. In a post-hoc analysis they used finer distinctions for different kinds of product and market growth strategies with different levels of complexity, and found some evidence that more complex market growth strategies, such as

internationalization, led to slower growth than less complex product growth strategies, such as minor product line extensions.

Barringer et al. (2005) examined the attributes and behaviours that allow firms to achieve and maintain rapid growth. Rapid growth was defined as at least a 15 per cent increase in the number of employees per year. They used an in-depth literature review and a pilot study of 25 narratives to identify the relevant variables. They then used the ATLAS/ti software program to code the narratives at the sentence level into the four categories of founder characteristics, firm attributes, business practices and human resource management practices. Finally, they used a one-tailed, Fisher's exact test to test the differences between the frequencies for the rapid-growth and slow-growth firms. They concluded that the founders of rapid-growth firms were better educated than their slow-growth counterparts, that rapid-growth firms had a stronger commitment to growth, were more involved in interorganizational relationships, and had a more growth-oriented mission statement than slow-growth firms.

Areas for future research

Content analysis is a useful tool for studying entrepreneurship, particularly areas such as entrepreneurial values, intentions and cognition; identifying stakeholders' perceptions and how they influence firm value and the evolution and social construction of entrepreneurial markets; and the rhetorical strategies entrepreneurs use to influence stakeholders' perceptions. It can be used to study entrepreneurial phenomena at the individual, firm and field levels. However, we have only begun to scratch the surface. Future research needs to be more ambitious in the types of analyses conducted, the constructs measured and the datasets used. We also need to take advantage of new and different sources of text that allow us to ask different kinds of questions and to study different kinds of entrepreneurs and entrepreneurial firms.

Analyses and constructs considered

Many of the studies we described are largely descriptive, establishing that firms are different on some dimension such as entrepreneurial orientation, but stopping there. The next step for entrepreneurship scholars is to take these dictionaries and measures that have been validated in early research and use them as independent and dependent variables in more sophisticated analyses to develop and test new theories about their roles in resource acquisition, firm growth and innovation.

We also need to move beyond coding relatively simple constructs like positive/negative tenor, and develop ways to code more sophisticated constructs, particularly using CATA. Kennedy (2008) provides us with an excellent example of this type of research; but if the requirement is writing your own software, as he did, progress will be slow.

Different sources of text

With the propagation of the internet and social media, scholars have newfound access to rich data sources amenable to content analysis. For example, scholars can now content analyse entrepreneurs' blog and Twitter posts, as well as more traditional business plans. This presents an opportunity to compare the thought processes of different groups of entrepreneurs. It also creates opportunities to conduct longitudinal studies of how perceptions evolve over time, as well as how firm images, stories and even their business models are socially constructed via conversations among groups of actors.

Scholars interested in studying entrepreneurial processes in new industries can also make use of historical archives that are maintained by groups such as industry associations, governments and university historians (Forbes and Kirsch 2011). According to Forbes and Kirsch (2011), these historical sources – such as the Firm and Industry Evolution and Entrepreneurship Project, the Charles Babbage Institute (CBI) and the archives of the Chemical Heritage Foundation (CHF) in the US – allow researchers to look at the different stages of industry evolution. Kirsch has also developed websites such as the Business Plan Archive (www.businessplanarchive.org/) and the Dot Com Archive (www.dotcomarchive.org/) to collect and preserve data that can be used to study a wide array of issues via content analysis. Future research should make use of these valuable and largely untapped sources of data.

Different types of entrepreneurs

Most, but not all, the entrepreneurship research has focused on entrepreneurial firms that are public, or were in the process of going public. While extremely valuable and worthy of further study, these firms represent only a small slice of the types of entrepreneurs that we can study. Expanding the types of entrepreneurs studied can also expand the types of questions asked.

For example, scholars interested in social entrepreneurship have been challenged regarding whether and how social entrepreneurship differs from traditional entrepreneurship or non-profit activities (Dacin et al. 2011). Content analysis of social entrepreneurs' accounts through social media and other traditional sources of text data could reveal variation in how they perceive the differences in what they do that can be used to clarify the definitional differences between social entrepreneurship and other types of ventures, or create a typology of different types of social entrepreneurship ventures.

Content analysis can also be used to understand how social entrepreneurs gain legitimacy when they do not fit easily into existing cognitive categories (Dacin et al. 2011; Kennedy 2008), how what it means to be a social entrepreneur has evolved over time (Kennedy et al. 2012), and the extent to which these processes are different than for other types of entrepreneurs or new industries. Because we can content analyse the media to tap into societal beliefs and values, entrepreneurship scholars can investigate changes in collective beliefs about different entrepreneurial practices over time and the institutional contexts within which these firms exist.

Other types of entrepreneurs we can study include ethnic entrepreneurs, entrepreneurs in developing countries, and entrepreneurs with small, private firms. These types of firms often participate in developmental or capacity-building programmes, such as those backed by governments or non-governmental agencies. To justify the need for these programs, the sponsoring agencies often keep records that include narrative data, which entrepreneurship scholars can content analyse to explore questions related to institutional and/or societal justifications for promoting business ownership.

The purpose of this chapter is to introduce entrepreneurship scholars to a powerful method for studying perceptions, attitudes and cognitive processes. As technologies become more advanced and the availability of machine-readable text increases, content analysis becomes more and more viable as a means for studying a wide range of questions and phenomena relevant to entrepreneurship. We encourage you to avail yourself of these opportunities and advance entrepreneurship scholarship by finding innovative ways to employ content analysis in your own research. For more detailed information on the methods of content analysis, see Krippendorff (2012), Neuendorf (2002), Weber (1990) and West (2001).

Recommended readings

Krippendorff, K. (2012) *Content Analysis: An Introduction to its Methodology*, 3rd edition, Thousand Oaks, CA: Sage.

Martens, M.L., Jennings, J. and Jennings, P.D. (2007) 'Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition', *Academy of Management Journal* 50 (5): 1107–32.

Pfarrer, M.D., Pollock, T.G. and Rindova, V.P. (2010) 'A tale of two assets: The effects of firm reputation and celebrity on earnings surprises and investors' reactions', *Academy of Management Journal* 53 (5): 1131–52.

Pollach, I. (2012) 'Taming textual data: The contribution of corpus linguistics to computer-aided text analysis', *Organizational Research Methods* 15 (2): 263–87.

Short, J.C. and Palmer, T.B. (2008) 'The Application of DICTION to Content Analysis Research in Strategic Management', *Organizational Research Methods* 11 (4): 727–52.

References

Baker, T. and Nelson, R. (2005) 'Creating Something from Nothing: Resource Construction through Entrepreneurial Bricolage', *Administrative Science Quarterly* 50 (3): 329–66.

Baron, R.A. and Ensley, M.D. (2006) 'Opportunity recognition as the detection of meaningful patterns: Evidence from comparisons of novice and experienced entrepreneurs', *Management Science* 52: 1331–44.

Barr, P.S. (1998) 'Adapting to unfamiliar environmental events: A look at the evolution interpretation and its role in strategic change', *Organization Science* 9: 644–69.

—— , Stimpert, J.L. and Huff, A.S. (1992) 'Cognitive change, strategic action, and organizational renewal', *Strategic Management Journal* 13: 15–36.

Barringer, B.R. and Greening, D.W. (1998) 'Small business growth through geographic expansion: a comparative case study', *Journal of Business Venturing* 13 (5), 467–92.

—— , Jones, F.F. and Neubaum, D.O. (2005) 'A quantitative content analysis of the characteristics of rapid-growth firms and their founders', *Journal of Business Venturing* 20 (5): 663–87.

Bednar, M. (2012) 'Watchdog or lapdog? A behavioral view of the media as a corporate governance mechanism', *Academy of Management Journal* 55: 131–50.

Bettman, J.R. and Weitz, B.A. (1983) 'Attributions in the board room: Causal reasoning in corporate annual reports', *Administrative Science Quarterly* 28: 165–83.

Brigham, K.H., Lumpkin, G., Payne, G.T. and Zachary, M.A. (2013) 'Researching long-term orientation: a validation study and recommendations for future research', *Family Business Review* 27 (1): 72–88.

Carpenter, M.A., Pollock, T.G. and Leary, M.M. (2003) 'Governance, the experience of principals and agents, and global strategic intent: Testing a model of reasoned risk taking', *Strategic Management Journal* 24: 803–20.

Dacin, M.T., Dacin, P.A. and Tracey, P. (2011) 'Social entrepreneurship: A critique and future directions', *Organization Science* 22 (5): 1203–13.

Deephouse, D.L. (2000) 'Media reputation as a strategic resource: An integration of mass communication and resource based theories', *Journal of Management* 26: 1091–112.

Duriau, V.J., Reger, R.K. and Pfarrer, M.D. (2007) 'A content analysis of the content analysis literature in organization studies: research themes, data sources, and methodological refinements', *Organizational Research Methods* 10 (1): 5–34.

Fiol, C.M. (1995) 'Corporate communications: Comparing executives' private and public statements', *Academy of Management Journal* 38: 522–36.

Forbes, D.P. and Kirsch, D.A. (2011) 'The study of emerging industries: Recognizing and responding to some central problems', *Journal of Business Venturing* 26 (5): 589–602.

Grégoire, D.A., Barr, P.S. and Shepherd, D.A. (2010) 'Cognitive processes of opportunity recognition: The role of structural alignment', *Organization Science* 21: 413–34.

Hart, R.P. (2000) *DICTION 5.0: The text-analysis program*, Thousand Oaks, CA: Sage.

Janis, I.L. and Fadner, R. (1965) 'The coefficient of imbalance', in: Lasswell, H., Leites, N. and Associates (eds.) *Language of politics*, Cambridge, MA: MIT Press: 153–69.

Kennedy, M.T. (2008) 'Getting counted: markets, media, and reality', *American Sociological Review* 73 (2): 270–95.

——— , Chok, J.I. and Liu, J. (2012) 'New criteria for assessing corporate reputation', in: Barnett, M.L. and Pollock, T.G. (eds.) *The Oxford Handbook of Corporate Reputation*, Oxford: Oxford University Press: 69–93.

Kirsch, D., Goldfarb, B. and Gera, A. (2009) 'Form or Substance: The role of business plans in venture capital decision making', *Strategic Management Journal* 30: 487–515.

Krippendorff, K. (2012) *Content Analysis: An Introduction to its Methodology*, 3rd edition, Thousand Oaks, CA: Sage.

Lounsbury, M. and Glynn, M.A. (2001) 'Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources', *Strategic Management Journal* 22: 545–64.

Lumpkin, G. T. and Dess, G.G. (1996) 'Clarifying the entrepreneurial orientation construct and linking it to performance', *Academy of Management Review* 21: 135–72.

MacMillan, I.C. and Narasimha, P.N.S. (1987) 'Characteristics distinguishing funded from unfunded business plans evaluated by venture capitalists', *Strategic Management Journal* 8 (6): 579–85.

Martens, M.L., Jennings, J. and Jennings, P.D. (2007) 'Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition', *Academy of Management Journal* 50 (5): 1107–32.

McKenny, A.F., Short, J.C., Zachary, M.A. and Payne, G.T. (2012) 'Assessing espoused goals in private family firms using content analysis', *Family Business Review* 25: 298–317.

Mishina, Y., Pollock, T.G. and Porac, J.F. (2004) 'Are more resources always better for growth? Resource stickiness in market and product expansion', *Strategic Management Journal* 25: 1179–97.

Nelson, R. (2003) 'On the uneven evolution of human know-how', *Research Policy* 32 (6): 909–22.

Neuendorf, K.A. (2002) *The content analysis guidebook*, Thousand Oaks, CA: Sage Publications.

Nicholls, A. (2010) 'The legitimacy of social entrepreneurship: reflexive isomorphism in a pre-paradigmatic field', *Entrepreneurship Theory and Practice* 34 (4): 611–34.

Pennebaker, J.W., Mehl, M.R. and Niederhoffer, K.G. (2003) 'Psychological aspects of natural language use: Our words, our selves', *Annual Review of Psychology* 54: 547–77.

Pfarrer, M.D., Pollock, T.G. and Rindova, V.P. (2010) 'A tale of two assets: The effects of firm reputation and celebrity on earnings surprises and investors' reactions', *Academy of Management Journal* 53 (5): 1131–52.

Pollach, I. (2012) 'Taming textual data: The contribution of corpus linguistics to computer-aided text analysis', *Organizational Research Methods* 15 (2): 263–87.

Pollock, T.G. (2004) 'The benefits and costs of underwriters' social capital in the U.S. initial public offerings market', *Strategic Organization* 2 (4): 357–88.

——— and Rindova, V.P. (2003) 'Media legitimation effects in the market for initial public offerings', *Academy of Management Journal* 46 (5): 631–42.

——— , Fund, B.R. and Baker, T. (2009) 'Dance with the one that brought you? Venture capital firms and the retention of founder-CEOs', *Strategic Entrepreneurship Journal* 3: 199–217.

——— , Rindova, V.P. and Maggitti, P.G. (2008) 'Market watch: Information and availability cascades among the media and investors in the U.S. IPO market', *Academy of Management Journal* 51 (2): 335–58.

Porac, J.F., Mishina, Y. and Pollock, T.G. (2002) 'Entrepreneurial narratives and the dominant logics of high growth firms', in: Huff, A. and Jenkins, M. (eds.) *Mapping Strategic Knowledge*, Thousand Oaks, CA: Sage, 112–36.

Rindova, V.P., Petkova, A.P. and Kotha, S. (2007) 'Standing out: how new firms in emerging markets build reputation', *Strategic Organization* 5 (1): 31–70.

——— , Pollock, T.G. and Hayward, M.L.A. (2006) 'Celebrity firms: The social construction of market popularity', *Academy of Management Review* 31 (1): 50–71.

Rodale, J.I. (1978) *The synonym finder*, Emmaus, PA: Rodale Press.

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

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Belmont, CA: Wadsworth Centage Learning.

Shamir, B., Arthur, M.B. and House, R.J. (1994) 'The rhetoric of charismatic leadership: a theoretical extension, a case study, and implications for research', *Leadership Quarterly* 5: 25–42.

Short, J.C. and Palmer, T.B. (2008) 'The Application of DICTION to Content Analysis Research in Strategic Management', *Organizational Research Methods* 11 (4): 727–52.

——— , Payne, G.T., Brigham, K.H., Lumpkin, G. and Broberg, J.C. (2009) 'Family firms and entrepreneurial orientation in publicly traded firms: A comparative analysis of the S&P 500', *Family Business Review* 22: 9–24.

Wade, J.B., Porac, J.F. and Pollock, T.G. (1997) 'Worth, words and the justification of executive pay', *Journal of Organizational Behavior* 18: 641–64.

Wasserman, N. (2003) 'Founder CEO succession and the paradox of entrepreneurial success', *Organization Science* 14 (2): 149–72.

—— (2006) 'Stewards, agents, and the founder discount: Executive compensation in new ventures', *Academy of Management Journal* 49 (5): 960–76.

Weber, R. (1990) *Basis content analysis*, 2nd edition, Thousand Oaks, CA: Sage Publications.

Welbourne, T.M. and Andrews, A.O. (1996) 'Predicting performance of initial public offerings: Should human resource management be in the equation?', *Academy of Management Journal* 39 (4): 891–919.

West, M.D. (ed.) (2001) *Theory, Method, and Practice in Computer Content Analysis*, Westport, CT: Ablex Pub.

Zacharakis, A.L. and Shepherd, D.A. (2005) 'A non-additive decision aid for venture capitalists' investment decisions', *European Journal of Operational Research* 162: 673–89.

Zachary, M.A., Brigham, K.H., Payne, G.T. and Lumpkin, G.T. (2012) *Long-term orientation and firm performance: Construct validation and comparative analysis in high-growth publically-traded and private firms*, working paper, Texas Tech University.

—— , McKenny, A.F., Short, J.C. and Payne, G.T. (2011b) 'Family Business and Market Orientation: Construct Validation and Comparative Analysis', *Family Business Review* 24 (3): 233–51.

—— , McKenny, A.F., Short, J.C., Davis, K.M. and Wu, D. (2011a) 'Franchise branding: an organizational identity perspective', *Journal of the Academy of Marketing Science* 39 (4): 629–45.

Zavyalova, A., Pfarrer, M., Reger, R. and Shapiro, D. (2012) 'Managing the message: the effects of firm actions and industry spillovers on media coverage following wrongdoing', *Academy of Management Journal* 55 (5): 1079–101.

Online resources

www.businessplanarchive.org/ (accessed 1 May 2014).
www.dotcomarchive.org/ (accessed 1 May 2014).

Notes

1   The Edgar website lists these documents based on their SEC-specified document type, rather than by category. Proxies are listed as DEF-14As, annual reports are listed at 10-Ks and IPO prospectuses are listed as S-1s.