Article

Strategic Organization
2024, Vol. 22(2) 269–296

DOI: 10.1177/14761270241231724
journals.sagepub.com/home/soq
Sage

The Long March: The quest for valid text-based indicators of exploration and exploitation

Nazlihan Ugur
Nyenrode Business University, The Netherlands; KU Leuven, Belgium

Rene Belderbos
KU Leuven, Belgium; Maastricht University, The Netherlands; UNU-MERIT, The Netherlands

Stijn Kelchtermans
KU Leuven, Belgium

Bart Leten
KU Leuven, Belgium; University of Hasselt, Belgium

Abstract
Since March outlined the importance of balancing exploration and exploitation in organizational learning, the exploration–exploitation paradigm has received substantial attention in the management literature. Recent studies have used computer-aided text analysis to construct measures of firms' inclination towards exploration or exploitation, using the original set of keywords proposed by James G. March. We propose a structured series of tests to assess the validity of computer-aided text analysis-based measures and demonstrate that the approach used in prior studies is unlikely to deliver valid indicators. We demonstrate that an alternative approach, which relies on a larger library of keywords, including synonyms of the March keywords and selects only those keywords that are informative and that pass validity tests, delivers valid computer-aided text analysis indicators – both for unstructured (news articles) and structured text-bases (annual reports). Our study contributes to the literature on construct validity and has broader implications for the development of computer-aided text analysis-based indicators in strategy and organization research.

Keywords
ambidexterity, computer-aided text analysis, content analysis, exploitation, exploration, firm performance

Corresponding author:
Nazlihan Ugur, Nyenrode Business University, Straatweg 25, 3620 AC Breukelen, The Netherlands.
Email: n.ugur@nyenrode.nl

Introduction

Inspired by the work of March (1991), an expanding literature has examined the question of how firms should balance exploration and exploitation in organizational learning and how such a balance affects firm performance (e.g. Bauer and Leker, 2013; Belderbos et al., 2010; Eisenhardt et al., 2010; Gatti et al., 2015; He and Wong, 2004; Jancenelle, 2020; Jansen et al., 2009; Jansen et al., 2006; Knott, 2002; Lavie et al., 2011; Lin et al., 2007; Luo et al., 2019; McKenny et al., 2018; Nosella et al., 2012; Stettner and Lavie, 2014; Tushman and O'Reilly, 1996; Tushman et al., 2011; Uotila et al., 2009; Vagnani, 2015; Van Looy et al., 2005; Walrave et al., 2011). Following a pioneering study by Uotila et al. (2009), several studies have used content analysis, and, in particular, computer-aided text analysis (CATA) of texts describing firms' activities (e.g. annual reports or news articles) to capture firms' endeavours in exploitation and exploration. Most studies that applied CATA to measure firms' exploration and/or exploitation activities used the terms proposed by March (1991) to define the concepts of exploration and exploitation, in some cases expanded with additional keywords, permutations or synonyms (Allison et al., 2014; Gatti et al., 2015; Heyden et al., 2015; Jancenelle, 2020; Kimbrough et al., 2014; Luger et al., 2018; Luo et al., 2019; McKenny et al., 2018; Moss et al., 2014; Oehmichen et al., 2017; Titus et al., 2017; Uotila et al., 2009; Vagnani, 2015; Walrave et al., 2011).

CATA has the advantage that it can deliver measures that (1) cover a broad range of firm activities, (2) can be calculated for a large number of firms across long time periods and (3) are applicable across different industries. Furthermore, CATA allows drawing on large amounts of text data and promises higher reliability than human coding (Neuendorf, 2002), hence holds the promise of generating broadly applicable indicators. In this article, we argue that the value of this approach in strategy and organization research in general and in research on exploration and exploitation, in particular, will crucially depend on the validity of CATA measures. However, construct validity tests have rarely been systematically applied in the literature.

In this article, we use a systematic validity protocol for CATA measures following Belderbos et al. (2017) and Short et al. (2010) to assess the validity of the dominant approach in the ambidexterity literature. We demonstrate that indicators based on the original March keywords are unlikely to pass these established validity tests and hence cannot be considered proper measures of exploration or exploitation. The March keyword-based measures appear to cover broader concepts and are not able to distinctively identify exploration and exploitation activities. The original contribution by March provides no specific foundations for the terms and it is fair to assume that he did not intend his lists to be used 'as is' for the construction of empirical indicators. The main reason for the use of the original keywords proposed by March appears to be their availability rather than their validity. We propose an alternative 'extended-informative March' CATA approach that utilizes a much larger library of keywords–including synonyms of March keywords – but that subsequently selects only those keywords that are informative (frequency of occurrence) and that pass validity tests. We demonstrate that this approach delivers valid CATA indicators both for unstructured (news articles) and structured (annual reports) text-bases.

The overall message of our analysis consists of both an endorsement and a warning for the use of CATA indicators in strategy and organization research. An endorsement because – as recent research has shown – the data sources for using content analysis are commonly available, easily accessible, and rich in content. This allows for the construction of indicators that are applicable across sectors and time, a feature that existing indicators typically cannot provide. However, the use of CATA indicators is also associated with a number of caveats. Text data are noisy, and indicators should be constructed using an informative and valid set of keywords. We recommend a structural approach to test the validity of CATA indicators as indispensable to safely unleash the value of text data and content analysis. Keywords should be curated for each unique text corpus to develop a CATA-based indicator. Only when the indicator passes a range of tests (sampling validity, content validity, discriminant validity, and predictive validity), we can make a judgement concerning construct validity (Edwards, 2011; Krippendorff, 2013). Our article responds to the call by Boyd et al. (2013) to give more attention to the measurement and validity of constructs in management studies.

In the next section, we provide the background of our research by reviewing prior studies on exploration, exploitation, firm performance, and existing literature using CATA in ambidexterity research. Subsequently, we describe a set of validity testing steps to be taken to construct valid indicators. We then discuss our data and methods, and report the findings when applying the validity steps to an indicator based on the March (1991) keywords for unstructured text (news articles). After demonstrating that this approach fails to pass all validity tests, we then turn to our approach based on a larger library of keywords that includes synonyms and the selection of informative and valid keywords from among these. We show that such an approach can lead to a valid exploration orientation indicator that has the expected curvilinear relation with firm performance. In supplementary analysis, we show that comparable results are obtained if we use structured text data in the form of annual reports. Finally, we discuss the implications of our findings and conclude.

Theoretical background

Exploration, exploitation and ambidexterity

Since March's (1991) seminal study, the terms exploration and exploitation have dominated debates on organizational learning (e.g. Levinthal and March, 1993), strategic management (e.g. Gatti et al., 2015; He and Wong, 2004; Heyden et al., 2015; Luger et al., 2018; Titus et al., 2017; O'Reilly et al., 2009; Uotila et al., 2009) and innovation management (e.g. Arora et al., 2020; Belderbos et al., 2010; Jancenelle, 2020; Jansen et al., 2006). Exploitation refers to the improvement and extension of existing capabilities by means of activities, such as standardization, upscaling and refinement, leading to predictable returns that are proximate in time. Exploration refers to the creation of new capabilities by means of activities, such as fundamental research, experimentation and search, implying uncertain returns that are more distant in time (March, 1991). More specifically, March (1991: 72) defines exploration as 'gaining new information about alternatives', and exploitation as 'using the information currently available'.

March viewed exploration and exploitation as two types of organizational learning that compete for scarce organizational resources. The central tenet of March's work is that firms need to be ambidextrous and maintain a balance between exploitation and exploration to prosper in both the short and long run. Organizations that engage only in exploration are likely to end up with too many undeveloped ideas and too little distinctive competencies (Levinthal and March, 1993). Conversely, firms that focus exclusively on exploitation might end up in competency traps (Levinthal and March, 1993; Levitt and March, 1988) and might experience that such core capabilities (Prahalad and Hamel, 1990) turn into core rigidities over time (Leonard-Barton, 1992; Walrave et al., 2011).

While balancing is crucial, pursuing exploration and exploitation simultaneously in an organization creates important tensions (Nosella et al., 2012). Both activities are competing for scarce resources, they have distinct objectives, and they require different mindsets and organizational routines, with flexible and organic structures being preferable for exploration purposes, and efficiency-oriented and mechanistic organizational practices being better suited for exploitation activities (Abernathy, 1991; Benner and Tushman, 2003; Burns and Stalker, 1961; Ghemawat et al., 1993; Jansen et al., 2006; Stettner and Lavie, 2014; Titus et al., 2017 Maclean et al., 2021). Accordingly, March (1991) defines exploration and exploitation as a trade-off, as two ends of a single continuum, such that, ambidexterity comes down to balancing these two ends in a zero-sum game (Chen, 2017; Wilden et al., 2018). This bi-polar view of ambidexterity has been widely adopted in the literature (Ahuja and Lampert, 2001; Ghemawat et al., 1993; He and Wong, 2004; Lavie and Rosenkopf, 2006; Raisch and Zimmermann, 2017) and has been proposed by Lavie et al. (2010) as the most appropriate representation.

The 'balanced' view has been subject to debate. The main criticism is that March (1991) based this trade-off on the assumption that firm resources are limited. Since not all resources are scarce (Grant, 1996), the exploration of new possibilities and the exploitation of existing expertise do not always have to constitute a trade-off. Some information-based resources are not delimited using or sharing but can be multiplied (Moreno-Luzon and Valls-Pasola, 2011). Studies also point out that the requirements for exploration are different from those for exploitation in terms of knowledge, culture, managerial practices and structure, such that, they are not directly competing for the same resources (Benner and Tushman, 2003; Katila and Ahuja, 2002; Kauppila, 2010; Miron-Spektor et al., 2018). Based on such arguments, some researchers conceptualize exploration and exploitation as two independent concepts (Benner and Tushman, 2003; Katila and Ahuja, 2002; O'Reilly and Tushman, 2011; Raisch and Birkinshaw, 2008). In this 'orthogonal' view, exploration and exploitation are two independent and compatible goals for firms. In this article, we focus on organizational learning and follow the dominant view in the field by treating exploration and exploitation as a continuum. We discuss the limitations of this approach in the final section of the article.

Measurement of exploration and exploitation

In the ambidexterity literature, three main approaches are followed: survey-, patent- and text-based measures. Each approach has its advantages and disadvantages regarding its potential for longitudinal analysis, data availability, and indicator validity.

Quite a few studies use surveys to measure exploration and exploitation (He and Wong, 2004; Mom et al., 2009 Koryak et al., 2018; de Visser et al., 2010) and interviews with managers to uncover detailed information about innovation processes (Maclean et al., 2021; Mohnen and Roller, 2005). Survey-based indicators are frequently used in contextual ambidexterity studies at the individual, team, and top management team levels (Koryak et al., 2018; Marabelli et al., 2012; Mihalache et al., 2014; Mom et al., 2009), in structural ambidexterity research, including interorganizational partnerships (Kauppila, 2010; Kiss et al., 2020; Kortmann, 2015; Ungureanu et al., 2020; de Visser et al., 2010) and in temporal ambidexterity studies (Uotila, 2018; Wang et al., 2019). However, survey-based indicators have drawbacks, such as a low level of international and cross-sectoral comparability, subjective judgements and common method bias (e.g. Kleinknecht et al., 2002).

A second empirical approach in ambidexterity research focuses on patent-based indicators (e.g. Belderbos et al., 2010; Choi et al., 2016; Geerts et al., 2018; Morandi Stagni et al., 2021; Quintana-García and Benavides-Velasco, 2008; Rosenkopf and Nerkar, 2001; Russo and Vurro, 2010; Wang and Li, 2008). Katila and Ahuja (2002) and Russo and Vurro (2010) examine citation patterns in patents to distinguish the exploitation of existing knowledge and the exploration of new knowledge. Choi et al. (2016) and Morandi Stagni et al. (2021) use the total number of new versus re-used citations in patent applications. Belderbos et al. (2010) use patent class information to identify exploration and exploitation. Patent-based indicators are easily available and provide longitudinal data but also have disadvantages. For instance, patents only provide a partial view of firms' technology activities since not all inventions are patented, and firms may choose other mechanisms to appropriate the returns on their R&D investments. The importance of patenting also varies by country and sector (Park and Park, 2006).

A third approach to generate exploration and exploitation indicators – and the focus of our study–is content analysis, which can be defined as a 'systematic, replicable approach for compressing words of a text into fewer content categories based on explicit rules of coding' (Weber, 1990). Content analysis holds the promise of generating indicators that have a more general applicability and generalizability. It is generally performed on archival textual data, such as CEO letters to shareholders, annual reports, press releases and news articles, website archives and mission statements (Arora et al., 2020; Duriau et al., 2007; Titus et al., 2017; Uotila et al., 2009). Recent advances in computer technologies have enabled researchers to perform content analyses on large bodies of texts using CATA. CATA can deliver higher reliability than human coding with a lower cost and greater speed (Neuendorf, 2002) and is seen as a valuable tool to process qualitative information, representing an opportunity to derive a more accurate measurement (McKenny et al., 2018). It assists researchers in filtering, categorizing, and processing information by combining the strengths of computer reliability and expert human judgement (Krippendorff, 2004). CATA can be used for texts that are common in an organizational context (e.g. annual reports and press releases), which provides a high internal validity (Aguinis and Edwards, 2014; Dalton and Aguinis, 2013; Short and Palmer, 2008; Short et al., 2010). It allows data collection across a greater set of circumstances (McKenny et al., 2013). Hence, CATA is seen as a promising approach to developing novel and rich measures (Aguinis and Edwards, 2014; Short and Palmer, 2008).

An impressive array of recent studies have performed CATA to derive measures of exploitation and exploration using March's (1991) original keywords[1] (Allison et al., 2014; Gatti et al., 2015; Heyden et al., 2015; Jancenelle, 2020; Kimbrough et al., 2014; Luger et al., 2018; Luo et al., 2019 (McKenny et al., 2018; Moss et al., 2014; Oehmichen et al., 2017; Titus et al., 2017; Uotila et al., 2009; Vagnani, 2015). In some studies, the set of March keywords was extended with additional terms and permutations (e.g. Moss et al., 2014; Walrave et al., 2011) or synonyms (e.g. Heyden et al., 2015; Matthews et al., 2022; Oehmichen et al., 2017).

In line with March's (1991) conceptualization of exploration and exploitation as two learning activities that compete for firm resources, most of the aforementioned studies measure firms' relative orientation towards exploration, as compared with exploitation.[2] Given the aim of this article to examine the validity of the CATA indicators used in past studies, our analysis will stay close to this prior work in terms of conceptualizing and operationalizing exploration and exploitation as points on a continuum. In particular, we follow the approach of the pioneering study of Uotila et al. (2009), of which the operationalization of ambidexterity in the form of firm-level exploration orientation has frequently served as a reference in follow-up research.

Validity in content analysis

A validity protocol for a CATA-based exploration orientation indicator

Validity, the degree that a measure accurately represents the focal construct, is a crucial feature of any empirical study (Cronbach's, 1971). There are different methods and techniques recommended in the literature to assure validity (Cronbach's and Meehl, 1955; Krippendorff, 1980, 2004; Messick, 1989, 1995), and studies have recommended protocols to examine the validity of CATA-based indicators (Belderbos et al., 2017; McKenny et al., 2018; Short et al., 2010). We draw on Short et al. (2010) and Belderbos et al. (2017) to introduce a set of validity checks specifically for a CATA-based exploration orientation indicator. We tailor validity analysis to the development of measures for the multidimensional construct of ambidexterity, and we propose a specific approach to arrive at valid CATA indicators. This structured set of validity tests that allow for developing valid indicators includes sampling validity, content validity, discriminant validity, and predictive validity.

The first step in the validity protocol for a CATA-based exploration indicator is sampling validity, which refers to the:

> the degree to which a collection of data are either statistically representative of a given universe or in some specific respect similar to another sample from the same universe so that the sample can be analysed in place of the universe of interest. (Krippendorff, 1980)

Hence, the body of texts on which content analysis is performed has to be informative about the focal construct. Poor sampling validity may lead to unreliable indicators. To achieve better scalability compared to a manual assessment and to avoid the ex-ante limitation to a single type of document, topic modelling techniques like LDA (latent Dirichlet allocation), a commonly used bottom-up topic modelling algorithm, can be used (Steyvers and Griffiths, 2007). Such techniques help researchers to find 'hidden structures' or 'latent topics' in the text corpus, allowing them to judge which topics are representative of the studied phenomenon (Blei, 2012).

The second validity test we will apply is *content validity*. Content validity refers to 'the extent to which a measure represents all facets of a given construct' (Nunnally and Bernstein, 1994). The dominant approach of content analysis in the management literature uses (a series of) keywords where the underlying assumption is that the words people use in narrative texts, such as news articles, annual reports, press releases or newspaper coverage, are a reflection of their thoughts and of the constructs under study (Maclean et al., 2021; Pennebaker et al., 2003; Short et al., 2010; Tetlock et al., 2008). Studies that performed CATA to build indicators on exploration orientation have used a 'deductive approach' where the keywords are rooted in theory, in particular March's (1991) seminal study (Uotila et al., 2009; Vagnani, 2015). A concern with this approach, which we address in this article, is that it assumes a universal interpretation and use of the March keywords across different settings and bodies of texts.

To assess the content validity of keywords, several complementary tests can be used, such as the agreement between human raters on the construct reflected in a series of documents, the analysis of the context in which keywords occur in documents ('Keyword-in-Context' or KWIC), or the precision, accuracy, and recall performance of a set of keywords in terms of identifying a construct in a document. Interrater agreement (percentage agreement) is a commonly recommended test of the consistency in ratings given by different expert raters, such as whether an article reflects exploration rather than exploitation activities of the firm (Krippendorff, 2013; McKenny et al., 2018). A KWIC analysis examines the relevance of each keyword taking into account the context in which it is used in the body of the text (Gatti et al., 2015; Heyden et al., 2015; Uotila et al., 2009; Vagnani, 2015). Accuracy, precision and recall tests are widely used diagnostics in machine learning and text analysis to assess content validity (Kimbrough et al., 2014). Finally, *face validity* and *semantic validity* – sometimes informally referred to as the 'smell test' – are also used to assess content validity but they are a rather subjective way of assessing content validity (Gravetter and Forzano, 2012).

The third step in our proposed validity testing protocol assesses the *discriminant validity* of keywords, which refers to the degree that the focal construct (identified using the selected keywords) is distinct from other constructs (Campbell and Fiske, 1959). Discriminant validity can be tested using factor analysis, cluster analysis or dimensionality checks that are based on the analysis of correlations both within and across constructs (Gatti et al., 2015; Short et al., 2010). In our empirical setting, discriminant validity means that the keyword-based indicators of exploration and exploitation should be sufficiently distinct from each other and that, for example, the correlations between keywords pertaining to the same construct should be higher than the correlations between keywords across the exploration and exploitation constructs. Note that this validity step is particularly crucial for a multidimensional construct like exploration orientation. In such cases, the term *construct dimensionality* is also used to refer to an indicator's ability to discriminate between the construct's multiple dimensions.

The final step in the validity testing protocol concerns *predictive validity*, which is the degree to which an indicator behaves in relation to an outcome in a manner predicted by theory (Krippendorff, 1980). Short et al. (2010) discuss many assessment techniques that can be used to evaluate predictive validity, including regression analysis and structural equation modelling. In the context of our study, we rely on the body of ambidexterity theory that predicts a balance between exploration and exploitation activities is positively associated with firm value.

While CATA methods have been subject to continuous refinement, prior work that developed text-based exploration and exploitation indicators have used a variety of validity checks without a common, structured approach. Table 1 summarizes the CATA-based articles in exploration–exploitation research, their text data sources, databases, keyword base and validity checks performed.[3] Papers report various, but rarely comprehensive validity checks, including interrater reliability, KWIC analysis, precision and recall tests, face/semantic validity, construct dimensionality, correlational validity, predictive validity, sampling validity and so on. Notably, papers often refer to prior studies to justify the validity of measures rather than actually testing the validity of the indicators used. Only a few articles performed a notable effort to use multiple tools to validate their text-based exploration–exploitation indicators (Moss et al., 2014; Matthews et al., 2022).

Prior studies used different types of text data, such as annual reports, shareholder letters, financial reports, news articles and website archives. Although the data sources display great variation, sampling validity analysis has been performed only rarely (Arora et al., 2020; Jancenelle, 2020). The most commonly reported validity test in terms of content validity is interrater reliability, which prior studies used to assess the extent to which there is agreement between (human or algorithmic) raters about the classification of texts as either exploration- or exploitation-oriented. This step is recommended to avoid potential measurement errors (McKenny et al., 2018), but it constitutes only a partial test of content validity in the absence of complementary tests, such as KWIC and assessments of accuracy, precision and recall. Only few articles also used other forms of validity tests, such as construct dimensionality (discriminant validity) (Gatti et al., 2015). This step is important since unlike constructs like entrepreneurial orientation (Short et al., 2010), global mindset (Belderbos et al., 2017) or long-term orientation (Brigham et al., 2014), ambidexterity is a two-dimensional construct, which needs to be taken into account when assessing whether the keywords of a CATA-based indicator distinctively capture the two dimensions.

The majority of the studies focus solely on March (1991) keywords with permutations (e.g. permutations for the 'search' keyword include 'searching', 'searches', 'research' . . . as in the studies of Allison et al., 2014; Jancenelle, 2020; Luger et al., 2018; Uotila et al., 2009). A few studies broaden the March keyword list by adding synonyms using dictionaries, such as the *Merriam–Webster Thesaurus* (Heyden et al., 2015; Luo et al., 2019), or broaden the keyword list identified in the body of texts through linguistic analysis (Kimbrough et al., 2014; Klonek et al., 2020; Moss et al., 2014).

Finally, we note that we consider exploration orientation a formative measure. Formative measures describe different dimensions of a construct and channel conceptually distinct measures into a single construct (Edwards, 2011). Each keyword (partially) captures a dimension of the construct, such that, the complete set of keywords for a dimension provides redundancy in measurement and therefore provides some resilience against measurement error (Bagozzi, 2011), which is

Table 1. Characteristics of CATA-based research on exploration and exploitation.

Reference	Data source	Database	Keyword base	Sampling validity	Content validity	Discriminant validity	Predictive validity	Other validity checks
Allison et al. (2014)	Shareholder letters	Factiva	March	No	No	No	Yes	No
Arora et al. (2020)	Archived website data	Wayback Machine	R&D vs local (geographic)	LDA	Face validity	No	Yes	No
Gatti et al. (2015)	Annual reports, letters to stockholders and financial reports	Lexis Nexis	March	No	KWIC/interrater agreement	Construct dimensionality	Yes	Temporal stability
Heyden et al. (2015)	Annual reports	UK & EU Pharma Corp Websites	March with synonyms	No	KWIC/face validity	No	No	No
Jancenelle VE (2020)	Annual reports	EDGAR	March	Removal of incomplete records	No	No	Yes	No
Kimbrough et al. (2014)	Not mentioned	Random	March & LIWC	No	Precision and recall	No	Yes	No
Klonek et al. (2020)	Observation	NA	LIWC	No	No	No	Yes	No
Luger et al. (2018)	News articles	Dow Jones index registered firms archived data	March	No	Interrater agreement	No	Correlational validity with other innovation measures	No
Luo et al. (2019)	Annual reports	HeXUN	March with synonyms	No	Interrater agreement	No	Yes	No
Reference	Data source	Database	Keyword base	Sampling validity	Content validity	Discriminant validity	Predictive validity	Other validity checks
Matthews et al. (2022)	Annual reports	UK FTSE 350 firms	March + selective inductive words	No	KWIC	No	Yes	Sensitivity tests
McKenny et al. (2018)	Shareholder letters	S&P500	March	No	Interrater agreement	No	Yes	No
Moss et al. (2014)	10-K reports management section	Family vs non-family business search in business week	LIWC & March	No	Interrater agreement	No	Yes	No
Oehmichen et al. (2017)	Annual reports	UK Pharma Firms	March	No	No	No	No	No
Titus et al. (2017)	10-K reports (CVC, JV, acquisitions)	US Security and Exchange Commission	March	No	No	No	Yes	No
Uotila et al. (2009)	News articles	Factiva	March	No	Interrater agreement/ KWIC/semantic validity	No	Yes	No
Vagnani (2015)	Annual reports, including presidents' letters to stockholders and management's analyses of financial results	Archival data	March	No	KWIC	No	Yes	Temporal stability

Notes: LIWC = linguistic inquiry and word count (software package mainly used for word frequency analysis), KWIC = keyword-in-context, LDA = latent Dirichlet allocation (topic clustering algorithm to identify the informative text-base).

a distinct advantage when working with large volumes of unstructured and 'noisy' data. The factor analysis explicitly assesses the correlations among keywords of a single dimension as well as between the exploration and exploitation dimensions, constituting a verification of internal consistency (Edwards, 2011). The factor analysis also ensures each retained keyword captures a unique part of the variance of its associated dimension. The KWIC analysis further weeds out ambivalent keywords to ensure each retained keyword represents either exploration or exploitation.

Data and methods

The text corpus for our study consists of 129,413 unique news articles[4] from the LexisNexis news archive database[5] for the period 1996–2003. This set of articles covers 161 manufacturing companies, drawn from the 2004 edition of the Industrial R&D Scoreboard, which provides a list of the 500 most R&D - intensive European and 500 most R&D - intensive US and Japanese firms. The firms selected are the top R&D spending firms in their region of origin and industry for which we could retrieve longitudinal financial and economic data. The firms have headquarters in the United States, Europe or Japan. Our sample covers medium to high R&D - intensive industries: chemicals, pharmaceuticals and biotechnology, electronics and electrical engineering, IT hardware (computers and communication equipment) and engineering and non-electrical machinery.

We follow recent studies (Heyden et al., 2015; Lavie et al., 2010; Sidhu et al., 2007; Stettner and Lavie, 2014; Uotila et al., 2009; Volberda et al., 2001) and operationalize the exploration orientation of a firm by dividing the number of March's exploration keywords found in the texts published in a year for a particular firm by the total number of March's exploration and exploitation keywords for the firm in that year (the 'balanced view'). Since the operationalization of exploration orientation is a ratio, an article - length standardization is not required.

To make useful comparisons with prior literature, our specification stays close to prior studies using CATA and to study the performance effects of balanced exploration and exploitation. For the predictive validity tests, we measure firm performance as Tobin's Q, which captures both short-term performance and the firm's long-term prospects of value creation (Allen, 1993; Lavie et al., 2011; Lubatkin and Shrieves, 1986; Uotila et al., 2009). Tobin's Q is defined as the market value of the firm divided by the book value of assets (Bebchuk and Cohen, 2005; Belderbos et al., 2010). In line with prior work (e.g. Uotila et al., 2009), the models of firm performance include the following control variables, taken from Worldscope, Compustat and corporate annual reports: the one-period lagged value of Tobin's Q, the logarithm of firm size (total assets), R&D intensity (R&D expenses divided by total sales) and a set of sector and year dummies. We additionally control for patent intensity (patents divided by R&D expenses) and R&D collaboration. We proxy R&D collaborations by the share of co-owned patents of a firm with external partners (Belderbos et al., 2014). We use patent data from the European patent office and collected patents at the consolidated firm level, taking into account yearly changes in firms' group structures over time. We also include media coverage, defined as the frequency by which firms' business activities are reported on in the press. From research on the relation between media coverage and market value, we infer an inverted U-shaped relationship: initial media coverage creates liquidity and investor recognition (Barber and Odean, 2008; Huberman and Regev, 2001; Nguyen, 2015) while intensive coverage in the media tends to capture episodes of adverse events, referred to as 'bad news bias' in media reporting (Bonsall et al., 2020; Tetlock, 2007).

We estimate system Gaussian mixture model (GMM) to relate firm performance to exploration orientation and past performance (Blundell and Bond, 1998). This approach is suitable for 'small T, large N' panel data that include lagged values of the dependent variable (Arellano, 2003). Furthermore, this dynamic panel data estimation can handle the endogeneity of regressors, unobserved firm- and industry-specific effects, and autocorrelation, using multiple instrumental variables (March and Sutton, 1997; Richard et al., 2009; Roodman, 2009). We treat sector and year dummies as exogenous variables, while we instrument for lagged Tobin's Q, exploration orientation, R&D intensity, patent intensity, R&D collaboration, media coverage and firm size. To avoid instrument proliferation, we limit the number of lags for the instruments to 2 years for predetermined variables. After adding control variables, we can utilize an imbalanced panel data set of 955 observations on 161 firms.

Validity analysis of a CATA indicator of exploration orientation using March's keywords

In this section, we examine the validity of the March (1991) keyword-based indicator of exploration orientation, following the structured set of validity checks described above.

Sampling validity

The principle of sampling validity suggests that uninformative articles have to be removed from the body of text to which content analysis is applied. Besides the primary objective of constructing a relevant text-base – particularly important when working with unstructured data, such as news articles – pruning the data to obtain a valid sample reduces the workload of later validity steps that involve manual coding. In the construction of the keyword-based indicators, we relied on those LexisNexis articles that are informative of firms' exploration and exploitation activities. Due to the wide variety of texts in these unstructured texts in the LexisNexis archive, many of these might not be informative for the constructs.

To arrive at a valid body of texts, we need to identify a set of articles that relate to exploration and/or exploration in organizational learning. Content not focussing on organizational learning, such as routine sales reports, financial statements without guidance or interpretation, mere notifications of management and board changes (in the absence of the reasons for these changes), and similar categories of texts provide no relevant content and are omitted in a sample validity step. For that purpose, we performed an LDA analysis.

To prepare the LexisNexis text corpus for LDA, we performed a set of cleaning and preparation steps. We removed punctuation characters and *pro forma* clauses, such as statements on copyrights. We trained the LDA using 1000 articles and employed $K=2$-10 topics over 1000 iterations with smoothing parameters $\alpha=50/K$ and $\beta=0.01$, respectively (Steyvers and Griffiths, 2007). Our initial analysis showed that $K=7$ is the best fitting number of topics since it provides the most consistent set of keywords for each topic. The labels assigned to the seven topics were confirmed by four independent expert raters. The topic labels as well as the 10 most common keywords for each topic and their word probabilities are shown in Table 2. From a sampling validity perspective, five topics are considered relevant for our analysis of firms' exploration and exploitation activities. Articles that focus on the topics of *financials/sales* or *management and external relations* are omitted. The latter two topics relate to articles without further interpretation covering sales reports, financial reports, notifications of management, and board changes. Articles that have a joint probability greater than 0.5 to fall within these two categories are omitted. This effort to increase the validity of the sample reduced the number of articles from a total of 138,755 articles to 101,041.[6]

Table 2. Sampling Validity: LDA topic clustering results ($K = 7$) with topic labels and 10 most common words for each topic.

	ICT & networks			Design & technology			Drug development			R&D Pharma			Business development			Financials/sales*			Management& external relations *	
	Words	Probability		Words	Probability		Words	Probability		Words	Probability		Words	Probability		Words	Probability		Words	Probability
1	Services	0.0136	1	Technology	0.0106	1	Study	0.0075	1	Research	0.0135	1	Development	0.0101	1	Sales	0.0189	1	Research	0.0077
2	Network	0.0115	2	Digital	0.0083	2	Treatment	0.0112	2	Development	0.0134	2	Products	0.0081	2	Million	0.0166	2	President	0.0057
3	Systems	0.0107	3	Products	0.0071	3	Clinical	0.0082	3	Products	0.0129	3	Production	0.0075	3	Quarter	0.0133	3	University	0.0048
4	Software	0.0099	4	Design	0.0063	4	Study	0.0075	4	Agreement	0.0085	4	Research	0.0066	4	Market	0.0131	4	Development	0.0040
5	Internet	0.0093	5	Systems	0.0058	5	Drugs	0.0071	5	Technology	0.0065	5	Market	0.0065	5	Percent	0.0121	5	Director	0.0036
6	Solutions	0.0093	6	Software	0.0057	6	Disease	0.0058	6	Pharmaceutical	0.0063	6	Plant	0.0065	6	Share	0.0116	6	Million	0.0036
7	Technology	0.0093	7	Market	0.0050	7	Trials	0.0056	7	Statements	0.0063	7	Manufacturing	0.0062	7	Billion	0.0099	7	Programme	0.0033
8	Mobile	0.0085	8	Product	0.0047	8	Cancer	0.0053	8	Announced	0.0054	8	Joint	0.0062	8	Growth	0.0072	8	People	0.0032
9	Service	0.0080	9	Devices	0.0046	9	Phase	0.0051	9	Product	0.0052	9	Sales	0.0062	9	Shares	0.0067	9	Centre	0.0031
10	Wireless	0.0078	10	Memory	0.0046	10	Approval	0.0050	10	Medical	0.0051	10	Venture	0.0061	10	Stock	0.0063	10	Industry	0.0031

*Categories deemed non-informative for the measurement of exploration and exploitation.

Content validity

The content validity of a CATA indicator is assessed on the basis of the accuracy of the keywords used to construct this indicator. To assess content validity, we randomly selected (across firms and years) news articles and two independent human raters classified them as dealing with either exploration or exploitation. We relied on March's (1991: 85) conceptualization of exploitation and exploration as distinct learning activities, where 'the essence of exploitation is the refinement and extension of existing competencies, technologies and paradigms' while 'the essence of exploration is experimentation with new alternatives' and gave the raters clear instructions to take these conceptualizations as their basis for rating. In the context of our study, exploitation captures activities, such as improvements of machinery and manufacturing systems, the extension and routine differentiation of existing product lines and the optimization and updating of work practices. Exploration captures activities like experimentation with and introduction of novel products, technologies, services or processes or entry into new markets.

We used a set of 200 randomly selected articles and asked the raters to classify these as either exploitative or explorative. Classifying articles as distinctively exploration- or exploitation-related was feasible since most articles referred to specific business events, such as a market expansion or a new product introduction. In the rare case that an article did not conclusively report on either exploration or exploitation, it was discarded, and a new article was drawn. We continued to draw articles until a set of 100 exploitation and 100 exploration articles was established. This set of 200 articles was roughly evenly distributed across sectors and years and fairly representative of the sample of firms. The interrater reliability (percentage agreement) was equal to 0.76 (corresponding to a Kappa of 0.56), which is commonly considered an intermediate to good agreement, as it is above the generally accepted threshold of 0.75 (Fleiss et al 2003).[7] We note that researchers can make other decisions on the implementation of the rating scheme. We limited our categorization to a binary one since manual categorization of articles is a complex task and a simple but clear dichotomous approach is less prone to error than a complex Likert - type scoring. Given that articles can contain both explorative and exploitative contents, a Likert rating would also have been conceivable. Similarly, an alternative approach to discarding ambiguous articles would be to classify undetermined cases both as explorative and exploitative.

Subsequently, we examined the *accuracy*, *precision* and *recall* of the CATA text-based indicator, using the manually classified interrater-agreed 152 articles as a benchmark. An article was classified as 'exploration' when it contained more exploration keywords than exploitation keywords and as exploitation otherwise. Table 3 shows the classification of the articles by the human raters and the text-based indicator as well as the classification performance indicators. *Accuracy* is defined as the share of articles that are classified correctly by the CATA text-based indicator (i.e. in agreement with the two raters) as exploration (and hence also as exploitation). The *precision* and *recall* indicators are specific to exploration and exploitation. For exploration, *precision* tells us, for all articles classified as exploration by the CATA text-based indicator, what share was also identified as such by the human raters (and analogous for exploitation). *Recall* tells us, for all articles classified as exploration by the human raters, what share was classified correctly using the March text-based indicator (and again analogous for exploitation).The resulting values for *accuracy* (49%), *precision* (56% for exploration and 31% for exploitation) and *recall* (70% for exploration and 19% for exploitation) suggest that content validity of the March text-based indicators is rather poor. This finding is in line with McKenny et al.'s (2018) comparison of manual and CATA encoding of ambidexterity using the March keywords on a data set of Management Discussion and Analysis sections of 10K filings of 205 firms.

Table 3. Classification of articles by human raters versus CATA March and extended-informative March indicator.

		Classification by human raters	
		Exploration	Exploitation
Classification by *March* CATA indicator	Exploration	63	50
	Exploitation	27	12
Accuracy	TP + TN / TP + FP + FN + TN	(63 + 12) / (63 + 50 + 27 + 12) 49%	(12 + 63) / (12 + 27 + 50 + 63) 49%
Precision (P)	TP/TP + FP	63/(63 + 50) 56%	12/(12 + 27) 31%
Recall (R)	TP/TP + FN	63/(63 + 27) 70%	12/(12 + 50) 19%
F1	2×((P × R)/(P + R))	0.62	0.24
		Classification by human raters	
		Exploration	Exploitation
Classification by *extended-informative March* CATA indicator	Exploration	68	39
	Exploitation	22	23
Accuracy	TP + TN / TP + FP + FN + TN	(68 + 23) / (68 + 39 + 22 + 23) 60%	(12 + 63) / (23 + 22 + 39 + 68) 60%
Precision (P)	TP/TP + FP	68/(68 + 39) 64%	23/(22 + 23) 51%
Recall (R)	TP/TP + FN	68/(68 + 22) 76%	23/(23 + 39) 37%
F1	2 × ((P × R)/(P + R))	0.69	0.43

Notes: TP = true positive; TN = true negative; FP = false positive; FN = false negative. The F1 statistic combines precision and recall in a single performance measure.

Finally, a KWIC analysis (Manning and Schutze, 1999) was used to examine whether keywords had an unambiguous meaning across contexts (Krippendorff, 2004). Details are provided in Table A.1, which contains examples of the March keywords and the extended-informative March keywords (see below) where these keywords are used in an ambiguous or in a correct context. Twenty random occurrences of each March keyword and their 100 surrounding contextual words in the news articles were manually inspected using NVIVO software to examine whether they were used in a relevant context. The KWIC scores for the March keywords are presented in Table 4. The ambiguity scores indicate that several exploration and exploitation keywords are often used out of their context and do not represent exploration or exploitation activities accurately. This pattern is most pronounced for the exploration keyword *play* and the exploitation keyword *execution* that are in more than half of the cases not used in the correct intended context, reflected in ambiguity scores that exceed 0.5. Table 4 also presents that the frequency of occurrence of some of the March keywords is rather low in the text corpus, with the keywords *variation and refine* represented in less than 1% of the texts.

Discriminant validity

To test the discriminant validity of the 17 keywords of March (1991), we performed a factor analysis to identify those keywords that are associated with the two underlying constructs of exploration

Table 4. Content validity: KWIC analysis and ambiguity scores.

Keywords	Search string	Total number of hits	Ambiguity score
March keywords			
Choice	Choice	5346	0.05
Efficiency	Efficien	13,435	0.25
Execution	Execut	24,278	0.8
Exploitation	Exploit	1386	0.05
Implementation	Implemen	12,936	0.1
Production	Production	28,825	0.05
Refine	Refine	948	0.05
Selection	Select	13,975	0.15
Discovery	*Discover*	22,255	0.1
Experimentation	*Experiment*	2682	0
Exploration	*Explor*	3222	0.25
Flexibility	*Flexib*	97,96	0.15
Innovation	*Innovat*	26405	0
Risk taking	*Risk*	18,234	0.3
Search	*Search*	95,846	0.15
Play	*Play*	34,345	0.6
Variation	*Variation*	699	0.5
Extended-informative March keywords			
Advanced	Advance	26,761	0.05
Control	Control	30,316	0.1
Enhance	Enhanc	19,100	0
Improve	Improv	25,211	0.05
Collaboration	*Collaborat*	18,237	0
Discover	*Discover*	22,255	0.1
Search	*Search*	95,846	0.15

Notes: Exploration keywords are represented in *italics*. The ambiguity score is calculated as the percentage of cases in which a keyword is used in an ambiguous context. When retrieving keywords, prefixes or postfixes (e.g. eco-innovation, refinement) were also included.

and exploitation (Short et al., 2010). The Kaiser–Meyer–Olkin measure of sampling adequacy (0.8, well above the commonly used benchmark value of 0.5) indicates that our sample lends itself to factor analysis (Field, 2009). The factor solution for the March keywords is shown in Table 5 and indicates that March's keywords load into no fewer than six different factors. Moreover, most underlying factors combine both exploration and exploitation keywords. For example, factor two includes the exploration keyword *flexible* together with the exploitation keywords *efficien and implement*. The conclusion is that the exploration and exploitation constructs are not captured distinctively with the March sets of keywords.

Predictive validity

In the final validity step, we examined how the March keywords, and the indicator derived from them, behave in an analysis of the relationship between exploration orientation and firm performance, measured by (the logarithm of) Tobin's Q. Based on the 'balanced view' on ambidexterity, we expect an inverted U-shaped relationship between exploration orientation and firm performance. Table 6 shows the descriptive statistics and correlations and Table 7 shows the empirical

Table 5. Discriminant validity: Factor analysis for March's keywords.

Eigenvalues	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Uniqueness
	(1.60)	(1.45)	(1.16)	(1.05)	(1.03)	(1.00)	
Exploration							
Discover	0.7315						0.4531
Experiment					0.4006		0.6723
Explor						0.6650	0.4702
Flexible		0.5176					0.6594
Innovate							0.6828
Play			0.7792				0.3660
Risk	0.4263						0.7304
Search	0.7186						0.4746
Variation					0.6132		0.5853
Exploitation							
Choice							0.6159
Efficien		0.5966					0.5617
Execute						−0.4640	0.5019
Exploit					−0.5025		0.6162
Implement		0.6586					0.5399
Production				0.6865			0.4491
Refine				0.4175			0.6648
Select							0.6442

Notes: Blanks represent loadings lower than the 0.4 threshold; varimax rotation applied.

Table 6. Descriptive statistics for the predicted validity test.

	M	*SD*	1	2	3	4	5	6	7	8
1. Log (Tobin's Q)	0.30	0.78	1.000							
2. Lagged log 'Tobin's Q)	0.31	0.79	0.830	1.000						
3. Log (employees)	10.01	1.33	−0.226	−0.226	1.000					
4. R&D intensity	0.07	0.05	0.404	0.385	−0.359	1.000				
5. Patent propensity	0.02	0.03	−0.128	−0.134	0.098	−0.321	1.000			
6. R&D collaboration	0.05	0.08	−0.084	−0.076	−0.167	0.018	−0.089	1.000		
7. Media coverage	0.21	0.08	0.204	0.143	−0.003	0.183	−0.110	−0.117	1.000	
8. Exploration Orientation (March)	0.59	0.20	0.250	0.249	−0.070	0.329	−0.156	0.038	0.018	1.000
9. Exploration orientation (E-I March)	0.51	0.26	−0.106	−0.121	0.087	−0.171	0.115	−0.174	0.246	−0.540

Notes: N = 995. E-I March stands for extended-informative March

results. The results for the control variables indicate significance for the lagged value of Tobin's Q, R&D intensity, and R&D collaboration and indicate an inverted U-shaped association with media coverage (in model 2). The negative sign for R&D collaboration measured as co-patenting is consistent with prior research (Belderbos et al., 2014) and related to the shared claims on the value of patented inventions. The results reported in model 2 in Table 7 do not reveal a statistically significant relationship between exploration orientation and firm performance for the March keyword-based indicator: the estimates of $\beta = -0.957$ for the linear coefficient and $\beta = 0.737$ for the squared coefficient are both insignificant. We conclude that the March keyword-based indicator does not pass the predictive validity test. Given the lack of validity demonstrated by the previous tests, the

Table 7. Predictive validity – the relationship between Tobin's Q and exploration orientation.

	Model 1	*Model 2*	*Model 3*
	Controls only	Exploration orientation with March keywords	Exploration orientation with extended-informative March keywords
Exploration orientation		-0.957	1.472**
		(1.323)	(0.646)
Exploration orientation²		0.737	-1.161**
		(1.313)	(0.524)
Log (Tobin's Q_{t-1})	0.604***	0.590***	0.547***
	(0.0713)	(0.0960)	(0.0791)
Log (employees)	0.0411	0.0619	0.0168
	(0.0380)	(0.0548)	(0.0502)
R&D intensity	2.324***	1.662	1.882*
	(0.828)	(1.139)	(1.016)
Patent propensity	-0.199	3.393	-0.664
	(2.061)	*(2.333)*	(2.702)
R&D collaboration	-0.459***	*-0.921***	-1.136**
	(0.159)	(0.399)	*(0.528)*
Media coverage	-0.156	2.996***	0.983
	(0.884)	(1.135)	(1.085)
Media coverage²	0.503	-3.653**	-1.049
	(1.290)	(1.538)	(1.466)
Year dummies	*Included*	*Included*	*Included*
Sector dummies	*Included*	*Included*	*Included*
Constant	0.000	0.000	0.000
	Omitted	Omitted	Omitted
Observations	995	995	995
Wald χ^2 (*df*)	1126.93	1490.25	1128.15
	(0.000)	(0.000)	(0.000)
AR2	−0.77	−0.87	−0.82
No. of instruments	98	107	107
Hansen J (prob $> \chi^2$)	0.052	0.155	0.222
Hansen *J*-test	97.14	96.06	92.52

Notes: Media coverage is scaled by a factor 1000. R&D collaboration is scaled by a factor 100. Robust standard errors in parentheses. ***$p < 0.01$, ** $p < 0.05$, * $p < 0.1$. AR2 = Arellano–Bond AR(2) test for autocorrelation.

likely reason is that an exploration orientation indicator based on the March keywords fails to accurately capture the underlying phenomena of exploration and exploitation.

Towards a solution: An extended-informative March indicator

The difficulty with the March keyword approach is twofold: the keywords are not necessarily informative for the text corpus at hand, with few occurrences, and the keywords may be used out of context and do not allow grouping them into two factors for a multidimensional indicator, such as exploration orientation. In this section, we present an alternative approach to construct a valid CATA indicator of exploration orientation that addresses these issues. The logic of this approach is to extend the number of keywords but at the same time maintain a strict adherence to validity. We label this approach as the 'extended-informative' March approach, since we rely on an 'extended' list of March keywords but only select those keywords that are 'informative' in the sense that they

Table 8. Factor analysis for the extended-informative March keywords.
Panel A: Factor analysis on 20 keywords.

Eigenvalues	Total number of hits	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
		(1.76)	(1.71)	(1.70)	(1.66)	(1.47)	
Exploration							
New	181,157						0.6186
Search	95,846	0.662					0.5296
Chang	30,659			0.5219			0.6305
Innovat	26,405		0.4998				0.6994
Discover	22,255	0.7897					0.3533
Release	20,969			0.709			0.4869
Expand	20,398				0.5456		0.6662
Patent	19,121						0.836
Collaboration	18,237	0.7434					0.4394
Acquisition	12,206				0.5646		0.6467
Exploitation							
Grow	44,227				0.7085		0.4701
Standard	34,710						0.7253
Continue	34,704				0.5188		0.5987
Control	30,316		0.4271				0.754
Advanced	26,761		0.4891				0.7264
Improve	25,211		0.4826				0.6999
Speed	23,407					0.7885	0.3745
Fast	20,469					0.7635	0.3941
Enhance	19,100		0.5851				0.6475
Certain	17,787			0.7423			0.3882

Panel B: Factor analysis on seven keywords.

Eigenvalues	Factor 1	Factor 2	Uniqueness
	(1.76)	(1.71)	
Exploration			
Collaboration	0.7563		0.4280
Discover	0.8111		0.3415
Search	0.7563		0.5243
Exploitation			
Advanced		0.5652	0.6802
Control		0.5456	0.6992
Enhance		0.6214	0.6124
Improve		0.6029	0.6245

Notes: Blank cells represent factor loadings lower than 0.4. A varimax rotation was applied.

occur frequently in a text-base and pass the validity tests. We use the extended list of 211 (126 exploitation + 85 exploration) keywords developed in Matthews et al. (2022), who add a series of synonyms based on the Merriam-Webster Thesaurus and add a number of inductive keywords.[8] In our approach, we select only those keywords that are most frequently occurring in the text sample of news articles for our sample firms. We limit ourselves to the 10 most frequent exploration and exploitation keywords to illustrate the approach and stay close to the number of 17 original March keywords. The list of 20 keywords–together with the number of times they occur in the text sample – are presented in Table 8. In contrast with some of the original March keywords, all the 20 keywords occur frequently in the text corpus, with the exploration keyword *new* occurring the most frequently (181,157 times).

The next step of our approach consists of a factor analysis to identify those keywords that are associated with the two underlying constructs of exploration and exploitation.[9] Table 8 (Panel A) shows that the keywords load into five different factors. The first factor contains the exploration keywords *search*, *discover*, and *collaboration*. The second factor combines four exploitation keywords (*improve*, *enhance*, *advanced* and *control*) and one exploration keyword (*innovat*). Since our objective is to establish a dictionary of keywords in which each keyword is unambiguously linked to either exploration or exploitation, we reduce the factor solution to two separate dimensions by removing exploration (exploitation) keywords that are not grouped with other exploration (exploitation) keywords. Hence, we dropped the keyword *innovat*. Conducting a second factor analysis with the seven keywords, we obtained a meaningful two-factor solution whereby each factor consists exclusively of exploration or exploitation keywords that are informative for the text-base at hand (Table 8, panel B).

We subsequently subjected the extended-informative March indicator to content and predictive validity tests. The content validity tests are reported in Tables 3 and 4, alongside the results for the standard set of March keywords. The extended-informative March keywords all have low KWIC ambiguity scores, ranging between 0 and 0.15, suggesting that the keywords are most often used in the right context. The extended-informative March CATA indicator also improves significantly over the March CATA indicator in terms of accuracy, precision, and recall. Accuracy improves from 49% to 60%.

Finally, the extended-informative March CATA indicator also passes the predictive validity test. As shown in model 3 of Table 7, we find a significant inverted U-shaped relationship between exploration orientation and firm performance, confirming the ‘balanced view’ of ambidexterity, with estimates of $\beta = 1.472$ for the linear coefficient and $\beta = -1.161$ for the squared coefficient. Overall, the extended-informative March CATA indicator passes all validity steps in our suggested approach and can be considered a valid indicator of a firm’s exploration orientation.

Supplementary analysis

We have shown that the use of unstructured text creates challenges for sampling validity, such that, researchers need to ensure that they work with an informative text-base. Structured text-bases, such as company annual reports or letters to shareholders, present fewer challenges in this regard. We examine whether the March keyword approach and our suggested approach also work in such a context. The results of our analyses are reported in a separate online appendix. We focused the analysis on the US companies in our sample, as for these we were able to systematically collect their 10-K reports through EDGAR, the electronic data gathering, analysis and retrieval system for company filings hosted by the US Securities and Exchange Commission; 10-K reports are comprehensive, mandatory reports filed annually by publicly traded US companies that report in detail on their business activities, corporate structure, and financial performance. These reports can be considered an informative source on firms’ exploration and exploitation activities. Our corpus of text of 10-K reports is an unbalanced panel (1996–2003) of 308 observations on 55 US companies.

In constructing the extended-informative March indicator, we started again from the extended keyword list of Matthews et al. (2022). After selecting the 10 exploration and 10 exploitation keywords that occurred most frequently in the 10-K reports, and implementing the iterative factor analysis, we found two meaningful factors. The first factor contained the four exploration keywords *search*, *patent*, *trademark* and *discover*; the second factor was composed of the six exploitation keywords *continu*, *proxim*, *control*, *standard*, *improve* and *existing*. While the extended-informative March indicator passes the discriminant validity test by construction, this was not the case for the March keyword indicator: the 17 March keywords loaded into six different factors that all combined exploration and exploitation keywords (see the online appendix, Table A.1). Moreover, it was not possible to choose an informative subset of March keywords that loaded into two distinct factors.

In the next step, we subjected the March and extended-informative March keywords to a KWIC content validity analysis. This analysis showed that the extended-informative March exploitation keyword *proxim* was never used in the right context, and we removed it from the list of exploitation keywords. Several March keywords scored poorly on the KWIC analyses, with exploitation keywords *execution* and *selection* and exploration keyword *variation* having very high ambiguity scores close to 0.9.

Subsequently, we examined the predictive validity of the March and the extended-informative March exploration orientation indicators. While the March exploration orientation indicator was insignificant in Tobin's Q analysis, we again observed a significant inverted U-shaped relationship between the extended-informative March indicator and firm performance. In sum, the supplementary analyses using a structured text-base of 10-K reports confirmed the findings of the main analyses based on the unstructured text-base of news articles.

Finally, we examined whether the use of the unaltered keyword dictionary of Matthews et al. (2022) could deliver valid indicators of exploitation orientation. This list of keywords also failed the discriminant analysis and predictive validity tests (see the online appendix, Table A.6). Neither the use of the March keywords nor the use of an indiscriminate expanded list of keywords using synonyms, provides valid CATA indicators of exploration orientation. We conclude that to obtain valid CATA indicators, it is important to work with a set of informative keywords that occur sufficiently frequent in the particular corpus of text and that pass the validity tests that we propose in our framework. This set of keywords is context-specific and has to be adapted to the text corpus.

Discussion and conclusion

The primary objective of this study was to examine the validity of CATA indicators of exploration and exploitation and to propose an approach that allows researchers to develop valid indicators. Our study shows that March's original keyword list pertaining to exploration and exploitation is unlikely to deliver valid and generally applicable indicators of firms' exploration orientation derived from large unstructured text-bases, such as news articles, or structured text-bases, such as annual reports. We argue that the derivation of valid and representative CATA indicators of exploration and exploitation should be based on a structured series of validity tests and tailored to the text corpus at hand. We demonstrate that an approach expanding the March keywords with synonyms, but selecting those keywords that are informative in a particular text-base and that pass validity tests for this text-base, can lead to a valid construct of exploration orientation that has the expected curvilinear relationship with firm performance. Moreover, we demonstrate not only that this is the solution when working with structured texts in the form of annual reports but that the approach can also produce a valid indicator when working with unstructured texts (new articles).

Our results indicate that a given set of keywords derived from theory may not necessarily generate accurate indicators. A universal set of keywords to measure exploration and exploitation that is valid for each and every context does not exist. We recommend research utilizing CATA indicators to follow a structured set of validity tests, carefully curated for the data source and operationalization, to arrive at valid and better interpretable indicators specific to the research context. This implies using broader dictionaries and from this larger set of keywords extract those that are informative and prove valid in the formation of the construct. We see this as a prerequisite to exploit the full advantages of text-based indicators, for example, their accessibility, and comparability across sectors, firms and years, and their potential application to a variety of management questions. The approach described in our article offers a tangible and actionable approach to conducting CATA-based research on ambidexterity.

Our study adds to the scarce literature emphasizing the importance of sound methodologies and construct validity in management, which has been concerned with a lack of proper measurement of management constructs (e.g. Boyd et al., 2013; Ketchen et al., 2008). Credible research contributing to cumulative knowledge building requires that measures capture the constructs that they claim to capture (Belderbos et al., 2017; Ketchen et al., 2008). The failure to ensure the validity of measures opens the door to type II errors, such as 'false positives' (Boyd et al., 2013).

Our analysis shows that accounting for idiosyncrasies of the textual data sources is crucial for valid construct measurement and hypothesis testing. Broadly speaking, we can identify three approaches in the literature utilizing text data to deal with the challenge. First, categorizing all articles manually (e.g. Stettner and Lavie, 2014). This has the advantage that it bypasses issues of measurement error in CATA but is generally only applicable to a limited number of texts. Second, limiting the sampling to a homogeneous group of texts, such as annual statements to shareholders in annual reports (e.g. Vagnani, 2015). This will reduce measurement errors but limits the perspective on firm behaviour and may leave a wealth of relevant information untouched. Third, applying CATA to large, unstructured, and heterogeneous text databases, such as news archives. This approach uses the richest information on firms but is most prone to measurement error and potential invalidity of applying identical sets of keywords across contexts. Here, a series of structured validity tests is essential to preserve the richness of the data while limiting measurement error (Belderbos et al., 2017).

First, venturing into new data sources and developing novel techniques and indicators always spawn ideas for further robustness checks and refinements. We highlight here what we believe to be the main caveats of the approach to indicator development described in our study, and how further research might address them. A key data limitation is that we do not observe whether the source of the article can be traced back to the firm, or whether it is produced by external observers such as journalists or analysts. This distinction is of interest since there may be structural differences in their objectivity or more generally the way of reporting about firm activities (Morris, 1994).

Second, while we followed the dominant approach for content analysis in the management literature in the sense that we counted individual keywords (Bligh et al., 2004; Lyon et al., 2000; Uotila et al., 2009; Vagnani, 2015), the text analysis literature offers several ways for refinement. For example, there is an inherent trade-off between precision and recall in automatic text classification: if an indicator is very 'imprecise' and produces more false positives then it will typically yield fewer false negatives, that is, the recall will be higher. This trade-off can possibly be mitigated by the use of weights, such as term frequencies and inverse document frequencies (Liu et al., 2009; Salton and Buckley, 1988). Third, we used the operationalization of exploration orientation as a continuum, as applied in Uotila et al. (2009), which is the dominant approach in the literature (Raisch and Zimmermann, 2017). Future studies could compare different operationalizations, such as orthogonal measures of exploitation and exploration, and examine their predictive validity. In addition, different ways of standardization taking into account (relative) article length (e.g. Allison et al., 2014; Luo et al., 2019) could improve our understanding of valid indicator development.

Finally, we note that we relied on human judgement at several points in our procedure, most notably to classify an initial set of articles, which was then used to validate the March keywords. While we do not think that it is feasible or desirable to eliminate human judgement from the process, data classification techniques offer ways to leverage that human effort. For example, Li (2010) uses Bayesian learning algorithms for categorizing the tone and content of forward-looking statements in firms' 10-Q and 10-K filings. Supervised machine learning approaches start with creating a training set of documents that report on exploration or exploitation activities. Unlike in the keyword-based approach, one can subsequently rely on word co-occurrence algorithms (e.g. latent semantic analysis) or word embeddings (e.g. *word2vec*) to determine the similarity of any document with the documents in the training set based on the document vector representations. Machine learning approaches are a promising avenue for developing indicators based on unstructured text but also come with caveats. First, supervised approaches still require informed raters to generate a training set for the specific text corpus. Koshute et al. (2021) suggest that reliable training sets for machine learning are typically large and should cover more than 3000 data points. Second, some types of documents may not lend themselves well to word embedding algorithms. The use of deep learning algorithms, such as large language models (LLMs) is appropriate when the training set is large since such models have millions of parameters that need to be trained. Still, they may provide poor word embeddings – that is, predictions of the context of a word – if the training set is small or if it consists of short documents.

Conclusion and implications

In this study, our objective is to examine the validity of a commonly used CATA-based indicator to measure exploration and exploitation activities of firms. The majority of CATA-based indicators are built on the narratively defined set of keywords developed by March (1991) without going through a structured series of validity tests. Our literature review shows that past work that has relied on these indicators used few and divergent validity checks but lacked a systematic approach. Our analysis demonstrates that the use of the keywords proposed by March to build CATA indicators has severe validity problems. We believe that research on ambidexterity, strategy and organization studies in general, should adopt a structured validity approach when applying CATA indicators, using broad dictionaries but selecting those keywords that are informative and valid in the context of the text corpus. Hence, the deductive 'theory-based' keywords (such as those proposed by March) are augmented with keywords that are synonyms, or identified bottom-up in the text corpus (Matthews et al., 2022; McKenny et al., 2018; Moss et al., 2014), after which a rigorous selection is made based on the validity steps in our protocol, including KWIC analysis, discriminant validity tests, and precision and recall tests. Our effort is not intended towards recommending a 'one-size-fits-all' approach or keyword dictionary that must be strictly adhered to for every data source and construct, but rather a recommendation for following a structured protocol that encourages researchers to question their indicators. Such an approach could constitute a major impetus for future management research using CATA and for studies on exploration and exploitation in particular.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iDs

Nazlihan Ugur https://orcid.org/0000-0001-7781-2910

Rene Belderbos https://orcid.org/0000-0002-4083-3387
Stijn Kelchtermans https://orcid.org/0000-0002-9977-1926

Supplemental material

Supplemental material for this article is available online.

Notes

1. March's exploration keywords are search, variation, risk, experimentation, play, flexibility, discovery and innovation. The exploitation keywords are refinement, choice, production, efficiency, selection, implementation and execution.
2. Exceptions are Moss et al. (2014), Vagnani (2015), Gatti et al. (2015), Klonek et al. (2020), and Matthews et al, (2022).
3. The following keyword combinations were used to perform the bibliographic search in the Web of Knowledge (1990–2020): CATA and exploration, CATA and ambidexterity, content analysis and exploration, content analysis and ambidexterity, text-mining and exploration, text-mining and ambidexterity.
4. We only include the articles in LexisNexis where the company name is mentioned in the headline of the article, following Uotila et al. (2009). Articles that mention multiple company names are included for each firm in the data set, such that, the number of non-unique articles is 138,755.
5. Key sources for LexisNexis include major disseminators of news releases like Business Wire, PR Newswire Association, Information Bank Abstracts (by *The New York Times*), Information Access Company (Thomson Corp.) and so on, as well as more specialized information brokers like Reuters Health Medical News, Intellectual Property Today, Espicom Business Intelligence (for the pharmaceutical sector), among many others. It is fair to assume that our data offer a comprehensive coverage of publicly known codified information about corporate business activities.
6. We examined whether broadening our text-base by including uninformative articles in the discarded categories influenced our results. While the March CATA indicator was insignificant, the extended-informative March CATA indicator had the predicted inverted U-shape relationship with firm performance.
7. We note that one of the expert raters was an author of the paper, but that it is preferable to assign rating tasks to two raters who are both outside the research team.
8. We note that this is the broadest dictionary in the literature, which is likely to contain a range of keywords with limited validity. If these keywords are less distinctive, our systematic set of steps will identify them as such, and they will not be included in the final keyword list used.
9. We perform the factor analysis before the KWIC analysis, as (potential) reduction in the set of keywords based on the factor analysis reduces the volume of labour - intensive KWIC analysis.

References

Abernathy W (1991) *The Productivity Dilemma*. Baltimore, MD: John Hopkins University Press.
Aguinis H and Edwards JR (2014) Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies* 51(1): 143–174.
Ahuja G and Lampert CM (2001) Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal* 22(6/7): 521–543.
Allen F (1993) Strategic management and financial markets. *Strategic Management Journal, Winter Special Issue* 14: 11–22.
Allison TH, McKenny AF and Short JC (2014) Integrating Time Into Family Business Research: Using Random Coefficient Modeling to Examine Temporal Influences on Family Firm Ambidexterity. *Family Business Review* 27(1): 20–34.
Arellano M (2003) *Panel Data Econometrics*. New York: Oxford University Press.
Arora SL, Li Y, Youtie J, et al. (2020) Measuring dynamic capabilities in new ventures: Exploring strategic change in US green goods manufacturing using website data. *The Journal of Technology Transfer* 45(5): 1451–1480.

Bagozzi RP (2011) Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly*: 261–292.
Barber BM and Odean T (2008) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2): 785–818.
Bauer M and Leker J (2013) Exploration and exploitation in product and process innovation in the chemical industry. *R&D Management* 43(3): 196–212.
Bebchuk L and Cohen A (2020) The costs of entrenched boards. *Journal of Financial Economics* 78: 409–433.
Belderbos R, Cassiman B, Faems D, et al. (2014) Co-ownership of intellectual property: Exploring the value-appropriation and value-creation implications of co-patenting with different partners. *Research Policy* 43(5): 841–852.
Belderbos R, Faems D, Leten B, et al. (2010) Technological activities and their impact on the financial performance of the firm: Exploitation and exploration within and between firms. *Journal of Product Innovation Management* 27(8): 69–882.
Belderbos R, Grabowska M, Kelchtermans S, et al. (2017) On the use of computer-aided text analysis in international business research: Global mindset and firm performance. *Global Strategy Journal* 7(3): 312–331.
Benner MJ and Tushman ML (2003) Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review* 28: 238–256.
Blei DM (2012) Probabilistic topic models. *Communications of the ACM* 55(4): 77.
Bligh MC, Kohles JC and Meindl JR (2004) Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11Th terrorist attacks. *The Leadership Quarterly* 15: 211–239.
Blundell RW and Bond SR (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–143.
Bonsall SB, Green J and Muller KA (2020) Market uncertainty and the importance of media coverage at earnings announcements. *Journal of Accounting and Economics, Elsevier* 69(1).
Boyd BK, Bergh DD, Ireland RD, et al. (2013) Constructs in strategic management. *Organizational Research Methods* 16(1): 3–14.
Brigham KH, Lumpkin GT, Payne GT, et al. (2014) Researching long-term orientation: A validation study and recommendations for future research. *Family Business Review* 27(1): 72–88.
Burns T and Stalker GM (1961) *The Management of Innovation*. London: Tavistock.
Campbell DT and Fiske DW (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56: 81–105.
Chen Y (2017) Dynamic ambidexterity: How innovators manage exploration and exploitation. *Business Horizons* 60: 385–394.
Choi B, Shyam Kumar MV and Zambuto F (2016) Capital structure and innovation trajectory: The role of debt in balancing exploration and exploitation. *Organization Science* 27(5): 1183–1201.
Cronbach LJ (1971) Test validation. In: Thorndike RL (ed.) *Educational measurement*, 2nd edn. Washington, DC: American Council on Education, pp. 443–507.
Cronbach LJ and Meehl PE (1955) Construct validity in psychological tests. *Psychological Bulletin* 52(4): 281–302.
Dalton DR and Aguinis H (2013) Measurement malaise in strategic management studies: The case of corporate governance research. *Organizational Research Methods* 16(1): 88–99.
de Visser M, Faems D, De Weerd-Nederhof P, et al. (2010) Structural ambidexterity in NPD processes: A firm-level assessment of the impact of differentiated structures on innovation performance. *Technovation* 30(5/6): 291–299.
Duriau VJ, Reger RK and Pfarrer MD (2007) A content analysis of the content analysis literature in organizational studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods* 10: 5–34.
Edwards JR (2011) The fallacy of formative measurement. *Organizational Research Methods* 14: 370–388.

Eisenhardt K, Furr N and Bingham C (2010) Microfoundations of performance: Balancing efficiency and flexibility in dynamic environments. *Organization Science* 21(6): 1263–1273.

Field A (2009) *Discovering Statistics Using SPSS*. 3rd edn. Thousand Oaks, CA: Sage.

Fleiss JL, Levin B and Paik MC (2003) Statistical Methods for Rates and Proportions. Hoboken, NJ: Wiley.

Gatti C, Volpe L and Vagnani G (2015) Interdependence among productive activities: Implications for exploration and exploitation. *Journal of Business Research* 68(3): 711–722.

Geerts A, Leten B, Belderbos R, et al. (2018) Does spatial ambidexterity pay off? On the benefits of geographic proximity between technology exploitation and exploration. *Journal of Product Innovation Management* 35(2): 151–163.

Ghemawat P, Ricart I and Costa JE (1993) The organizational tension between static and dynamic efficiency. *Strategic Management Journal, Winter Special Issue* 14: 59–73.

Grant RM (1996) Toward a knowledge-based theory of the firm. *Strategic Management Journal* 17: 109–122.

Gravetter FJ and Forzano LB (2012) *Research Methods for the Behavioral Sciences*. 4th edn. Belmont, CA: Wadsworth, p. 78.

He ZL and Wong PK (2004) Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis. *Organization Science* 15(4): 481–494.

Heyden MLM, Oehmichen J, Nichting S, et al. (2015) Board background heterogeneity and exploration-exploitation: The role of the institutionally adopted board model. *Global Strategy Journal* 5(2): 154–176.

Huberman G and Regev T (2001) Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance* 56: 387–396.

Jancenelle VE (2020) Relative exploration and firm performance: Exploring curvilinear relationships and the role of industry, instability, and munificence. *Long Range Planning* 53(6): 101926.

Jansen JJP, Tempelaar MP, Van Den Bosch FAJ, et al. (2009) Structural differentation and ambidexterity: The mediating role of integration mechanisms. *Organization Science* 20(4): 797–811.

Jansen JJP, Van Den Bosch FAJ and Volberda HW (2006) Exploratory innovation, exploitative innovation, and performance: Effects of organizational antecedents, environmental moderators. *Management Science* 52(11): 1661–1674.

Katila R and Ahuja G (2002) Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal* 45(6): 1183–1194.

Kauppila OP (2010) Creating ambidexterity by integrating and balancing structurally separate interorganizational partnerships. *Strategic Organization* 8(4): 283–312.

Ketchen D, Boyd BK and Bergh DD (2008) Research methodology in strategic management: Past accomplishments and future challenges. *Organizational Research Methods* 11(4): 643–658.

Kimbrough SO, Chou C, Chen YT, et al. (2014) On developing indicators with text analytics: Exploring concept vectors applied to English and Chinese texts. *Information Systems and E-business Management* 12(3): 385–415.

Kiss AN, Libaers D, Barr PS, et al. (2020) CEO cognitive flexibility, information search, and organizational ambidexterity. *Strategic Management Journal* 41(12): 2200–2233.

Kleinknecht A, Van Montfort K and Brouwer E (2002) The non-trivial choice between innovation indicators. *Economics of Innovation and New Technology* 11(2): 109–121.

Klonek FE, Volery T and Parker SK (2020) Managing the paradox: Individual ambidexterity, paradoxical leadership and multitasking in entrepreneurs across firm life cycle stages. *International Small Business Journal* 39(1): 40–63.

Knott AM (2002) Exploration and exploitation as complements. In: Bontis N and Choo CW (eds) *The Strategic Management of Intellectual Capital and Organizational Knowledge: A Collection of Readings*. New York: Oxford University Press, pp. 339–358.

Kortmann S (2015) The mediating role of strategic orientations on the relationship between ambidexterity-oriented decisions and innovative ambidexterity. *Journal of Product Innovation Management* 32(5): 666–684.

Koryak O, Lockett A, Hayton J, et al. (2018) Disentangling the antecedents of ambidexterity: Exploration and exploitation. *Research Policy* 47(2): 413–427.

Koshute P, Zook J and McCulloh I (2021) Recommending training set sizes for classification. *Arxiv Preprint*. Available at: https://arxiv.org/abs/2102.09382.

Krippendorff K (1980) *Validity in Content Analysis*. In: Mochmann E (ed.) *Computerstrategien fÃ¼r die kommunikationsanalyse*. Frankfurt/New York: Campus, pp. 69–112.

Krippendorff K (2004) *Content Analysis: An Introduction to Its Methodology*, 2nd edn. Thousand Oaks, CA: Sage.

Krippendorff K (2013) *Content Analysis: An Introduction to Its Methodology*, 3rd edn. Thousand Oaks, CA: Sage.

Lavie D and Rosenkopf L (2006) Balancing exploration and exploitation in alliance formation. *Academy of Management Journal* 49: 797–818.

Lavie D, Kang J and Rosenkopf L (2011) Balance within and across domains: The performance implications of exploration and exploitation in alliances. *Organization Science* 22(6): 1517–1538.

Lavie D, Stettner U and Tushman ML (2010) Exploration and exploitation within and across organizations. *The Academy of Management Annals* 4(1): 109–155.

Leonard-Barton D (1992) Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal* 13: 111–125.

Levinthal D and March JG (1993) The myopia of learning. *Strategic Management Journal* 14: 95–112.

Levitt B and March JG (1988) Organizational Learning. *Annual Review of Sociology* 14: 319–340.

Li F (2010) The information content of forward-looking statements in corporate filings – A Naïve Bayesian machine learning approach. *Journal of Accounting Research* 48: 1049–1102.

Lin Z, Yang H and Demirkan I (2007) The performance consequences of ambidexterity in strategic alliance formations: Empirical investigation and computational theorizing. *Management Science* 53(10): 1645–1658.

Liu Y, Loh HT and Sun A (2009) Imbalanced text classification: A term weighting approach. *Expert Systems with Applications* 36: 690–701.

Lubatkin M and Shrieves RE (1986) Toward reconciliation of market performance measures to strategic management research. *Academy of Management Review* 11: 497–512.

Luger J, Raisch S and Schimmer M (2018) Dynamic balancing of exploration and exploitation: The contingent benefits of ambidexterity. *Organization Science* 29(3): 449–470.

Luo CY, Kumar S, Mallick DN, et al. (2019) Impacts of exploration and exploitation on firms performance and the moderating effects of slack: A panel data analysis. *IEEE Transactions on Engineering Management* 66(4): 613–620.

Lyon D, Lumpkin G and Dess G (2000) Enhancing entrepreneurial orientation research: Operationalizing and measuring a key strategic decision making process. *Journal of Management* 26: 1055–1085.

McKenny AF, Aguinis H, Short JC, et al. (2018) What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. *Journal of Management* 44(7): 2909–2933.

McKenny AF, Short JC and Payne GT (2013) Using computer-aided text analysis to elevate constructs: An illustration using psychological capital. *Organizational Research Methods* 16(1): 152–184.

Maclean M, Harvey C, Golant BD, et al. (2021) The role of innovation narratives in accomplishing organizational ambidexterity. *Strategic Organization* 19(4): 693–721.

Manning CD and Schutze H (1999) *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press.

Marabelli M, Frigerio C and Rajola F (2012) Ambidexterity in service organizations: Reference models from the banking industry. *Industry and Innovation – IND INNOV* 19: 109–126.

March JG (1991) Exploration and exploitation in organizational learning. *Organization Science* 2(1): 71–87.

March JG and Sutton RI (1997) Organizational performance as dependent variable. *Organization Science* 8: 698–706.

Matthews L, Heyden ML and Zhou D (2022) Paradoxical transparency? Capital market responses to exploration and exploitation disclosure. *Research Policy* 51(1): 104396.

Messick S (1989) Test Validity: A matter of consequence. *Social Indicators Research* 45(1–3): 35–44.

Messick S (1995) Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice* 14(4): 5–8.

Mihalache O, Jansen J, Van Den Bosch F, et al. (2014) Top management team shared leadership and organizational ambidexterity: A moderated mediation framework. *Strategic Entrepreneurship Journal* 8(10): 128–148.

Miron-Spektor E, Ingram A, Keller J, et al. (2018) Microfoundations of organizational paradox: The problem is how we think about the problem. *Academy of Management Journal* 61(1): 26–45.

Mohnen P and Roller LH (2005) Complementarities in innovation policy. *European Economic Review* 49(6): 1431–1450.

Mom TJ, Van Den Bosch FA and Volberda HW (2009) Understanding variation in managers' ambidexterity: Investigating direct and interaction effects of formal structural and personal coordination mechanisms. *Organization Science* 20(4): 812–828.

Morandi Stagni R, Fosfuri A and Santaló J (2021) A bird in the hand is worth two in the bush: Technology search strategies and competition due to import penetration. *Strategic Management Journal* 42(8): 1516–1544.

Moreno-Luzon MD and Valls-Pasola M (2011) Ambidexterity and total quality management: Towards a research agenda. *Management Decision* 49(6): 927–947.

Morris R (1994) Computerized content analysis in management research: A demonstration of advantages & limitations. *Journal of Management* 20(4): 903–931.

Moss TW, Payne GT and Moore CB (2014) Strategic consistency of exploration and exploitation in family businesses. *Family Business Review* 27(1): 51–71.

Neuendorf KA (2002) *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.

Nosella A, Cantarello S and Filippini R (2012) The intellectual structure of organizational ambidexterity: A bibliographic investigation into the state of the art. *Strategic Organization* 10(4): 450–465.

Nunnally JC and Bernstein IH (1994) *Psychometric Theory*, 3rd edn. New York: McGraw–Hill.

Nguyen BD (2015) Is more news good news? Media coverage of CEOs, firm value, and rent extraction. *Quarterly Journal of Finance* 5(04): 1550020.

O'Reilly III, CA, Harreld JB and Tushman ML (2009) Organizational ambidexterity: IBM and emerging business opportunities. *California Management Review* 51(4): 75–99.

O'Reilly CA and Tushman ML (2011) Organizational ambidexterity in action: How managers explore and exploit. *California Management Review* 53(4): 5–22.

Oehmichen J, Heyden MLM, Georgakakis D, et al. (2017) Boards of directors and organizational ambidexterity in knowledge-intensive firms. *International Journal of Human Resource Management* 28(2): 283–306.

Park G and Park Y (2006) On the measurement of patent stock as knowledge indicators. *Technological Forecasting and Social Change* 73(7): 793–812.

Pennebaker JW, Mehl MR and Niederhoffer KG (2003) Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology* 54: 547–577.

Prahalad CK and Hamel G (1990) The core competence of the corporation. *Harvard Business Review* 68(3): 79–91.

Quintana-García C and Benavides-Velasco CA (2008) Innovative competence, exploration and exploitation: The influence of technological diversification. *Research Policy* 37: 492–507.

Raisch S and Birkinshaw J (2008) Organizational ambidexterity: Antecedents, outcomes, and moderators. *Journal of Management* 34(3): 375–409.

Raisch S and Zimmermann A (2017) A process perspective on the exploration–exploitation paradox. In: Smith WK, Lewis MW and Jarzabkowski P, et al. (eds) *The Oxford Handbook of Organizational Paradox*. Oxford: Oxford University Press, pp. 315–332.

Richard P, Devinney T, Yip G, et al. (2009) Measuring organizational performance: Towards methodological best practice. *Journal of Management* 35: 718–804.

Roodman D (2009) How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal* 9(1): 86–136.

Rosenkopf L and Nerkar A (2001) Beyond local search: Boundary-spanning, exploration, impact in the optical disk industry. *Strategic Management Journal* 22: 287–306.

Russo A and Vurro C (2010) Cross-boundary ambidexterity: Balancing exploration and exploitation in the fuel cell industry. *European Management Review* 7(1): 30–45.

Salton G and Buckley C (1988) Term weighting approaches in automatic text retrieval. *Information Processing and Management* 24(5): 513–523.

Short JC and Palmer TB (2008) The application of DICTION to content analysis research in strategic management. *Organizational Research Methods* 11(4): 727–752.

Short JC, Broberg JC, Cogliser CC, et al. (2010) Construct validation using Computer-Aided Text Analysis (CATA). An Illustration Using Entrepreneurial Orientation. *Organizational Research Methods* 13(2): 320–347.

Sidhu JS, Commandeur HR and Volberda HW (2007) The multifaced nature of exploration and exploitation: Value of supply, demand, and spatial search for innovation. *Organization Science* 18(1): 20–38.

Stettner U and Lavie D (2014) Ambidexterity under scrutiny: Exploration and exploitation via internal organization, alliances, and acquisitions. *Strategic Management Journal* 35(13): 1903–1929.

Steyvers M and Griffiths T (2007) Probabalistic topic models. In: Landauer T, McNamara D and Dennis S, et al. (eds) *Latent Semantic Analysis: A Road to Meaning*. Hillsdale, NJ: Laurence Erlbaum, pp. 77–84.

Tetlock PC (2007) Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62: 1139–1168.

Tetlock PC, Saar-Tsechansky M and Macskassy S (2008) More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63(3): 1437–1467.

Titus V, House JM and Covin JG (2017) The influence of exploration on external corporate venturing activity. *Journal of Management* 43(5): 1609–1630.

Tushman M and O'Reilly C (1996) Ambidextrous organizations: Managing evolutionary, revolutionary change. *California Management Review* 38(4): 8–30.

Tushman M, Smith WK and Binns A (2011) The ambidextrous CEO. *Harvard Business Review* 89(6): 74–80.

Ungureanu P, Bertolotti F, Mattarelli E, et al. (2020) Collaboration and identity formation in strategic interorganizational partnerships: An exploration of swift identity processes. *Strategic Organization* 18(1): 171–211.

Uotila J (2018) Punctuated equilibrium or ambidexterity: Dynamics of incremental and radical organizational change over time. *Industrial and Corporate Change* 27(1): 131–148.

Uotila J, Maula M, Keil T, et al. (2009) Exploration, exploitation, financial performance: Analysis of S&P 500 corporations. *Strategic Management Journal* 30: 221–231.

Vagnani G (2015) Exploration and long-run organizational performance: The moderating role of technological interdependence. *Journal of Management* 41(6): 1651–1676.

Van Looy B, Martens T and Debackere K (2005) Organizing for continuous innovation: On the sustainability of ambidextrous organizations. *Creativity and Innovation Management* 14(3): 208–221.

Volberda H, Baden-Fuller C and Van den Bosch FAJ (2001) Mastering strategic renewal: Mobilizing renewal journeys in multi-unit firms. *Long Range Planning* 34: 159–178.

Walrave B, van Oorschot KE and Romme AGL (2011) Getting trapped in the suppression of exploration: A simulation model. *Journal of Management Studies* 48(8): 1727–1751.

Wang H and Li J (2008) Untangling the effects of overexploration and overexploitation on organizational performance: The moderating role of environmental dynamism. *Journal of Management* 34(5): 925–951.

Wang SL, Luo Y, Maksimov V, et al. (2019) Achieving temporal ambidexterity in new ventures. *Journal of Management Studies* 56(4): 788–822.

Weber RP (1990) *Basic Content Analysis*, 2nd edn. Newbury Park, CA: Sage.

Wilden R, Hohberger J, Devinney TM, et al. (2018) Revisiting James March (1991), whither exploration and exploitation? *Strategic Organization* 16(3): 352–369.

Author Biographies

Nazlihan Ugur is an assistant professor of Strategy, Organization and Leadership at Nyenrode Business University (The Netherlands). Her research focuses on ambidexterity, exploration, innovation strategies and global strategies of multinational, R&D - intensive firms. Her expertise lies within the domain of strategy but

with particular focus on computer-aided text analysis, patent-based research and construct development. Her research has been published in journals, such as the *Global Strategy Journal* and *AOM proceedings*.

René Belderbos is full professor of Strategy at KU Leuven and part-time professor of International Corporate Strategy at Maastricht University and UNU-MERIT (The Netherlands). His research focuses on innovation strategies and international business strategies of multinational firms, including location strategies, global knowledge sourcing, top management teams and internationalization, international R&D collaboration, intellectual property rights and innovation, and knowledge spillovers. He has published widely in journals, such as the *Strategic Management Journal*, *Journal of Management*, *Journal of Management Studies*, *Journal of Product Innovation Management*, *Research Policy*, the *Journal of International Business Studies*, *Global Strategy Journal*, *Journal of Business Venturing*, *Regional Studies* and *Strategic Organization*, among others. His work has received more than 10,000 citations. He has been a consultant to the European Commission, the OECD, the Flemish, Belgian and Dutch governments, and the National Institute of Science and Technology Policy (NISTEP, Japan).

Stijn Kelchtermans is a professor of Strategy & Innovation at KU Leuven and (part-time) at Hasselt University. His research deals with the economics of science and innovation, addressing topics like industry-science links, technology markets and scientific productivity. In addition, he does research and consults on R&D policy, primarily addressing the design and impact of funding schemes. His research has been published in journals, such as the *Journal of International Business Studies*, *Journal of Product Innovation Management*, *Global Strategy Journal*, *The Review of Economics and Statistics*, the *Journal of Applied Econometrics*, the *Journal of Public Economics* and the *Journal of Technology Transfer*.

Bart Leten is full professor of Strategy and Innovation at KU Leuven and Hasselt University. His research focuses on innovation and international business strategies of multinational firms, addressing topics, such as industry-science links, R&D internationalization, intellectual property strategies, ambidexterity, ideation, migration and top management teams. His work has been published in journals, such as the *Journal of International Business Studies*, *Global Strategy Journal*, *Journal of Management*, *Journal of Management Studies*, *California Management Review*, *Journal of Product Innovation Management* and *Research Policy*. His work has received 3000 citations in Google Scholar.