Small Bus Econ
https://doi.org/10.1007/s11187-022-00655-0

The identity of social impact venture capitalists: exploring social linguistic positioning and linguistic distinctiveness through text mining

Laura Toschi · Elisa Ughetto · Andrea Fronzetti Colladon

Accepted: 2 June 2022

Abstract Impact investing is gaining momentum as an investment practice that optimizes both financial and social outcomes. However, the market is still in its emerging stage, and there is ambiguity regarding the definition of players and practices. In this paper, we adopt an investor identity perspective and use a linguistic approach to explore how social impact venture capitalists (SIVCs) communicate their identities and actions to their external stakeholders. Through a text mining analysis of the websites of 195 investors worldwide, our results reveal four types of investors who differ in terms of their social linguistic positioning and linguistic distinctiveness. Finally, by training a tree boosting machine learning model, we assess the extent to which the use of different linguistic styles is associated with website traffic.

Plain English Summary Using the Semantic Brand Score and other innovative methods and tools of text mining and social network analysis, we analyzed the website communication of 195 social impact venture capitalists (SIVCs). We investigate what language reveals about their identity byintroducing two dimensions that measure the strength of their social positioning and the distinctiveness of their language. These two factors are particularly relevant as they can reduce uncertainty and equivocality in the audience's information processing. Our results reveal four types of investors: Smart Heroes (with high linguistic distinctiveness and a solid social positioning), Naïve Dreamers (with a strong social positioning but a language similar to the mass), Illusionists (showing distinctive language but poor social identity), and Blabbers (with low scores on both dimensions). In addition, we use network topic modeling to identify the main discourse themes and discuss the differences among the four groups. Communicating impact investing is most relevant for Smart Heroes and Naïve Dreamers. Illusionists dedicate more attention to the characteristics of their target ventures, management teams, and the environmental impact of their investments, whilesustainable solutions are the main focus of Blabbers. Lastly, we show that the websites of Smart Heroes have more web

All the authors contributed equally to this work.

L. Toschi
Department of Management, University of Bologna, Via Capo di Lucca 34, 40126 Bologna, Italy

E. Ughetto
Politecnico di Torino & Bureau of Entrepreneurial Finance, Corso Duca degli Abruzzi 24, Torino 10129, Italy

A. Fronzetti Colladon (✉)
Department of Engineering, University of Perugia, Via G. Duranti 93, 06125 Perugia, Italy
e-mail: andrea.fronzetticolladon@unipg.it

A. Fronzetti Colladon
Department of Management, Kozminski University, Jagiellonska 57/59, 03-301 Warsaw, Poland

Published online: 11 July 2022

traffic than the other three groups. Our study presents a new approach for analyzing the (online) communication of SIVCs and demonstrates that social enterprises must wisely choose the language with which they communicate their identity. We maintain that managers should be aware of the power of language and carefully elucidate the social impact they seek to achieve.

Keywords Social impact venture capital · Investor identity · Linguistic style · Text mining · Semantic brand score · Semantic network analysis

1 Introduction

Impact investors support social enterprises (Agrawal & Hockerts, 2019; Austin et al., 2006; Barber et al., 2021; Chen & Harrison, 2020; M. Lee et al., 2020; Pache & Santos, 2010)[1] "with the intention to generate positive, measurable social and environmental impact alongside a financial return" (Global Impact Investing Network).[2] Like social enterprises, impact investors have been recognized as hybrid organizations (Cetindamar & Ozkazanc-Pan, 2017; Chen & Harrison, 2020; M. Lee et al., 2020), as they incorporate competing institutional logics (financial versus social-oriented). Given this dual nature, their identity is hard to both define and understand, resulting in mistaken perceptions by external stakeholders (Clegg et al., 2007; Durand & Paolella, 2013). Moreover, the field of impact investing exists at an early stage of development (Calderini et al., 2018; Ormiston et al., 2015), where the definitions of players, practices, and standards remain uncertain and ambiguous. These characteristics have opened a space for the diffusion of "impact washing" occurrences (Harji & Jackson, 2012; Höchstädter & Scheck, 2015), which are deleterious to the field's long-term legitimacy.

The management literature suggests that, under conditions of ambiguity in identity, organizations should engage in extensive communication efforts to improve perceptions of their key attributes among external audiences (Bishop et al., 2019). In this regard, scholars have focused on the use of language as a powerful communication tool that can help organizations "give a sense" of their identity and "make sense" of others' communicated identity (Clarke & Cornelissen, 2011; Gioia & Chittipeddi, 1991; Maitlis & Christianson, 2014; Petkova et al., 2013).[3] By combining identity theory and language theory, we investigate what language reveals about the identity of social hybrid investors operating in an emerging market category. We contribute to a recent stream of research that has started to investigate the relevance of investors' identity (Fisher, 2012; Pontikes, 2012; Smith & Bergman, 2020) and the motives that drive their investments (Allison et al., 2015; Barber et al., 2021; Block et al., 2021; Fisch et al., 2021; Geczy et al., 2021; Vismara, 2016). To our knowledge, prior literature fails to show the role of language as a sensegiving practice for the identity construction of social hybrid investors. In the field of social financing, extant research is confined to the analysis of sensegiving practices adopted by social hybrid ventures (Moss et al., 2018; Parhankangas & Renko, 2017) and the consequent sensemaking endeavors by investors (M. Lee et al., 2020; Miller & Wesley II, 2010; Moss et al., 2018).

Our work contributes in several ways to the extant literature, suggesting that resource exchange dynamics are bidirectional (L. Huang & Knight, 2017) and that "both sides of the coin" (investors and investees) are important for an understanding of resource provision (Smith & Bergman, 2020). First, adopting a view of investors as active players in sensegiving practices allows us to better understand the generation of strong identity relationships between the parties based on mutual meaning, understanding, and a common feeling of belonging, which provide the necessary governance for resource exchange. Second, an investor perspective allows us to unveil the complex nature of the impact investing field

[1] Social ventures deviate from conventional organizational forms focused on the maximization of either financial gain or social welfare and represent valuable vehicles for addressing societal problems (Battilana et al., 2017; Fosfuri et al., 2016; Luo & Kaul, 2019). They face difficulties in attracting funding due to a mismatch between their hybrid nature and the practices of conventional investors who provide capital with the goal of maximizing either financial returns (traditional investors) or social impact (philanthropic investors).

[2] https://thegiin.org/impact-investing/need-to-know/#what-is-impact-investing

[3] *Sensemaking* is the "process through which people work to understand issues or events that are novel, ambiguous, confusing […]" (Maitlis & Christianson, 2014), while *sensegiving* refers to "[attempts] to influence the sensemaking and meaning construction of others toward a preferred redefinition of organizational reality" (Gioia & Chittipeddi, 1991, p. 442).

in its nascent stages, which may undermine the above-mentioned mutual understanding among actors. Indeed, the collective identity of a market comprises principles and practices that reflect the identities of the organizations that belong to the industry (Stigliani & Elsbach, 2018). During the critical phase of industry emergence, organizations act on behalf of the market and thus play a key role in shaping not only their own identities but also the identity and trajectory of the whole industry (Gustafsson et al., 2016).

We focus our investigation on social impact venture capitalists (SIVCs), a particular type of impact investing funding source that is transforming the way social entrepreneurs gain resources (Miller & Wesley II, 2010; Randjelovic et al., 2003). SIVCs represent an optimal setting for our study: The impact investing field is populated by several actors with different institutional arrangements. It is, thus, important to focus on one type of impact investor as "each form's logic will vary rather than reflect a unitary set of logics and practices" (Cetindamar & Ozkazanc-Pan, 2017, p. 258). In particular, SIVCs—which adopt the traditional practice of for-profit VCs to finance mission-driven companies—perfectly represent the hybrid nature at the base of our theorizing that in other types of impact investors (e.g., foundations, charities, or governmental funds) is less pronounced or absent.

Applying text mining techniques[4], we analyze the linguistic style[5] of 195 SIVCs' active websites in order to depict the nature of these social hybrid investors through their use of language. Through an exploratory inductive approach, we identify four types of linguistic styles that distinguish SIVCs in terms of their *social linguistic positioning* and *linguistic distinctiveness*. These two measures are particularly relevant for the study of social hybrid organizations' identity, as they respectively capture how SIVCs "fit in" the category (i.e., by demonstrating a strong commitment toward the core social mission characterizing the market) and "stand out" in the category (i.e., using a language to frame their identity in a way that differs from one of their peers) (Cao et al., 2017; Navis & Glynn, 2011). We propose that both the content of the text (i.e., focus on the social domain) and the style used (i.e., the distinctiveness of the language) jointly contribute to the definition of the social hybrid investor's identity. Finally, as additional analysis, by training a tree boosting machine learning model, we assess the extent to which the use of different linguistic styles is associated with website traffic as a proxy of impact on the external audience's attention.

Our results reveal four types of investors who differ in terms of their social linguistic positioning and linguistic distinctiveness: *Smart Heroes* (with high levels of both social identity and linguistic distinctiveness), *Naïve Dreamers* (with a strong social identity and a language similar to the mass), *Illusionists* (showing distinctive language but poor social identity), and *Blabbers* (with low levels of both social identity and linguistic distinctiveness). We analyze differences in terms of language style, sentiment, readability, and communication intensity among the four groups, finding significant differences for the dimension of sentiment, but not for the others. Moreover, we identify the main discourse topics that emerge on SIVCs' websites, showing differences among the four groups: communicating impact investing is most relevant for *Smart Heroes* and *Naïve Dreamers*; *Illusionists* dedicate more attention to the characteristics of their target ventures, management team, and the environmental impact of their investments; while sustainable solutions are the main focus of *Blabbers*. Finally, we assess the impact of different linguistic styles on web traffic, showing that *Smart Heroes* have significantly more page views than the other three groups.

Our study makes an important contribution to entrepreneurial finance and impact investing in particular. First, we build on the emerging view of investor identity as a precious perspective in resource exchange processes to better understand the nature of an emerging context characterized by high uncertainty and definitional ambiguity, whose dynamics differ from traditional financial contexts. In the consolidated traditional financial domain, rules and practices are well known, and players are well informed. Meanwhile, investors—who are driven by a single goal (i.e., obtaining financial

[4] Text mining techniques have led to empirical breakthroughs when dealing with qualitative information, such as what appears on websites (Archak et al., 2011; Das & Chen, 2007; D. Lee et al., 2018; Netzer et al., 2012). In fact, recent editorials in the field of management have endorsed such empirical methods to advance the study of strategy and managerial decision-making (Arts et al., 2018; Bettis et al., 2015; Choudhury et al., 2019; Kaplan & Vakili, 2015).

[5] Please note that stylistics as a linguistic discipline pays attention to lexicon and other language attributes, such as grammar, semantics, phonology, or syntax (Bradac et al., 1979; Wales, 2006). In this work, we focus exclusively on the lexical component of the linguistic style. Thus, linguistic style is used only to refer to the lexical profiles of the SIVCs' websites.

returns)—are likely to be homogeneous in their communication style, as they conform to the clear prototype of their market category (Czarniawska & Wolff, 1998). Instead, the emerging domain of social impact investing is characterized by high uncertainty and ambiguity in members' identities; thus, there is a greater variety in communication approaches that deserves deeper investigation. Moreover, we suggest that the analysis of investors' sensegiving practices provides researchers with a different perspective regarding the traditional view of sensegiving for new ventures. In the case of new ventures, the goal of sensegiving has generally been linked to the organization's ability to obtain valuable resources for its growth. In the case of investors, sensegiving is associated with the formation of the entire industry's identity, which comes to reflect the identity of the organizations that belong to the market. In addition, in impact investing, investors' engagement in sensegiving practices becomes a key success factor for the effective generation of social value.

Second, our study builds on the importance of language for studying topics of management science. In particular, we are inspired by previous contributions incorporating language in entrepreneurial finance. The majority of existing works on the role of language in new ventures have mainly focused on narratives and stories (Aldrich & Fiol, 1994; Allison et al., 2015; Downing, 2005; Lounsbury & Glynn, 2001; Manning & Bejarano, 2017; Martens et al., 2007; O'Connor, 2002; Wry et al., 2011), or metaphors and linguistic frames (Benford & Snow, 2000; Glaser et al., 2011; Hsu & Hannan, 2005; Navis & Glynn, 2010; Pan et al., 2020), with few exceptions on linguistic style (Moss et al., 2018; Parhankangas & Renko, 2017). Our work departs from these studies and introduces more micro-level linguistic elements related to linguistic styles that are based on "content words"—words with semantic content contributing to the meaning of the message. In doing so, we respond to calls for more research on the nature of organizations through their language, rather than on language through organizations (Boje et al., 2004).

2 The role of organizational identity

2.1 The concept of organizational identity

Organizational identity has become a burgeoning domain of investigation in organization studies (see Gioia et al. (2013) for a review). By moving the concept of identity from an individual level (Baumeister, 1998) to an organizational one (Albert & Whetten, 1985), research has responded to the question: "who are we as an organization?." The foundational paper by Albert and Whetten (1985, p. 265) suggests that organizational identity reflects an organization's self-referential claims with respect to its "central character, distinctiveness, and temporal continuity." Concretely, organizational identity represents the pool of features of an organization that, in the eyes of its members, are relevant and core to describing the organization's "self-image" and making it distinctive from other organizations, and are viewed as having continuity over time (Albert & Whetten, 1985). Scholars largely agree that establishing an organization's identity is critical to building a viable organization because it affects internal self-perceptions and how an organization presents itself to the external environment (Gioia et al., 2013). Accordingly, the organizational identity reflects how the organization views itself (Navis & Glynn, 2011), which results from a set of consensual and collective self-referential claims by the organization's members (Nag et al., 2007) about what they see as their organization's central, enduring, and distinctive elements (e.g., Albert & Whetten, 1985; Corley et al., 2006; Livengood & Reger, 2010; Whetten, 2006). Strategically, identity becomes a valuable signaling device, providing a means for positioning the organization in the minds of external stakeholders such as customers, suppliers, job seekers, and partners, and enabling the organization to acquire resources. This body of literature views organizational identity as founded upon institutionalized routines and binding organizational commitments (T. J. Brown, 2006).

Intrigued by the study of organizational identity, researchers have developed the original construct to account for its multiple facets (Gioia et al., 2013). However, there is currently no universal way to operationalize, conceptualize, and theorize organizational identity (Corley & Gioia, 2004; Khor, 2020). Instead, disparate disciplines (e.g., organization, strategy, sociolinguistic) have applied their precise angles to uncover and define the distinctive characteristics of organizational identity.

Organizational identity has been explored under different labels, ranging from organization-centric to audience-centric perspectives. Indeed, it includes both inwardly and externally oriented facets: as such, it has been defined as the identity perceived by the insiders,

the one that insiders want the outsiders to perceive, the one communicated to outsiders, and the one perceived by outsiders (T. J. Brown, 2006). Going a step further, scholars have suggested that organizations maintain multiple co-existing identities (Balmer & Greyser, 2002; Pratt & Foreman, 2000). According to Balmer and Greyser (2002), "actual, conceived, ideal, desired and communicated" identities define an organization.[6] In this paper, we adhere to the definition of communicated identity as the one "clearly revealed through 'controllable' corporate communication" (Balmer & Greyser, 2002, p. 74).

Starting from the paper by Corley et al. (2006), much of the literature on organizational identity has been devoted to studying identity dynamics (Hatch & Schultz, 2002), using multiple quantitative and qualitative approaches. Theory and research have depicted the processes that underlie the formation and changes in organizational identity. Regardless of how organizational identity is conceived, there is a certain agreement that it is not static, even if there is a tendency toward resilience to change (Chreim, 2005; Fiol, 2002). This body of literature recognizes that organizational identity can also change (in its meaning) over relatively short time horizons, but insiders tend to perceive it as stable and, as a consequence, act as if it was enduring (Gioia et al., 2010).

A relatively smaller and more recent stream of literature has focused on the processes underlying organizational identity formation (Ashforth et al., 2011; Glynn & Watkiss, 2012; Scott & Lane, 2000). Identity formation is a complex process that is affected by a number of factors, which are both internal and external to the organization (Gioia et al., 2010). These studies have also pointed out that the building of an organization's identity is pivotal to providing sensemaking in organizations in order to obtain and maintain acceptance and legitimacy in the environment (Clegg et al., 2007).

2.2 The role of investor identity in resource exchange processes

The literature on entrepreneurship has widely recognized the importance of creating an identity for new ventures, as this impacts the organization's ability to secure resources, recruit talent from the job market, and create networks with customers, suppliers, investors, and potential partners (Burns et al., 2016; Younger & Fisher, 2020). While the extant focus on identity for new ventures has been useful, it has also created a disproportional emphasis on "one side of the coin" in resource exchange processes (i.e., new ventures searching for funding) while disregarding "the other side" (i.e., capital providers) (Smith & Bergman, 2020). However, resource exchange is a complex and bidirectional process, where entrepreneurs acquire resources and investors provide them to valuable businesses (L. Huang & Knight, 2017).

Moved by the need to understand the identity process in resource exchange dynamics, a nascent stream of research has recently started to analyze the role of investor identity in mobilizing resources (Pontikes, 2012; Smith & Bergman, 2020). These studies complement the view of investors as passive actors, mainly engaged in sensemaking practices (Gioia & Chittipeddi, 1991; Navis & Glynn, 2010), with an active view where investors adopt practices of sensegiving to define who they are and what they do. The "investor identity work" (Smith & Bergman, 2020, p. 5), used to elaborate a sensegiving process, encompasses investors' directive organizational identity claims and actions towards both external actors (i.e., entrepreneurs and other investors) and internal actors (i.e., employees, management, boards), which serve to initiate and sustain resource provision.

Understanding the role played by organizational identity—from both the demand and supply side of the funding process—is fundamental for supporting the alignment between the two sides of the coin and facilitating resource exchange (Smith & Bergman, 2020). If the identity of the new venture is well understood, investors are more able to select "cherry-picking" instead of "frog-kissing" investments (Bertoni et al., 2016). Likewise, new ventures may actively seek those investors with clear-cut identities that overlap with that of the organization (Fisher, 2012). In this way, ventures can better grasp how to meet investors' expectations and potentially be considered for financing (Parhankangas & Renko, 2017). These two interrelated steps (Eckhardt et al., 2006) generate a process of mutual "identification", such that organizations identify themselves more strongly with another organization when their "self-concept contains the same attributes as those in the

[6] Often, the constructs of organizational identity include components of image, reputation, and branding (Balmer & Greyser, 2002). An attempt to propose a consistent terminology is provided by Brown (2006), who illustrates a framework that distinguishes between identity, intended image, and construed image.

perceived organizational identity" (Dutton et al., 1994, p. 239).

This matching process was originally articulated in psychology through the "similarity-attraction paradigm" (Byrne, 1971) and in sociology through the "homophily principle" (Mcpherson et al., 2001). In the literature on entrepreneurial finance, it is known as "positive sorting" (Sørensen, 2007): Reputable investors tend to match with the best companies in the market, moved by a reciprocal search for quality. Even if the matching depends on a different driving factor (quality rather than identity), research implies that, when faced with multiple options, investors and ventures both care about the attributes of their respective partners (Sørensen, 2007). Taken together, these studies provide initial evidence that investor identity is relevant to resource exchange processes. However, there is still limited work on the topic and, in particular, on hybrid investors' identities in emerging market categories. In such cases, the process of resource exchange is likely to face additional challenges due to the lack of a definitive and widespread consensus on the organizational identity of market members.

2.3 The identity of social hybrid investors in emerging market categories

Defining a clear identity for members of an emerging category is not a straightforward task (Clegg et al., 2007; Durand & Khaire, 2017; Durand & Paolella, 2013) and deserves particular attention. In well-established fields, identity formation is easier to achieve (Czarniawska & Wolff, 1998), but in newly formed domains, the rules guiding behaviors are preliminary and unclear (Durand & Khaire, 2017; Durand & Paolella, 2013). Secondly, during the critical phase of industry emergence, low barriers to entry translate to a heterogeneous pool of actors with different interpretations of the organizational identity (Jensen, 2010; Lamont & Molnár, 2002; Nicholls, 2010). Finally, organizational identity formation is pivotal not only for single organizations obtaining and maintaining acceptance in the market (Clegg et al., 2007) but also in defining the identity of the whole industry, which ultimately reflects the identities of the member organizations (Gustafsson et al., 2016; Stigliani & Elsbach, 2018). Indeed, "the formation of a new market category is an active, social project that likely involves the interpretations and actions of [market actors]" (Navis & Glynn, 2010, p. 441) and originates as "unstable, incomplete, and disjoined conceptual systems held by market actors" (Rosa et al., 1999, p. 64).

The lack of clear boundaries and a precise definition of the impact investor prototype has created space in the market for opportunistic behaviors (Busch et al., 2021; Findlay & Moran, 2019). Some actors, moved by the goal of maintaining their level of competitiveness by leveraging the trend towards sustainability (Findlay & Moran, 2019; Freireich & Fulton, 2009), deviated from the social and transformative mission that should be the vital goal of these investments (Harji & Jackson, 2012; Hehenberger et al., 2019; Höchstädter & Scheck, 2015). Thus, the interpretation of investors' social-oriented identity may be driven either by the perception of the investor's genuine concern about generating social impact or the investor's attempt to "window dress" to seize a new market opportunity. Previous studies have tried to capture such a controversial nature and dual identities by analyzing the criteria used in evaluating potential investment opportunities, distinguishing between "social sector criteria" and "traditional entrepreneurial sector criteria" (Miller & Wesley II, 2010). As hybrid organizations systematically integrate civil society and markets, they possess different conceptualizations of their identity (i.e., regarding what is central, distinctive, and enduring about their organization) and borrow distinctive elements from both the social and commercial sectors (Pharoah et al., 2004; Pratt & Foreman, 2000). For these organizations, the use of language in describing their nature becomes even more relevant to capture the emphasis given to one sphere (society) with respect to the other (market). In light of these considerations, social hybrid investors need to communicate their identity in a clear and precise manner to activate the proper stakeholder perceptions about "what the organization is" and "what the market is."

3 Communicating the organizational identity

3.1 The use of language to communicate the organizational identity

One of the most important decisions that organizational managers can make is how the organization's position and identity are communicated to external stakeholders (Pan et al., 2018). This clearly affects inter-organizational relationships—between firms,

investors, and other stakeholders. Managers typically concentrate on a limited set of attributes they would like to convey to external stakeholders (T. J. Brown, 2006) and then take actions to strategically communicate this organizational identity—what we refer to as a "communicated identity" (Balmer & Greyser, 2002).

This effort can involve different channels that are more (e.g., advertising, marketing materials, press releases, or websites) or less (e.g., word of mouth, media commentaries) controllable by the organization (Botero et al., 2013). When adopting controllable communication means, such as official websites, managers have to strategically decide what and how to communicate to clearly define the organization's attributes while simultaneously differentiating it from others in the eyes of external stakeholders (Scott & Lane, 2000). In these contexts, language plays a relevant role (Packard & Berger, 2017; Riley & Luippold, 2015; Thibodeau & Boroditsky, 2013; Younger & Fisher, 2020).

In the past decades, language scholars have scrutinized the role of language in constructing and defining identities (and identity types) using different approaches (e.g., spanning corpus linguistics, sociolinguistics, or psycholinguistics) (De Fina, 2019; Edwards, 2012; Zenker, 2018). Despite the wide range of perspectives adopted, there is an overall agreement that language is inextricably linked with identity (De Fina, 2019; Khor, 2020). Language, broadly understood, has the power to shape how organizations identify and present themselves (Eckert, 2000; O'Connor, 2002), which, in turn, affects the different inferences that audiences make about an organization's identity (Cutolo et al., 2020; Lounsbury & Glynn, 2001; Martens et al., 2007). Said differently, the narrative of organizational identity is similar to the one adopted by a writer who may describe an event using different options—with colorful words and unusual syntax or with a more formal language. As a consequence, the same story can emphasize different perspectives and features with different stylistic nuances, representing distinctive identity footprints.

Much of the work done in sociolinguistic and psycholinguistic research depends on the recognition that language is a critical strategic asset in the formation and dispersion of an organization's identity among the social and psychological spheres of a target audience (Graffin et al., 2011; Tannen, 1995; Tausczik & Pennebaker, 2010; Toma & D'Angelo, 2015). An organization can communicate its values, beliefs, missions, goals, and knowledge by adopting different perspectives of self-representation. For example, the subsidiaries of a multinational corporation can use multiple languages and linguistic styles in their lines of communication, thus showing different identity types (Iwashita, 2022).

However, the use of communication strategies to inform a target audience of an organization's identity has neither the same relevance nor the same impact across different organization types. Communicating a comprehensible organizational identity is especially crucial when organizations do not conform to extant categories with a high degree of institutionalization (Navis & Glynn, 2011).

3.2 Content and style components of language

Any communication includes both content and style. Content refers to the meaning, while style mainly refers to how something is said. The relationship between content and style in identity formation has been a topic of interest for language scholars for several decades. Consequently, the field now features a variety of theoretical and methodological methods focused on whether content matters more than style (or vice versa) in identity formation. Subfields have also emerged that provide even more nuanced perspectives (De Fina, 2019).

Taking inspiration from the communication and linguistic literature, organizational and management research has incorporated a language into its groundwork to emphasize the relevance of both content and style in the communication of an organization's identity (Larrimore et al., 2011; C. H. Miller et al., 2007). Thus, some research has devised conscious strategies that managers can use to convey precise communication content to a target audience (Bolino et al., 2008). Other studies have suggested that, above and beyond content, communication embeds language attributes (e.g., language diversity, intensity, and concreteness) that shape audience impressions about an organization's identity (Pan et al., 2018; Pennebaker, 2011; Pennebaker et al., 2003). For example, research has found that concrete language—referred to as the degree to which the words in a message provide context-specific and detailed information (Hansen & Wänke, 2010; ter Doest et al., 2002)—is a powerful means of informing the audience, especially in risky situations (Larrimore et al., 2011; Pan et al., 2018) marked by high uncertainty and informational voids.

However, the linguistic turn has only recently found its way into entrepreneurship and entrepreneurial

finance research (Clarke & Cornelissen, 2011; Cutolo et al., 2020; Martens et al., 2007). A stream of studies in entrepreneurship has focused on the role played by narratives, stories, metaphors, and linguistic frames in conveying a comprehensible identity to entrepreneurial ventures when raising capital (Aldrich & Fiol, 1994; Benford & Snow, 2000; Hsu & Hannan, 2005; Lounsbury & Glynn, 2001; Martens et al., 2007; Navis & Glynn, 2010; Wry et al., 2011). A number of works have also explored to what extent entrepreneurs' linguistic styles (e.g., the way they describe their project in a business plan, during an elevator pitch, or in a crowdfunding campaign) can affect the fundamental, yet challenging, task of attracting external resources (Allison et al., 2015; Clarke & Cornelissen, 2011; Manning & Bejarano, 2017; Parhankangas & Renko, 2017). All these studies provide strong evidence of the crucial role that language plays in the management sciences and its potential to explain dynamics in the field of entrepreneurial finance.

3.3 Linguistic styles of social hybrid investor's identity in an emerging market

3.3.1 Social linguistic positioning

Several academics and practitioners stress that impact investors must be characterized by "intentionality" in their actions, as the generation of social value cannot be an "incidental side-effect of a commercial deal" (A. Brown & Swersky, 2012, p. 3). While the goal of generating social value should be central to an organization in the field of impact investing (Miller & Wesley II, 2010), the market shows a fragmented scenario where categories vary in terms of their social orientation (Freireich & Fulton, 2009; Höchstädter & Scheck, 2015; Moore et al., 2012).

Investigating the notion of social linguistic positioning—which we define as investors' linguistic intensity in presenting their social orientation—is extremely relevant to capturing these investors' social intentionality. Linguistic positioning is the conscious and planned strategy used to convey the most relevant signs of an organization's identity (Bolino et al., 2008). In short, it is the quick and concise communication of who an organization is and what it does (Bart et al., 2001; Ireland & Hirc, 1992). Since actions are influenced by the aims communicated to stakeholders (Bart et al., 2001; O'Gorman & Doran, 1999), a communicated identity directed towards social impact is interpreted as the organization's driving force and a reflection of its underlying social purpose. Thus, how social hybrid investors linguistically emphasize and communicate their social orientation to external stakeholders may be useful for capturing their public declaration of social intentionality.

3.3.2 Linguistic distinctiveness

One way for organizations to be recognized within a market category is by distinguishing themselves from their peers (Brickson, 2005; Gioia et al., 2010). Organizational studies have defined the concept of distinctiveness as the level to which an organization is perceived as different from, rather than interchangeable with, other category members (Brickson, 2005; Gioia et al., 2010; Navis & Glynn, 2011; Younger & Fisher, 2020; Zhao et al., 2017).[7]

Language can play an important role here, not only by conveying more salient information (Rindova et al., 2007) but also by creating new and informative content with respect to what has been communicated by others, thereby increasing the perceived comprehensibility and credibility of identity (Guo et al., 2020). Thus, linguistic distinctiveness refers to the use of language to frame identity in a way that separates one from their peers (Cao et al., 2017; Navis & Glynn, 2011). In the context of impact investing, language may shape the relationship between investor and investee and, therefore, may be a major factor in a market's effectiveness.

It is worth highlighting that the paradigm of impact investing implies a radical change from investors' traditional approach, which is exclusively motivated by the generation of financial returns (Pache & Santos, 2010). In the entrepreneurial finance literature, several works have recently explored the motives that guide investments in specific assets or new instruments (Allison et al., 2015; Barber et al., 2021; Block et al., 2021; Fisch et al., 2021; Geczy et al., 2021; Vismara, 2016).

Impact investors seek to serve society through a form of governance where decisions are motivated by the investor's personal commitment to specific social challenges. The introduction of the social dimension has, thus, foregrounded

[7] The idea that organizations need to not only fit in with, but also differentiate from, peers of the same market category has recently been endorsed by management and organization scholars in studies on entrepreneurship (Fisher, 2012; Haans, 2019; Taeuscher et al., 2021; Taeuscher & Rothe, 2021; Younger & Fisher, 2020).

investor-investee relationships based on collaboration, mutual understanding, and reciprocal engagement, which are fundamental drivers of market effectiveness (Agrawal & Hockerts, 2019). Distinctive language can help foster these inter-relationships by emphasizing key and unique attributes of the organization's value (Bishop et al., 2019). Indeed, scholars broadly recognize that salient and novel information, rather than old and familiar, produces a more comprehensive understanding (Boswijk & Coler, 2020; Ellis, 2016, 2017; Giora, 2003; Tomlin & Myachykov, 2015).

4 The context of social impact venture capitalists

Impact investing consists of a broad range of financial institutions that use a heterogeneous pool of financial tools (Boni et al., 2021; M. Lee et al., 2020; Revelli & Viviani, 2015).[8] Given its breadth, social impact finance has been recognized as a potential new financial paradigm (Nicholls, 2010) and has recently received significant attention from financial bodies, private and public companies, and the press (Fink, 2020, 2018; The Economist, 2017; Zingales, 2018). This growing interest has accompanied the rapid evolution of the market for social impact investments. According to the Global Impact Investing Network (GIIN), nearly 10,000 social impact investments worth more than 46 billion dollars were financed in 2019 (Hand et al., 2020).

Despite these numbers, several observers believe that the social impact market has evolved into the "market building" stage (Ormiston et al., 2015) and that it is still far from reaching efficient global functioning (Lehner & Nicholls, 2014). One of the major issues affecting its aftermath is the lack of definitional and conceptual clarity (Nicholls & Daggers, 2016), which can lead to the concept becoming diluted: the so-called impact washing risk (Harji & Jackson, 2012; Hehenberger et al., 2019; Höchstädter & Scheck, 2015). Some global networks, such as the aforementioned GIIN and the Impact Investing Policy Collaborative (IIPC), have emerged in an attempt to provide the field with precise boundaries.[9] For the field to maintain its original transformative power, scholars have highlighted the need to intentionally generate social and environmental impact into practices (Höchstädter & Scheck, 2015; Truong & Nagy, 2021). Recently, scholars have suggested that definitions of impact investing need also to consider "additionality"—to guarantee that the social or environmental outcome generated goes beyond what would otherwise have occurred (Hebb, 2013; So & Staskevicius, 2015).[10] However, there is still no definitive consensus on what "social impact" means. Research in this field thus remains hobbled by the lack of clear boundaries and a unifying paradigm around social impact.

Considering the existence of multiple types of investors in the social impact field, we focus on financial organizations that invest equity capital to become partners of their invested companies (Miller & Wesley II, 2010): SIVCs. SIVCs operate in a market largely characterized by uncertainty, information asymmetries, and blurred boundaries. Contrary to the burgeoning literature on socially responsible investments (Arjaliès & Durand, 2019; Cheng et al., 2014; Hawn et al., 2018; Ioannou & Serafeim, 2015; Yan et al., 2019), studies on SIVCs are still in their infancy. The early study by Miller and Wesley II (2010) opened the door to a growing area that investigates how financial organizations might not only generate financial value, but also advance broader societal goals (Nicholls, 2010). Miller and Wesley II (2010) narrowly focus on the decision rules of funds in the sphere of venture philanthropy. Only recently has research started to explore the emerging category of SIVCs. These works have analyzed several aspects, ranging from SIVCs' intentional willingness to pay for impact (Barber et al., 2021), to their contracting practices toward both investors and

[8] Social impact finance utilizes several financing instruments, ranging from alternative currencies, community investment, crowdfunding, ethical banking, microfinance to social impact bonds, social impact investing, social responsible investment, social entrepreneurship, and venture philanthropy (Allison et al., 2015; Howard, 2012; Périlleux, 2015; Rizzi et al., 2018).

[9] These networks aim to match the supply and demand of capital in social impact finance, as well as set new standards and metrics to evaluate social impact investments (e.g., Impact Reporting and Investment Standards (IRIS) and the Global Impact Investing Rating System (GIIRS)) (Ormiston et al., 2015).

[10] Lee et al. (2020) ran experiments to demonstrate that impact investors must transcend their "categorical cognition" to realize the full potential of their investments.

portfolio companies (Geczy et al., 2021), to the criteria adopted when screening social enterprises (Block et al., 2021) and when making capital allocation decisions (M. Lee et al., 2020).

In their governance structures and investment modes, SIVCs align with the dictates of the VC industry (Cetindamar & Ozkazanc-Pan, 2017), except that they aim to generate a positive and measurable social and environmental impact alongside a financial return. They have been labeled "pragmatic idealists" (Bocken, 2015) because, while rigorously implementing investment practices that are common knowledge in the VC industry, they tend to be more patient in their exit strategy and target a range of returns that can be below or equal to the market rate, depending on the investors' strategic goals. They typically invest in businesses that are expected to tackle domestic or global social problems, operating in sectors such as education, microfinance, energy, and accessible basic services (e.g., housing, water procurement, and healthcare). To operate in line with social mandates, SIVCs' social returns must be *a priori* defined and ex-postevaluated (Calderini et al., 2018). Thus, SIVCs embody a new practice in the investing arena that exemplifies the hybrid logic of combining financial sustainability with social welfare goals (Battilana et al., 2017; Battilana & Lee, 2014). In short, they represent an ideal context for our work.

5 Methodology

To identify SIVCs that are active worldwide, we extrapolated information from ImpactBase, an online database managed by the GIIN.[11] Of the 445 active investors reported as social impact in the database, social impact in the data we only selected those included in the "private equity" and "venture capital" categories. Subsequently, we complemented the information using a second commercial database, Thomson One Banker, managed by Thomson Financial. We obtained a final sample of 195 SIVCs. We evaluated investors' communication by collecting all the text on their official websites (excluding external links and attachments) and manually crawling their content. We limited our analysis to the content of homepages, "about us" sections, and the pages describing the fund activities and investments. This allows us to obtain a corpus of comparable documents, focus on the most featured information provided to stakeholders, and limit the analysis to the language used by SIVCs (and not, for example, by media outlets). All websites had English as a common language. Official websites are "controllable" channels that organizations use to communicate their identity (Acs et al., 2021; Balmer & Greyser, 2002; Tietze et al., 2003). With the rapid spread and growing popularity of the Internet, organizational and management research has emphasized the role of website language (i.e., as text, narrative, story, and discourse) as a tool for delivering useful information to the public and ensuring that the intended messages are interpreted correctly (Botero et al., 2013; Gatti, 2011; Kent & Taylor, 1998; Taylor et al., 2001; Wirtz & Zimbres, 2018). In our context, websites occupy a relevant position for SIVCs' communication strategy, as they are designed to convey the investor's social impact identity and capture the different degrees of language distinctiveness (Aral & Van Alstyne, 2011; Gloor, 2017).

5.1 Evaluating social linguistic positioning

To measure SIVCs' social linguistic positioning, we used a metric of text mining and social network analysis—the Semantic Brand Score (SBS) (Fronzetti Colladon, 2018)—which is specifically designed to evaluate a concept's textual importance. The SBS is a novel measure of semantic importance inspired by well-known brand equity models (e.g., Keller, 1993).[12] In this research, we used it to measure the importance of thematically relevant terms such as "social," "impact," and their synonyms. Indeed, these words were

[11] The GIIN is globally recognized as a network that provides an "official certification" to social impact investors. Data were extracted from ImpactBase in 2019. Later on, the dataset was closed by GIIN and replaced by other data sources, such as Impact Space or Impact Assets. We carefully checked that analyzed SIVCs were reported in one of the two datasets.

[12] Although the measure was originally used to analyze the strength of a brand, it can be applied to any word or set of words, as it is calculated with textual data.

often used on SIVCs' websites to emphasize the social impact part of the investment. We call this set of words "Social Impact Words" (SIWs). To determine the best set of words, we used a double approach. Firstly, we referred to words related to the social impact of a SIVC's activity by looking at words used in past research (Barber et al., 2021). Secondly, we used an automated approach to extract keywords from our corpus. In particular, we employed the TF-IDF metric (Jurafsky & Martin, 2008).[13] Following this double approach, we derived a list of keywords that were subsequently evaluated by two experts in social impact investing. The two experts first worked independently to select the final SIWs and then met to find agreement on a few discordant cases. The final list of SIWs used for this study includes the following: community invest, disadvantaged, ethical invest, ethical objectives, ethically conscious, ethically motivated, impact, impact investing, impoverished, invest ethical, investing ethically, minority community, mission driven, mission investing, mission oriented, mission related, poverty, S.R.I, socially responsible, social objectives, socially, social finance, social good, social impact, socially motivated, socially responsible, socially conscious, sustainable, sustainable development, sustainable economic development, and sustainable investment.

Accordingly, a high score indicates a website communication that highly emphasizes the social impact side of investments. We looked at the distribution of the scores by plotting the quantiles and breaking them down into quarters. Since quantile plots get steeper above the upper quartile, we took the 75th percentile as a threshold for high values of the metric. In other words, the points above, which lie one-quarter of the data, indicate that website communication largely emphasizes social impact.

The SBS is calculated based on three dimensions: prevalence, diversity, and connectivity, which respectively account for how often a concept is mentioned, how rich its textual associations are, and how strongly it can bridge connections across different topics in the discourse. Prevalence measures the frequency with which SIWs appear on each SIVC website: the more frequently they are mentioned, the higher their prevalence. The idea is that the frequency of a word in a text could increase its potential for activation.[14] Website visitors will be aware of words representing the social theme when they read them, which is reflected in SIVCs' use of these words when they communicate. The second dimension, diversity, measures the heterogeneity of the words co-occurring with SIWs; a richer discourse entails higher diversity. A concept/keyword could be "mentioned frequently in a discourse, thus having a high prevalence, but always used in conjunction with the same words, being limited to a very specific context" (Fronzetti Colladon, 2018, p. 152). This measure is higher when textual associations are more diverse. Notably, previous research has shown that a higher number of associations has a positive effect on brand strength (Grohs et al., 2016). The third component, connectivity, expresses how often a word serves as an indirect link between all the other pairs of words while constructing a co-occurrence network. It reflects the embeddedness of the words related to the social impact theme in a SIVC website and can be considered an expression of their connective power (i.e., their ability to indirectly link different topics). While the social impact theme could be frequently mentioned (high prevalence) and might have heterogeneous associations with other concepts (high diversity), SIWs could still be peripheral and disconnected from the core of online communication.

To assess the measure, we first needed to process textual data to remove stop-words (i.e., those words that usually provide a little contribution to the meaning of a sentence, such as the word "and"), punctuation, and special characters. We changed every word to lowercase and extracted stems by removing word affixes (Jivani, 2011) using the NLTK Snowball Stemmer algorithm (Perkins, 2014). The next step was to transform text

[13] We also tested other approaches for determining the initial set of keywords—such as the TextRank algorithm (Mihalcea & Tarau, 2004)—but these did not yield better results.

[14] This effect could depend on factors other than term frequency. For example, rare terms, or terms that surprise, seem to have greater priming power—that is, the power to positively influence the probability that the same type of term will be found subsequently (Ellis, 2017; Hilpert, 2017; Rosemeyer & Schwenter, 2019).

documents into social networks where nodes are words that appear in the text. An arc exists between a pair of nodes if their corresponding words co-occur at least once; the frequency of co-occurrence determines arc weights. Following this procedure, we obtained 195 networks—one for each website. We adopted a five-word window for determining the maximum co-occurrence range and filtered out negligible co-occurrences. Nodes representing SIWs were merged into a single node in order to calculate their aggregated level of importance. Figure 1 provides an example representing the co-occurrence network generated by the following sentence (after removing stop-words, while skipping stemming for the sake of readability): "We invest in innovative technology to solve problems and sustain growth in agriculture and animal health."

Prevalence was measured as the frequency with which SIWs were mentioned on each SIVC website. Diversity was operationalized through a measure of network centrality (distinctiveness centrality) that takes into account the number of textual associations, which corresponds to the degree of the SIWs' nodes, rescaled based on their degree of uniqueness (Fronzetti Colladon & Naldi, 2020). If we take the example of Fig. 1 and consider the node "health," diversity would attribute more importance to the connection of this node with the word "animal" than with the word "sustain." This happens because the word "sustain" is connected to all other words in the network, whereas the word "animal" is not, thus making this association less common. Connectivity reflects the "brokerage power" of SIWs on each website, calculated using weighted betweenness centrality. Specifically, we considered the inverse of arc

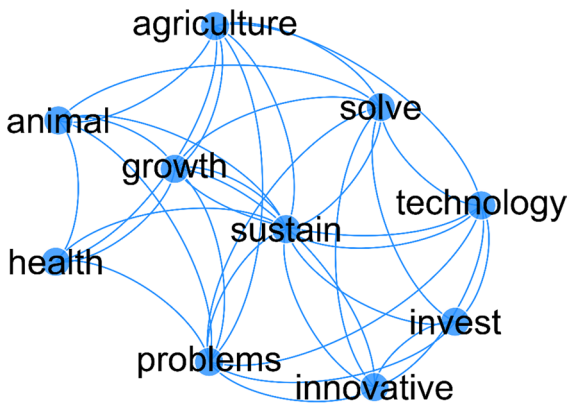

Fig. 1 Co-occurrence network

weights in determining the shortest network paths and then calculated weighted betweenness centrality using the algorithm proposed by Brandes (2001). In particular, we considered the network paths that interconnect the different words in the co-occurrence network. For example, in Fig. 1, we notice that the words "animal" and "invest" are not directly connected. Therefore, it is necessary to go through other words to connect these two words. Accordingly, connectivity is high when a word (node) frequently lies in the shortest network paths that interconnect the other words in the corpus.

To compare measures derived from different networks (i.e., one network per website), we standardized the values of prevalence, diversity, and connectivity. For each measure, we conducted standardization by considering the mean and standard deviation of scores obtained by all the words on the website. The SBS was subsequently calculated as the sum of the standardized values of its components. According to this standardization procedure, SBS can either be positive or negative based on the importance of the social impact words have on each website.

5.2 Measuring linguistic distinctiveness

The second part of the analysis involved measuring the distinctiveness of each website, which considers how much of a website's content is not already available on the majority of other SIVCs' websites. The distinctiveness indicator thus measures the extent to which a website uses rare (or new) words and introduces non-redundant information. While some websites only featured a single page with little information, others had multiple pages and were rich in content. However, the length of the text included in each website is not a good proxy of its distinctiveness, especially when considering that the majority of web users only read a reduced portion of content and sometimes stop on homepages. How much of a page is read can vary based on many factors, such as the availability of an abstract or the structure and design of websites (Nielsen, 1997, 2008; Nielsen & Loranger, 2006).

More precisely, we calculated the distinctiveness indicator based on the term frequency-inverse document frequency (TF-IDF) information retrieval metric (Jurafsky & Martin, 2008). Websites present new information only if they contain words that do not commonly appear on all other websites and if the message they convey is not lost in uninformative text blobs. Therefore, the frequency of each word's occurrence is

multiplied by the logarithmically scaled inverse fraction of the documents that contain that word. Distinctiveness for a specific website is calculated as:

$$\text{Distinctiveness} = \frac{1}{n}\sum_{w\in V} f_w \log\frac{N}{n_w}$$

where N is the total number of documents in the corpus (i.e., the number of websites); n is the total number of words that appear on a website, and V is their set; f_w is the frequency of word w, and n_w is the number of websites where the word w appears. We pre-processed the texts the same way as the SBS calculation (stemming, removal of stop-words, etc.). Similar to the case of the SBS, we looked at the distribution of the values of the metric and took the upper quartile to identify high values of distinctiveness.

6 Results

6.1 Four different ways to communicate a social hybrid investor identity

Thanks to the methodology presented in the previous section, we were able to classify SIVCs into four categories based on the intensity (low versus high) of the social impact theme (our measure of social linguistic positioning) and the distinctiveness of the language used on their websites. Clearly, neither of these two dimensions is an either-or proposition: Most SIVCs will fall somewhere along the spectrum of possible values between the two extremes of each pair of language characteristics. Still, we are confident that overlaying the two dimensions creates a useful map for understanding the nuances and complexity of the phenomenon. Our results show that, from the organizational identity perspective, the market is characterized by a high heterogeneity of linguistic style, which resembles characteristics of ambiguous and emerging market categories.

We found 16 SIVCs with high values for both social linguistic positioning and linguistic distinctiveness, 33 SIVCs that were only high in social linguistic positioning, and another 34 that were only high in linguistic distinctiveness. The remaining 112 SIVCs had a website communication that neither particularly emphasized the social impact theme nor provided new and non-redundant information with respect to competitors. We labeled the first group as *Smart Heroes*: "Smart" because the SIVC is able to communicate in a distinctive way (compared to others) and "Hero" because it has a strong social orientation and thus is more likely perceived as a champion in addressing social challenges. This category of SIVC strongly emphasizes its social conscience and does so by framing and proposing information to external stakeholders in a distinctive way. The second group, the *Naïve Dreamers*, has a strong social identity, similar to the *Smart Heroes*, but its communication differs little from the mass. By explicitly endorsing the social cause without being able to communicate its identity in a distinctive way, this SIVC appears to have organizational "dreams" regarding social issues but remains "naïve" about how to translate its focus into a distinct social identity. The third group is labeled *Illusionists*. This type of SIVC adopts a linguistic style that distinguishes it from others but without the social content that should characterize the declared identity of a social impact VC. Just as "illusionists" seek to enchant the audience with a range of tricks that mask reality, this type of SIVC tries to attract the external audience by adopting distinctiveness in its communication without ever showing a true social conscience. Finally, the *Blabbers* represent SIVCs that neither center on the social theme nor distinguish themselves from the masses in their communication. Figure 2 illustrates these four SIVC categories, while Table 1

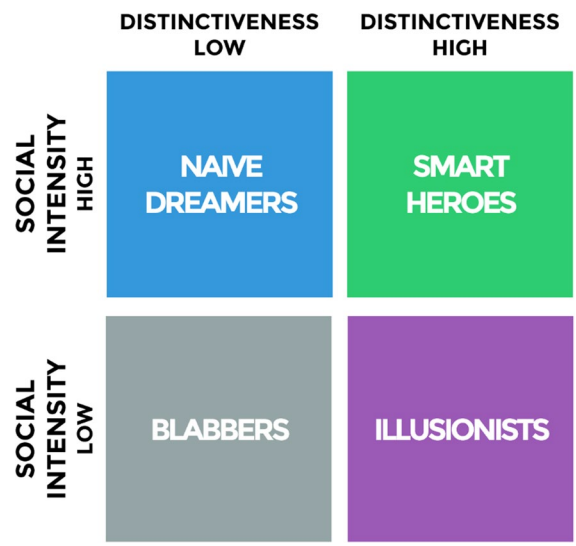

Fig. 2 A typology of SIVC's linguistic style

Table 1 Example of sentences for each SIVC's linguistic style

SIVC type	Social linguistic positioning	Linguistic distinctiveness	Representative sentences
Smart Heroes	High	High	"As early stage investors, we appreciate the importance of being strategically and operationally present with Founders on their journeys. […] While they [founders] follow their visions and chase individual dreams of impacting lives in their own unique ways, we aim to be there as facilitators, enablers, problem solvers, mediators and cheerleaders - always with an intent to catalyze progress."
Naïve Dreamers	High	Low	"We look for companies with a strong social orientation coupled with innovative business models. We invest in high quality entrepreneurs who demonstrate the intent to provide social returns to investors along with a healthy financial return."
Illusionists	Low	High	We are prepared to work hard to deliver what we promise and to add value that goes beyond what is expected. We always use our energy, skills and resources to deliver the best results. [...] Our approach is defined by "inspired ownership" and we strive to act as a meaningful strategic partner and not just as a financial sponsor.
Blabbers	Low	Low	"The fund provides private equity to entrepreneurial micro-finance organisations. […] Its goal is to increase the outreach of commercial micro-finance organisations and to provide a market rate of return to the investors. […] It has the ability to identify and invest in promising enterprises in the micro-finance sector."

shows some examples of sentences used by SIVCs in their websites for each linguistic style.

In our analysis, we considered multiple keywords to represent the social intensity dimension (see Section 5.1). We find no particular differences in the use of these words across groups, apart from the fact that Smart Heroes and Naïve Dreamers use them more and in more central positions in the semantic network. We also find that the most used term is referred to the social impact of the investment. Lastly, we observe that Illusionists are those who refer more to their goal of supporting minorities.

Blabbers represent the largest group in our analysis. To discuss within-group heterogeneity, we better explored the Blabbers category with respect to the dimensions of distinctiveness and social intensity. Results are presented in Fig. 3.

As Fig. 3 shows, Blabbers scores are rather homogeneously distributed with respect to the social intensity dimension—with some websites presenting a zero score (indicating that little importance was attributed to the communication of the "social side" of the investment). The same is true for the distinctiveness dimension, for which we provide a visual differentiation of low, high, and average scores. In general, we can neither recognize clear clusters nor group these observations with respect to other characteristics of SIVCs—such as their impact theme, geography, target geography, or stage of development.

6.2 Toward a better understanding of SIVCs' types

Table 2 presents the main characteristics of the SIVCs in our sample. In particular, we considered: their age (measured in years starting from their inception); the committed capital (measured in US$); the geographical area in which they operate; their geographical specialization (i.e., the target geographical area for capital allocation); and their impact agenda, codified according to the social area of intervention. With "theme," we denote the societal area that SIVCs intend to address. We distinguish between basic services (i.e., investments in companies whose mission is to improve people's access to food, water, and education), energy and environment (i.e., investments related to interventions to fight climate change and global warming), finance (i.e., investments in companies facilitating people's access to microcredit initiatives), and multiple impact (i.e., investments in companies targeting more than one societal challenge). Lastly, we considered the stated stage focus of SIVCs that can be: early stage (i.e., when SIVCs target ventures in their seed and early stages of entrepreneurial firm development), multi-stage (i.e., when SIVCs focus on both seed/early and growth/later stages of development ventures), or other (i.e., the stated focus includes mezzanine finance, PIPE/recap, or buyout).

As the table shows, most of the investors' target ventures are in their seed and early stages of entrepreneurial firm development. In terms of impact theme, the *Blabbers* and the *Illusionists* have the more diversified investments, with the latter having a larger focus

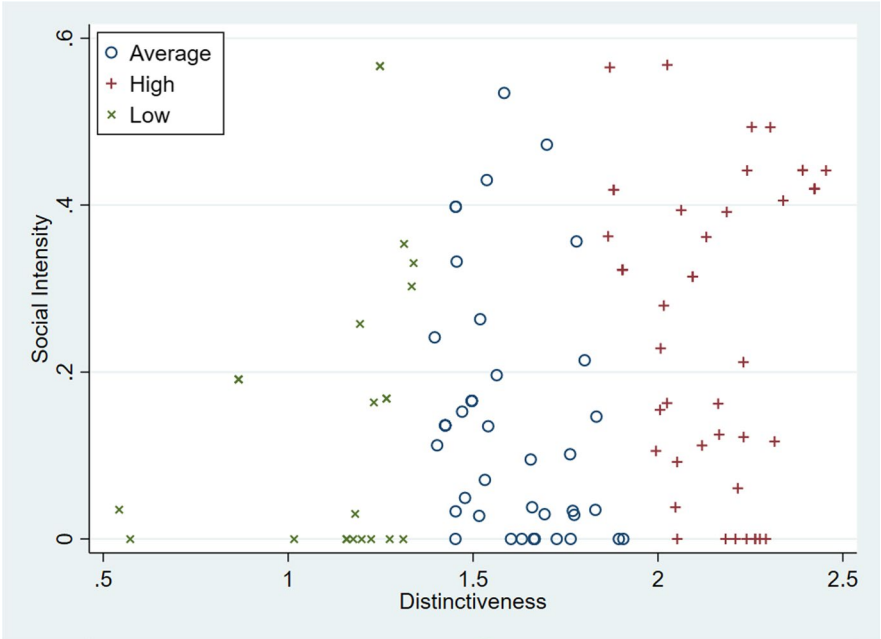

Fig. 3 Distinctiveness and social intensity of blabbers

on the energy and environment theme. *Naïve Dreamers* are those with the largest focus on finance. In general, the majority of SIVCs invest in companies targeting more than one societal challenge. Most SIVCs are located in North America and Europe, followed by Africa and Asia. These percentages are aligned with the target geographies, with some exceptions. For example, *Naïve Dreamers* are mostly located in Europe and have Asia as an important target geography. Many of these same SIVCs also target multiple geographical areas. *Blabbers* are primarily focused on Africa, whereas *Smart Heroes* are not. Lastly, we find no significant differences in terms of age and committed capital as the averages across groups are quite similar, also considering their standard errors. This is confirmed by the results of the Welch's robust test of equality of means ($p = 0.728$ for age and $p = 0.530$ for committed capital).

Figure 4 presents the results of a further analysis we carried out to study language similarity between SIVCs' websites. We wanted to understand if the four groups of Fig. 2 were using the same language or not. In particular, we used a bag-of-words approach and—after text pre-processing—represented each website as the bag of its words, disregarding order but preserving multiplicity. Accordingly, we constructed a document per term matrix, where each row represented a website, while columns represented the terms that appeared at least once in the corpus. Matrix cells were populated with term frequencies. To study websites' language similarity, we calculated a document per document distance matrix, using the cosine similarity metric that is typically employed in text mining (A. Huang, 2008). We subsequently plotted similarities in the two dimensions using the multidimensional scaling technique (Mead, 1992).[15] As the figure shows, *Smart Heroes* and *Naïve Dreamers* are much more clustered than *Blabbers* and *Illusionists*. The two former groups also have some degree of overlap, probably attributable to the importance they give to the social impact theme. *Blabbers* and *Illusionists*, on the other hand, spread across the entire graph without showing a consistent communication style.

Moreover, we calculated other well-known metrics of text analysis: sentiment, readability, and numerical intensity. Sentiment represents the positivity or negativity of the language used in communication, with values varying from −1 to +1—where −1 indicates a very negative valence of the text and +1 is a very positive one. Like the other metrics considered in this paper, sentiment was calculated through the SBS BI software (Fronzetti Colladon & Grippa, 2020),

[15] The picture shows a smaller number of points in respect to the size of our sample, as some are in perfect overlap, meaning that a few sites are a replication of each other.

Table 2 Descriptive statistics of the sample by SIVC type

	Smart Heroes Mean (SD)	Illusionists Mean (SD)	Naïve Dreamers Mean (SD)	Blabbers Mean (SD)
Age	8.19 (5.66)	7.56 (4.57)	6.73 (3.33)	7.18 (3.83)
Committed capital (US$ millions)	46.7 (86.1)	82.0 (145.0)	44.1 (50.8)	55.0 (75.9)
Stage				
Early	71.43%	74.19%	80.77%	69.57%
Multi	28.57%	22.58%	19.23%	27.17%
Other	0%	3.23%	0%	3.26%
Impact theme				
Basic services	0%	2.94%	0%	2.68%
Energy and environment	6.25%	8.82%	0%	5.36%
Finance	12.50%	0%	18.18%	7.14%
Multi	81.25%	88.24%	81.82%	84.82%
Geography				
Africa	12.50%	11.76%	15.15%	15.32%
Asia	12.50%	8.82%	9.09%	4.50%
Europe	18.75%	38.24%	51.52%	29.73%
North America	50%	41.18%	12.12%	47.75%
Central America	0%	0%	0%	1.80%
South America	6.25%	0%	9.09%	0.90%
Oceania	0%	0%	3.03%	0%
Target geography				
Africa	0%	17.65%	9.38%	18.75%
Asia	18.75%	14.71%	18.75%	8.93%
Europe	18.75%	14.71%	9.38%	9.82%
North America	37.50%	23.53%	9.38%	25%
Central America	0%	0%	0%	0%
South America	6.25%	0%	6.25%	2.68%
Oceania	0%	0%	0%	0%
Multi-area	18.75%	29.41%	46.88%	34.82%

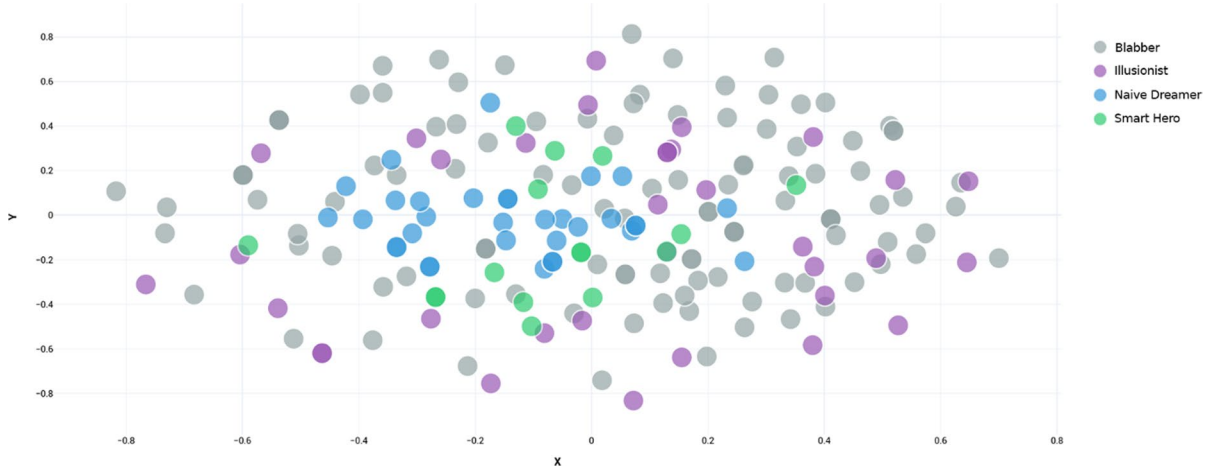

Fig. 4 Language similarity between SIVC's types

which uses the VADER lexicon for the English language (Hutto & Gilbert, 2014). Readability was calculated using the Gunning-Fog index, which proved useful, for example, in the analysis of crowdfunding campaigns (Du et al., 2015). Lastly, the dimension of numerical intensity was calculated to take into account the amount of quantitative information provided in SIVCs communication—i.e., counting numerical terms (including integers, numbers in lexical format, and terms referring to numerical operations) and dividing this number by the total word count (Hart, 2000; Henry, 2008; Short & Palmer, 2008). We compared the mean values of these indicators across groups and tested significant differences by the Welch's robust test of equality of means and the Games-Howell post hoc tests. The results of this additional analysis are presented in Table 3.

Not surprisingly, we notice that sentiment is very positive across all categories. However, the communication of *Blabbers* significantly differs from that of all the other groups, showing a lower positivity of the language used. The message sent by *Smart Heroes* is the most positive and significantly differs from all other categories, except for *Illusionists*, which also use a very positive language. On the other hand, differences are not significant in terms of readability and numerical intensity. In general, the communication of SIVCs is rather complex, with readability scores all indicating that a high level of education is required to properly understand the content of their websites.

As a final step, we complemented our analysis by exploring the main discourse topics that emerged on the SIVCs' websites. Topic modeling is increasingly used in management research to reveal constructs and conceptual relationships in textual data. This procedure can be used, for instance, to detect novelty and emergence or make sense of online audiences (Hannigan et al., 2019). In our context, we wanted to understand if there were prominent communication themes and how they were distributed across the four SIVC categories. Further, we investigated the differences in their use of language and the messages conveyed by their websites. We used a network approach to extract topics (e.g., Gerlach et al., 2018; Lancichinetti et al., 2015). Consistently with our previous analysis, we worked on the word co-occurrence network and found meaningful word clusters through the Louvain algorithm (Blondel et al., 2008).[16] Subsequently, we extracted the most representative words of each cluster by considering the weight of their connections and the proportion of internal and external links (Fronzetti Colladon & Grippa, 2020). Topics were manually labeled based on their keywords, as presented in Table 4.

Ten topics emerged from the analysis, with the most prominent (25%) being related to the profitability and positive social impact of investments (*Impact investing*). The sustainable solutions supported by

[16] We tested alternative approaches for network clustering, but did not achieve significantly different or better results.

Table 3 Welch's robust tests of equality of means

Variable	M	Welch's test significance (p value)	Games-Howell post hoc analysis			
			B	I	ND	SH
Sentiment						
B	0.9572	.000		*	*	*
I	0.9995		*		*	
ND	0.9990		*	*		**
SH	0.9997		*			**
Readability						
B	24.02	.136				
I	21.82					
ND	20.18					
SH	20.92					
Numerical intensity						
B	2.34%	.130				
I	2.00%					
ND	1.77%					
SH	2.05%					

*$p < .05$; **$p < .01$. *B*, blabbers; *SH*, Smart heroes; *ND*, Naïve dreamers; *I*, illusionists

Table 4 Topic modeling

Topic	Relevance	Top keywords
Impact investing	25.0%	investment, impact, returns, social, investors, financial, profit, risk, value, results, financing, economic, fund, responsible, positive
Sustainable solutions	17.5%	sustainable, technology, services, products, development, innovation, solutions, ecosmart, platform
Investor relations	17.0%	US, news, contact, newsletter, subscribe, information, email, send, updates, connect, LinkedIn
Target ventures	16.2%	entrepreneurs, small, medium, business, SMEs, micro, enterprises, beneficiary, venture, seeds, portfolio
Environmental impact	10.4%	forest, earth, land, conservation, sea, reforestation, world, WWC, water, Kyoto, solar, power, energy, renewable, electric, hydro, waste, recycling, clean, planet, deforestation, carbon, protection, nature, gas, fossil, glaciers
Geographical focus	4.9%	Africa, Nairobi, south, north, east, India, Bangalore, Saharan
Management team	3.7%	director, founder, CEO, associate, experience, partners, years, management, dr, board, advisory, chairman, committee
Community well-being	3.4%	ethnic, minority, women, education, urban, happy, youth, livelihoods, gender, housing, family
Financial instruments	1.2%	microfinance, microcredit, bank, debt, equity
Philanthropy	0.6%	donations, charitable, sponsors, campaigns, foundation

the SIVCs and the characteristics of the target ventures are also significantly discussed, accounting for 17.5% (*Sustainable solutions*) and 16.2% (*Target ventures*) of the discourse, respectively. All websites have sections meant to inform investors and promote contact (*Investor relations*). These sections present short news, contact information, and newsletters and roughly account for 17% of the communication. The social impact of the investments is also discussed in more detail, with a particular focus on the *Environmental impact* (10.4%) and the *Community well-being* (3.4%); notably, the former receives twice the attention. Some websites are more informative than others with respect to their *Management team* (3.7%) and the *Geographical focus* of their investments (4.9%)—even if these two topics are less relevant in the overall communication. Two minor topics concern philanthropic initiatives (0.6%) and specific *Financial instruments* (1.2%). Figure 5 shows the topic distribution across the different SIVCs categories, i.e., how much each topic is relevant in the communication of each group, on average.

We see that communicating impact investing is most relevant for *Smart Heroes* and *Naïve Dreamers*. This evidence is in line with the strong social identity of these two categories. Impact investing is an important topic for the other two categories but is less prominent. *Illusionists*, for example, devote much attention to the characteristics of their target ventures and management teams but very little to sustainable solutions. On the other hand, sustainable solutions are one of the main focuses of *Blabbers*. More than the other categories, *Illusionists* and *Blabbers* use their websites to promote news and general contact information. *Naïve Dreamers* are those with the most diversified communication skills but not the most distinctive—probably because the message sometimes remains general and covers too many topics. *Smart Heroes* dig deeper into the social benefits of the investments, putting much attention on the theme of community well-being—which is much less relevant in the communication of the other groups. This category and the *Illusionists* are the only ones that significantly promote the positive environmental impact of their investments.

7 Additional analysis: website traffic and linguistic styles

As an additional analysis, we assessed whether and to what extent the use of different linguistic styles is majorly effective at attracting the attention of external audiences by exploiting data on website traffic. We used the Amazon Alexa Global Ranking to estimate the differences in global website traffic for each SIVC (https://www.alexa.com/siteinfo), treating this as a proxy for SIVCs' ability to attract

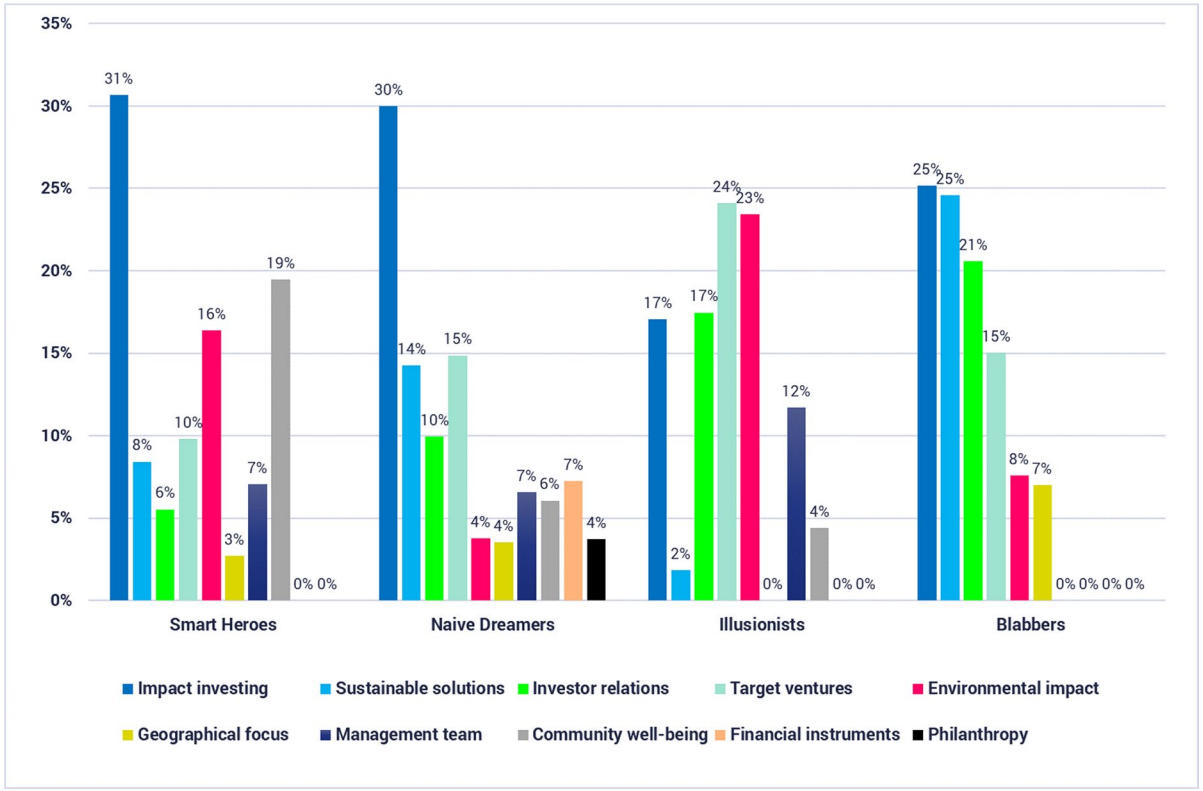

Fig. 5 Topic distribution by SIVC category

attention through their website communication strategies. Alexa, which provides a global ranking of more than 30,000,000 websites, is the most popular website traffic measurement system (Thakur et al., 2011). It is focused on traffic rather than on incoming links. A lower ranking is indicative of a higher ability to generate traffic, reflected in more page views and awareness of a SIVC. The Alexa database has been used in many studies, including the evaluation of page views for new venture companies' URLs (Goldfarb et al., 2007; Nuscheler et al., 2019; Reijden & Koppius, 2010; Winkler et al., 2019). The Alexa ranking was also used as a measure of venture capital customer traction (Hallen et al., 2014) or as one measure of performance for new ventures (Kerr et al., 2014). We manually crawled the Alexa database at the time of data collection. This analysis was restricted to 108 SIVCs that had a website with traffic data available.

Results suggest that *Smart Heroes* have significantly more page views for their URLs than the other three groups. Figure 6 shows the results of comparing the means of Alexa rankings for the four groups and comparing *Smart Heroes* with all others. According to the Welch's robust test of equality of means, the differences are statistically significant (p =0.045). Similarly, a t-test (with equal variances not assumed) comparing the websites of *Smart Heroes* with those of the other groups also produced a significant result (p = 0.010). We additionally carried out a non-parametric Mann-Whitney U test, obtaining a significant result at the 10% level ($p = 0.065$). Lower scores indicate higher rankings (i.e., more traffic).

In order to better evaluate the impact that linguistic styles could have on website traffic, we trained a tree boosting machine learning model designed for unbiased boosting with categorical features, namely CatBoost (Prokhorenkova et al., 2018). CatBoost is a computer-based machine learning method for dealing with "big data," such as large textual archives and repositories of images that enables the automatic extraction of knowledge and the implementation

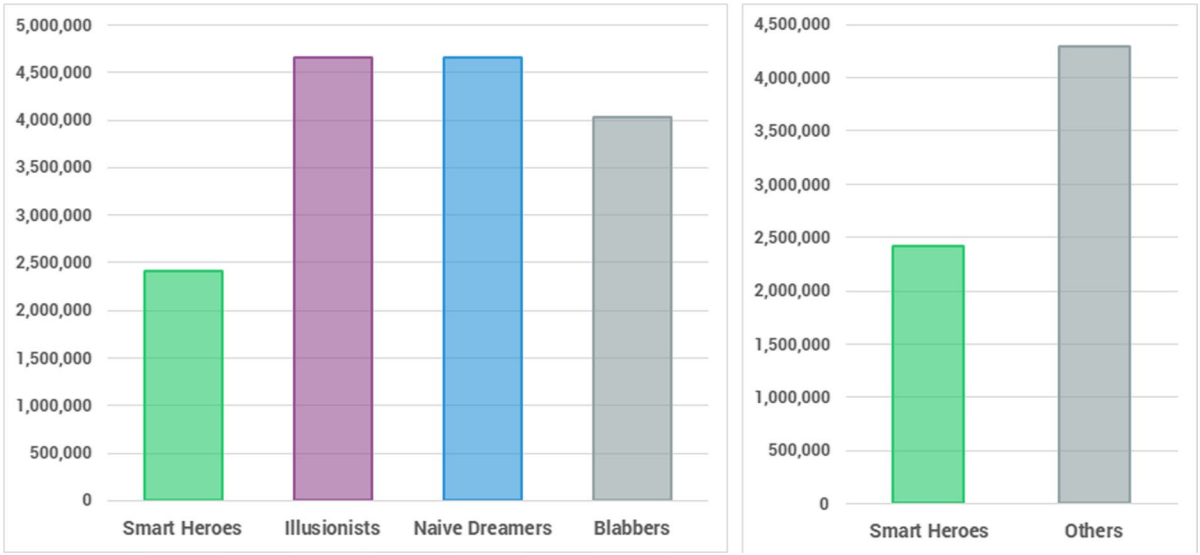

Fig. 6 Comparison of website traffic ranks

of optimization tasks (Choudhury et al., 2019; Cui et al., 2006). We chose this methodology over traditional OLS models as the relationships between our dependent variable and predictors do not necessarily follow regular curves. In addition, we wanted to use a nonparametric approach that is usually more powerful than OLS while making predictions[17]. Consistently, we did not evaluate model fitting on the in-sample, but we considered the out-of-sample model accuracy. Many fields (e.g., business, finance, and, more recently, strategy) have adopted machine learning methods as effective data mining instruments for extrapolating new and indistinct patterns of knowledge within a dataset, which can be used to improve predictive techniques and managerial decisions (Cui et al., 2006; Kleinberg et al., 2018).

With machine learning, we wanted to understand whether the variables of social linguistic positioning and linguistic distinctiveness could effectively support the prediction of SIVCs in the upper quartile of Alexa rankings (in our sample). In addition, we considered several other measures that could characterize SIVCs (e.g., age, geographical area and specialization, committed capital).

We validated the model results through Monte Carlo cross-validation (Dubitzky et al., 2007), with 500 random datasets split into training and test data. On average, the accuracy of predictions was 76%, and the area under the ROC curve was 0.7. Figure 7 shows the importance of each predictor, calculated as the average of its absolute Shapley values (Lundberg & Lee, 2017): The higher the score reported in the table, the more relevant the predictor. We considered the average model resulting from Monte Carlo cross-validation. SHapley Additive exPlanations (SHAP) is a well-known approach for determining feature importance, applicable to the output of different machine learning models. This method showed better consistency than previous approaches (Lundberg et al., 2020; Lundberg & Lee, 2017) and proved to be particularly appropriate for tree ensembles (Lundberg et al., 2018, 2019). These last analyses were carried out using the Python programming language, specifically the packages SHAP (Lundberg & Lee, 2017) and CatBoost (Prokhorenkova et al., 2018).

As the figure shows, SIVCs' geographical area and age are important determinants of website traffic. The third- and fourth-most important predictors are communication style and target geography. Meanwhile, committed capital, stage, and impact theme have a smaller effect on model predictions. The plots of Fig. 7 and Fig. 8 offer more detailed insights into the impact of each variable on model predictions and the contribution of each observation. For instance, for the first eight years of a SIVC's life, age has a

[17] Indeed, we tested OLS models without getting better results.

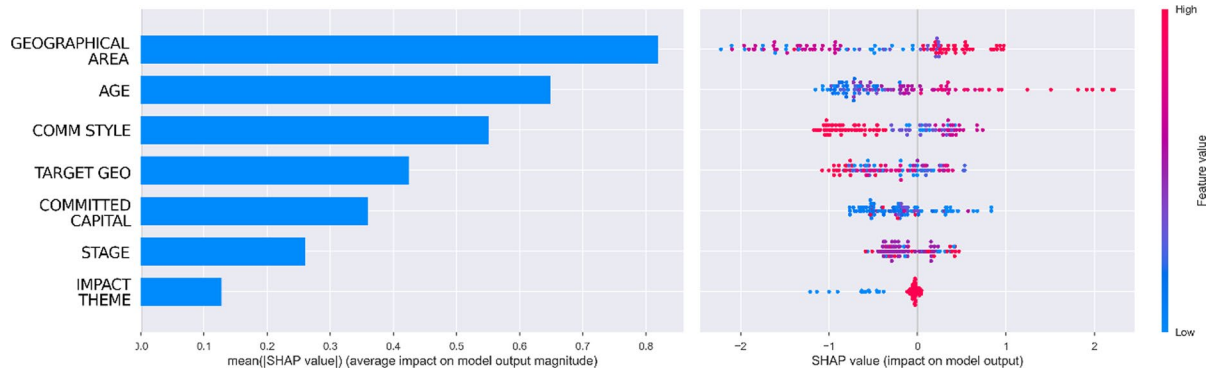

Fig. 7 Feature importance is measured through Shapley values

negative impact on the probability of being classified as a high-traffic website. After this threshold, SHAP values become positive with higher age values (i.e., for older funds). The older the fund, the higher the probability of having a high-traffic website, probably due to the time necessary to garner backlinks and better indexing from search engines. The effect of committed capital is mixed and has a smaller impact on model predictions, as does the impact on theme and stage of development. Being focused on terms of impact theme does not seem to improve predictions of website traffic, with SIVCs devoted to the finance theme having the lowest SHAP values. SIVCs located in Africa and Europe are also penalized, whereas those in North America have a higher probability of having more page views. This result could be partially explained by the higher experience of webmasters in designing websites and optimizing them for search engines or by the greater resources available to North America SIVCs to spend on digital marketing and website optimization. However, the websites of SIVCs that target North America are more likely to fall outside of the top rank. This may be because funds addressing social issues in developed countries are less remarkable than funds targeting geographical areas requiring greater urgency to resolve social issues. Finally, our classification of SIVCs in terms of communication strategies proved important for predictions. Being a *Blabber* has a strong impact and is indicative of a website with fewer visitors; by contrast, being a *Smart Hero* increases the probability of being in the top rank, as does being a *Naïve Dreamer*. This is a signal that social themes are attractive.

8 Conclusion

In recent decades, social enterprises and impact investors have emerged as an interesting market category—one that combines multiple organizational forms and institutional logics to create social value while generating economic returns. The hybrid nature of these actors is rooted in the concept of "blended value" (Battilana & Dorado, 2010; Battilana & Lee, 2014), which captures the idea that value is an indivisible integration of economic, social, and environmental returns from investments (Bugg-Levine & Emerson, 2011; Emerson, 2003).

In the domain of venture capital, these enterprises—known as SIVCs—resemble traditional VC investors in their governance structures and investment strategies but look for investments that emphasize social value with the aim of optimizing both financial and social outcomes (Barber et al., 2021; Hong & Kostovetsky, 2012). Unfortunately, research in this field has been dogged by the early-stage nature of the phenomenon, coupled with the difficulties in correctly defining and conceptualizing social impact. Indeed, the lack of universally accepted boundaries for the social impact concept has limited research mostly to the practitioner and press level. It is now time to call for new studies that leverage interdisciplinary approaches to explore the dynamics, complexities, and heterogeneity of the landscape inhabited by these new financial players.

In the present work, we have drawn from research on organizational identity, communication, and

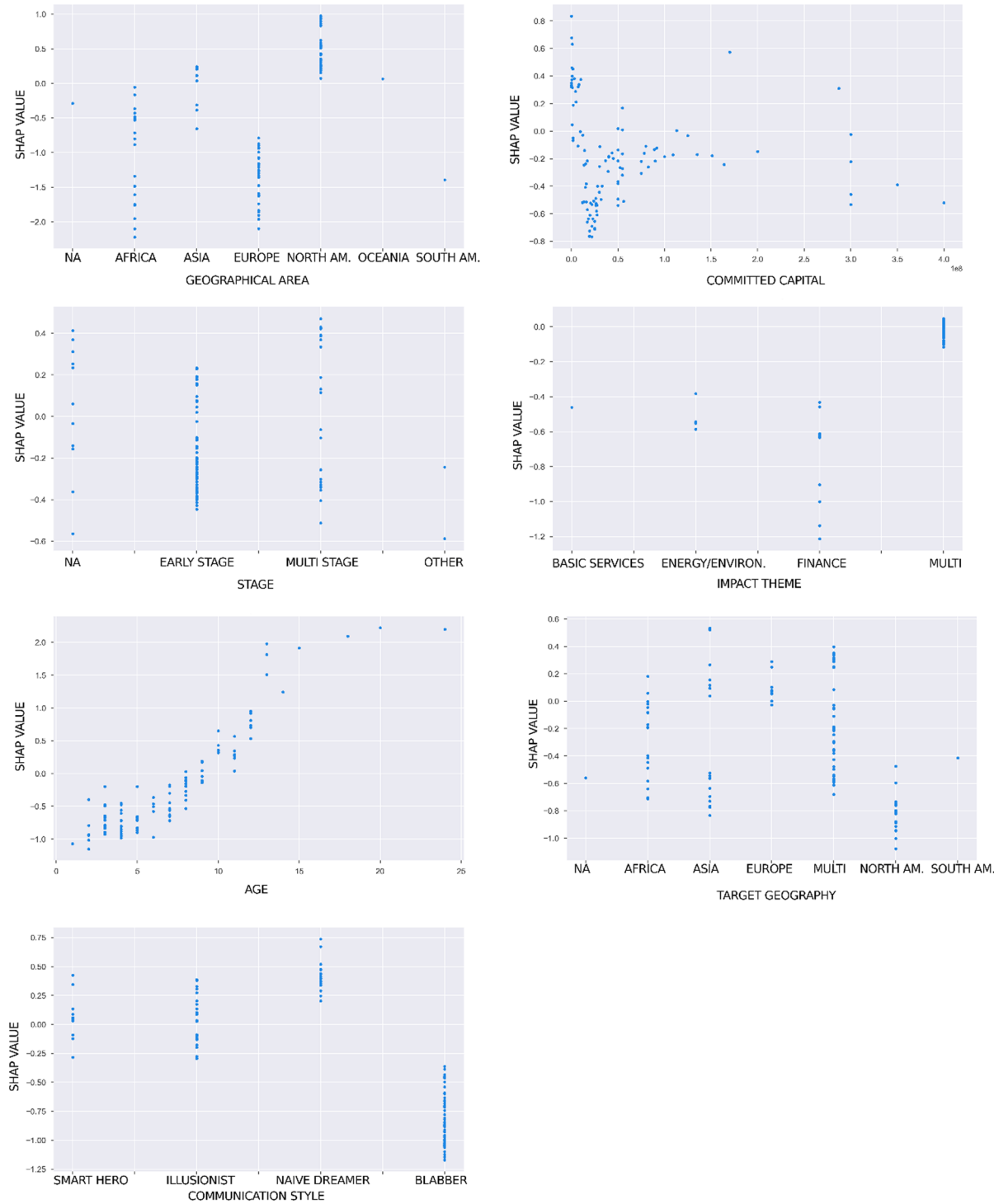

Fig. 8 Feature contribution to model predictions

language to speculate on how SIVCs use language to manage the complexity of their hybrid identity. In the emergent new field of social impact investing, the presence of information asymmetries and the lack of definitional and conceptual clarity about the concept of impact investing increase the uncertainty surrounding this new paradigm in finance.

To minimize stakeholder skepticism and ambiguity, social enterprises need to skillfully use language to communicate an identity (Chandra, 2014). Indeed, it is widely understood that organizations rely on language to convey their identity as well as influence the perceptions of others (Lounsbury & Glynn, 2001; Martens et al., 2007). Thus, we analyzed language from the investor perspective to complement traditional studies on identity for entrepreneurial ventures. By decreasing the perceived uncertainty about their identity, SIVCs may bolster a "positive sorting" match with potential investees.

Notably, we introduced two dimensions that respectively measure (i) the strength of the social positioning and (ii) the distinctiveness of the language. These two factors are particularly relevant in the context of social hybrid organizations, as they can reduce uncertainty and equivocality in the audience's information processing (Daft & Lengel, 1986): the former is critical for creating a social identity (Eckert, 2000), while the latter makes organizations distinctive in respect to their peers (Navis & Glynn, 2011). Combining these two linguistic characteristics is especially relevant in an emerging field such as social finance, where investors can use different emphases to shape their social identity and face high competition due to the low barriers to entry.

Using text mining techniques, we analyzed a sample of 195 SIVC websites. From this, we proposed a categorization of SIVCs according to their linguistic orientation that captures their heterogeneity in communication. A clustering of SIVCs revealed four different types of investors (*Blabbers*, *Smart Heroes*, *Naïve Dreamers*, and *Illusionists*), categorized according to the communicated intensity of their social impact theme and the distinctiveness of the language used on their websites. We also presented a topic modeling analysis to highlight the main themes covered by SIVCs on their websites. Finally, as an additional analysis, we examined website traffic to ascertain how linguistic distinctiveness and the importance attributed to the social impact theme could work to attract the attention of external audiences.

Despite its merits, the paper is not without limitations and leaves us with several unanswered questions that represent open avenues for future research.

The first limitation of our study is that we only focus on the distinctiveness of the language used. However, other aspects of linguistic style might be studied to better capture the nuances of communication. A further investigation could adopt other metrics to explore the interaction between "what" is said and "how" it is said. This could produce additional insights on how the communication coheres with different impact themes and the investor's impact mission. It would be valuable to clarify the development of SIVCs' communication strategies, which reflect how investors perceive the riskiness of financial versus social outcomes and the tension between such conflicting goals. Moreover, it would be interesting to explore if alternative communication modes supersede or hinder the action of the online communication tool. Specifically, scholars could explore the conditions under which, or the reasons why, specific language attributes guide the allocation of attention. We also hope that our findings may encourage future studies to assess the antecedents of different communication styles, the contingencies driving the effectiveness of language, and the extent to which these elements influence SIVCs' final goals.

Moreover, we cannot offer any insight into how the linguistic style used by SIVCs affects their selection process. Do SIVCs' different communication modes allow them to properly select target companies? What are the observed outcomes associated with this selection? For example, are *Smart Heroes* better than others in selecting higher-quality ventures? To what extent does the intensity of the social impact theme and the distinctiveness of the language help SIVCs balance the tension between social and financial outcomes? A future avenue for research could be to explore how different modes of communication drive SIVCs' selection process. Future research might address this issue by adopting more qualitative methods, case study approaches, or experiments. Since decision-making processes are difficult to observe and measure, scholars could leverage randomized control trials, an innovative methodology that overcomes the limitations of traditional techniques based on surveys and instrumental variables.

Additionally, we focused our study on SIVCs and thus disregarded other forms of social impact finance vehicles. However, our findings and theoretical insights might generalize to other organizational forms that leverage the power of web communication to allocate resources in the social impact arena—such as non-governmental organizations, charities, or philanthropic organizations. We encourage researchers to adopt our methods and analytical framework for future work that might compare how other investors in the social impact arena behave and communicate in order to build their own identity.

Lastly, our database is limited in size, as we could only gather information on active websites for 195 SIVCs. While this number is appropriate to carry out meaningful semantic network analysis, we encourage future research to broaden the data sources to validate our findings.

Our findings have several practical and managerial implications. The results suggest that the social impact intensity and the use of a distinctive language are two important and interrelated dimensions of language that contribute to the construction of identity in emerging market categories. Thus, we suggest that practitioners carefully manage the intentional use of language. Indeed, managers of SIVCs should be aware of the power of language and carefully elucidate the social impact they seek to achieve from the very beginning.

SIVCs operate in the private equity industry with the aim of simultaneously pursuing social and financial objectives; therefore, their communication approach has to be framed in this light. In a chaotic, dynamic, and changing context—where standards still have to be designed—it is important to create a strong social identity: one that conveys values, missions, and financing intentions in such a way that one is recognized as a member of the market category, but distinct from its competitors.

In this study, we wanted to demonstrate that the linguistic style is of paramount importance when dealing with a context characterized by ambiguous stakeholder perceptions, weakly defined boundaries, and difficult-to-measure outcomes. To the best of our knowledge, our study is the first to address this topic. However, more research is needed to better understand the role of communication in building organizational social impact identity. While we show that both the content and style of communication interact, our research is only a first step in uncovering the intricacies of those relationships.

Funding Open access funding provided by Università degli Studi di Perugia within the CRUI-CARE Agreement.

References

Acs, Z. J., Song, A. K., Szerb, L., Audretsch, D. B., & Komlósi, É. (2021). The evolution of the global digital platform economy: 1971–2021. *Small Business Economics, 57*(4), 1629–1659. https://doi.org/10.1007/s11187-021-00561-x

Agrawal, & Hockerts. (2019). Impact investing strategy: Managing conflicts between impact investor and investee social enterprise. *Sustainability, 11*(15), 4117. https://doi.org/10.3390/su11154117

Albert, S., & Whetten, D. A. (1985). Organizational identity. *Research in Organizational Behavior, 7*, 263–295.

Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *Academy of Management Review, 19*(4), 645–670. https://doi.org/10.5465/amr.1994.9412190214

Allison, T. H., Davis, B. C., Short, J. C., & Webb, J. W. (2015). Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues. *Entrepreneurship Theory and Practice, 39*(1), 53–73. https://doi.org/10.1111/etap.12108

Aral, S., & Van Alstyne, M. (2011). The diversity-bandwidth rrade-off. *American Journal of Sociology, 117*(1), 90–171. https://doi.org/10.1086/661238

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science, 57*(8), 1485–1509. https://doi.org/10.1287/mnsc.1110.1370

Arjaliès, D.-L., & Durand, R. (2019). Product categories as judgment devices: The moral awakening of the investment industry. *Organization Science, 30*(5), 885–911. https://doi.org/10.1287/orsc.2018.1267

Arts, S., Cassiman, B., & Gomez, J. C. (2018). Text matching to measure patent similarity. *Strategic Management Journal, 39*(1), 62–84. https://doi.org/10.1002/smj.2699

Ashforth, B. E., Rogers, K. M., & Corley, K. G. (2011). Identity in organizations: Exploring cross-level dynamics. *Organization Science, 22*(5), 1144–1156. https://doi.org/10.1287/orsc.1100.0591

Austin, J., Stevenson, H., & Wei-Skillern, J. (2006). Social and commercial entrepreneurship: Same, different, or both? *Entrepreneurship Theory and Practice, 30*(1), 1–22. https://doi.org/10.1111/j.1540-6520.2006.00107.x

Balmer, J. M. T., & Greyser, S. A. (2002). Managing the multiple identities of the corporation. *California Management Review, 44*(3), 72–86. https://doi.org/10.2307/41166133

Barber, B. M., Morse, A., & Yasuda, A. (2021). Impact investing. *Journal of Financial Economics, 139*(1), 162–185. https://doi.org/10.3386/w26582

Bart, C. K., Bontis, N., & Taggar, S. (2001). A model of the impact of mission statements on firm performance. *Management Decision, 39*(1), 19–35. https://doi.org/10.1108/EUM0000000005404

Battilana, J., & Dorado, S. (2010). Building sustainable hybrid organizations: The case of commercial microfinance organizations. *Academy of Management Journal, 53*(6), 1419–1440. https://doi.org/10.5465/amj.2010.57318391

Battilana, J., & Lee, M. (2014). Advancing research on hybrid organizing – insights from the study of social enterprises. *The Academy of Management Annals, 8*(1), 397–441. https://doi.org/10.5465/19416520.2014.893615

Benford, R. D., & Snow, D. A. (2000). Framing processes and social movements: An overview and assessment. *Annual Review of Sociology, 26*(1), 611–639. https://doi.org/10.1146/annurev.soc.26.1.611

Bertoni, F., D'Adda, D., & Grilli, L. (2016). Cherry-picking or frog-kissing? A theoretical analysis of how investors select entrepreneurial ventures in thin venture capital markets. *Small Business Economics, 46*(3), 391–405. https://doi.org/10.1007/s11187-015-9690-9

Bettis, R. A., Gambardella, A., Helfat, C., & Mitchell, W. (2015). Qualitative empirical research in strategic management. *Strategic Management Journal, 36*(5), 637–639. https://doi.org/10.1002/smj.2317

Block, J. H., Hirschmann, M., & Fisch, C. (2021). Which criteria matter when impact investors screen social enterprises? *Journal of Corporate Finance, 66*, 101813. https://doi.org/10.1016/j.jcorpfin.2020.101813

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bocken, N. M. P. (2015). Sustainable venture capital – catalyst for sustainable start-up success? *Journal of Cleaner Production, 108*, 647–658. https://doi.org/10.1016/j.jclepro.2015.05.079

Boje, D. M., Oswick, C., & Ford, J. D. (2004). Introduction to special topic forum: Language and organization: The doing of discourse. *The Academy of Management Review, 29*(4), 571. https://doi.org/10.2307/20159071

Bolino, M. C., Kacmar, K. M., Turnley, W. H., & Gilstrap, J. B. (2008). A multi-level review of impression management motives and behaviors. *Journal of Management, 34*(6), 1080–1109. https://doi.org/10.1177/0149206308324325

Boswijk, V., & Coler, M. (2020). What is salience? *Open Linguistics, 6*(1), 713–722. https://doi.org/10.1515/opli-2020-0042

Botero, I. C., Thomas, J., Graves, C., & Fediuk, T. A. (2013). Understanding multiple family firm identities: An exploration of the communicated identity in official websites. *Journal of Family Business Strategy, 4*(1), 12–21. https://doi.org/10.1016/j.jfbs.2012.11.004

Bradac, James J., Bowers, John Waite, & Courtright, John A. (1979). Three language variables in communication research: Intensity, immediacy, and diversity. *Human Communication Research, 5*(3), 257–269. https://doi.org/10.1111/j.1468-2958.1979.tb00639.x

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology, 25*(2), 163–177. https://doi.org/10.1080/0022250X.2001.9990249

Brickson, S. L. (2005). Organizational identity orientation: Forging a link between organizational identity and organizations' relations with stakeholders. *Administrative Science Quarterly, 50*(4), 576–609. https://doi.org/10.2189/asqu.50.4.576

Brown, T. J. (2006). Identity, intended image, construed image, and reputation: An interdisciplinary framework and suggested terminology. *Journal of the Academy of Marketing Science, 34*(2), 99–106. https://doi.org/10.1177/0092070305284969

Burns, B. L., Barney, J. B., Angus, R. W., & Herrick, H. N. (2016). Enrolling stakeholders under conditions of risk and uncertainty. *Strategic Entrepreneurship Journal, 10*(1), 97–106. https://doi.org/10.1002/sej.1209

Busch, T., Bruce-Clark, P., Derwall, J., Eccles, R., Hebb, T., Hoepner, A., Klein, C., Krueger, P., Paetzold, F., Scholtens, B., & Weber, O. (2021). Impact investments: a call for (re)orientation. *SN Business & Economics, 1*(2), 33. https://doi.org/10.1007/s43546-020-00033-6

Byrne, D. (1971). *The attraction paradigm*. Academic Press.

Calderini, M., Chiodo, V., & Michelucci, F. V. (2018). The social impact investment race: toward an interpretative framework. *European Business Review, 30*(1), 66–81. https://doi.org/10.1108/EBR-10-2016-0134

Cao, Z., Li, G., & Song, H. (2017). Modelling the interdependence of tourism demand: The global vector autoregressive approach. *Annals of Tourism Research, 67*, 1–13. https://doi.org/10.1016/j.annals.2017.07.019

Cetindamar, D., & Ozkazanc-Pan, B. (2017). Assessing mission drift at venture capital impact investors. *Business Ethics: A European Review, 26*(3), 257–270. https://doi.org/10.1111/beer.12149

Chandra, Y. (2014). Toward a meta-linguistic model of social entrepreneurship: Insights from computational linguistics. *Academy of Management Proceedings, 2014*(1), 13356. https://doi.org/10.5465/ambpp.2014.13356abstract

Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal, 35*(1), 1–23. https://doi.org/10.1002/smj.2131

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal, 40*(11), 1705–1732. https://doi.org/10.1002/smj.3067

Chreim, S. (2005). The continuity-change duality in narrative texts of organizational identity*. *Journal of Management*

Studies, 42(3), 567–593. https://doi.org/10.1111/j.1467-6486.2005.00509.x

Clarke, J., & Cornelissen, J. (2011). Language, communication, and socially situated cognition in entrepreneurship. *Academy of Management Review, 36*(4), 776–778. https://doi.org/10.5465/amr.2011.0192

Clegg, S. R., Rhodes, C., & Kornberger, M. (2007). Desperately seeking legitimacy: Organizational identity and emerging industries. *Organization Studies, 28*(4), 495–513. https://doi.org/10.1177/0170840606067995

Corley, K. G., & Gioia, D. (2004). Identity ambiguity and change in the wake of a corporate spin-off. *Administrative Science Quarterly, 49*(2), 173–208. https://doi.org/10.2307/4131471

Corley, K. G., Harquail, C. V., Pratt, M. G., Glynn, M. A., Fiol, C. M., & Hatch, M. J. (2006). Guiding organizational identity through aged adolescence. *Journal of Management Inquiry, 15*(2), 85–99. https://doi.org/10.1177/1056492605285930

Cui, G., Wong, M. L., & Lui, H.-K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science, 52*(4), 597–612. https://doi.org/10.1287/mnsc.1060.0514

Czarniawska, B., & Wolff, R. (1998). Constructing new identities in established organization fields. *International Studies of Management & Organization, 28*(3), 32–56. https://doi.org/10.1080/00208825.1998.11656739

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science, 53*(9), 1375–1388. https://doi.org/10.1287/mnsc.1070.0704

Downing, S. (2005). The social construction of entrepreneurship: Narrative and dramatic processes in the coproduction of organizations and identities. *Entrepreneurship Theory and Practice, 29*(2), 185–204. https://doi.org/10.1111/j.1540-6520.2005.00076.x

Durand, R., & Khaire, M. (2017). Where do market categories come from and how? Distinguishing category creation from category emergence. *Journal of Management, 43*(1), 87–110. https://doi.org/10.1177/0149206316669812

Durand, R., & Paolella, L. (2013). Category stretching: Reorienting research on categories in strategy, entrepreneurship, and organization theory. *Journal of Management Studies, 50*(6), 1100–1123. https://doi.org/10.1111/j.1467-6486.2011.01039.x

Dutton, J. E., Dukerich, J. M., & Harquail, C. V. (1994). Organizational images and member identification. *Administrative Science Quarterly, 39*(2), 239. https://doi.org/10.2307/2393235

Eckhardt, J. T., Shane, S., & Delmar, F. (2006). Multistage selection and the financing of new ventures. *Management Science, 52*(2), 220–232. https://doi.org/10.1287/mnsc.1050.0478

Edwards, J. (2012). *Language and Identity*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0599

Ellis, N. C. (2016). Salience, cognition, language complexity, and complex adaptive systems. *Studies in Second Language Acquisition, 38*(2), 341–351. https://doi.org/10.1017/S027226311600005X

Emerson, J. (2003). The blended value proposition: Integrating social and financial returns. *California Management Review, 45*(4), 35–51. https://doi.org/10.2307/41166187

Findlay, S., & Moran, M. (2019). Purpose-washing of impact investing funds: motivations, occurrence and prevention. *Social Responsibility Journal, 15*(7), 853–873. https://doi.org/10.1108/SRJ-11-2017-0260

Fiol, C. M. (2002). Capitalizing on paradox: The role of language in transforming organizational identities. *Organization Science, 13*(6), 653–666. https://doi.org/10.1287/orsc.13.6.653.502

Fisch, C., Masiak, C., Vismara, S., & Block, J. (2021). Motives and profiles of ICO investors. *Journal of Business Research, 125*, 564–576. https://doi.org/10.1016/j.jbusres.2019.07.036

Fosfuri, A., Giarratana, M. S., & Roca, E. (2016). Social business hybrids: Demand externalities, competitive advantage, and growth through diversification. *Organization Science, 27*(5), 1275–1289. https://doi.org/10.1287/orsc.2016.1080

FronzettiColladon, A. (2018). The semantic brand score. *Journal of Business Research, 88*, 150–160. https://doi.org/10.1016/j.jbusres.2018.03.026

FronzettiColladon, A., & Naldi, M. (2020). Distinctiveness centrality in social networks. *PLOS ONE, 15*(5), e0233276. https://doi.org/10.1371/journal.pone.0233276

Gatti, M. C. (2011). The language of competence in corporate histories for company websites. *Journal of Business Communication, 48*(4), 482–502. https://doi.org/10.1177/0021943611414543

Geczy, C., Jeffers, J. S., Musto, D. K., & Tucker, A. M. (2021). Contracts with (Social) benefits: The implementation of impact investing. *Journal of Financial Economics, in press.* https://doi.org/10.1016/j.jfineco.2021.01.006

Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation. *Strategic Management Journal, 12*(6), 433–448. https://doi.org/10.1002/smj.4250120604

Gioia, D. A., Patvardhan, S. D., Hamilton, A. L., & Corley, K. G. (2013). Organizational Identity Formation and Change. *The Academy of Management Annals, 7*(1), 123–193. https://doi.org/10.1080/19416520.2013.762225

Gioia, D. A., Price, K. N., Hamilton, A. L., & Thomas, J. B. (2010). Forging an identity: An insider-outsider study of processes involved in the formation of organizational identity. *Administrative Science Quarterly, 55*(1), 1–46. https://doi.org/10.2189/asqu.2010.55.1.1

Giora, R. (2003). On our mindsalience, context, and figurative language. *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780195136166.001.0001

Glaser, V., Fiss, P. C., & Kennedy, M. T. (2011). Rhetoric and resonance: framing strategies for institutionalizing new market conceptions. *Academy of Management Proceedings, 2011*(1), 1–6. https://doi.org/10.5465/ambpp.2011.65869627

Goldfarb, B., Kirsch, D., & Miller, D. (2007). Was there too little entry during the Dot Com Era?☆. *Journal of*

Financial Economics, 86(1), 100–144. https://doi.org/10.1016/j.jfineco.2006.03.009

Graffin, S. D., Carpenter, M. A., & Boivie, S. (2011). What's all that (strategic) noise? anticipatory impression management in CEO succession. *Strategic Management Journal, 32*(7), 748–770. https://doi.org/10.1002/smj.906

Grohs, R., Raies, K., Koll, O., & Mühlbacher, H. (2016). One pie, many recipes: Alternative paths to high brand strength. *Journal of Business Research, 69*(6), 2244–2251. https://doi.org/10.1016/j.jbusres.2015.12.037

Gustafsson, R., Jääskeläinen, M., Maula, M., & Uotila, J. (2016). Emergence of industries: A Review and Future Directions. *International Journal of Management Reviews, 18*(1), 28–50. https://doi.org/10.1111/ijmr.12057

Haans, R. F. J. (2019). What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories. *Strategic Management Journal, 40*(1), 3–27. https://doi.org/10.1002/smj.2978

Hallen, B. L., Bingham, C. B., & Cohen, S. (2014). Do accelerators accelerate? A study of venture accelerators as a path to success? *Academy of Management Proceedings, 2014*(1), 12955. https://doi.org/10.5465/ambpp.2014.185

Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals, 13*(2), 586–632. https://doi.org/10.5465/annals.2017.0099

Hansen, J., & Wänke, M. (2010). Truth from language and truth from fit: The impact of linguistic concreteness and level of construal on subjective truth. *Personality and Social Psychology Bulletin, 36*(11), 1576–1588. https://doi.org/10.1177/0146167210386238

Hatch, M. J., & Schultz, M. (2002). The dynamics of organizational identity. *Human Relations, 55*(8), 989–1018. https://doi.org/10.1177/0018726702055008181

Hawn, O., Chatterji, A. K., & Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). *Strategic Management Journal, 39*(4), 949–976. https://doi.org/10.1002/smj.2752

Hebb, T. (2013). Impact investing and responsible investing: what does it mean? *Journal of Sustainable Finance & Investment, 3*(2), 71–74. https://doi.org/10.1080/20430795.2013.776255

Hehenberger, L., Mair, J., & Metz, A. (2019). The assembly of a field ideology: An idea-centric perspective on systemic power in impact investing. *Academy of Management Journal, 62*(6), 1672–1704. https://doi.org/10.5465/amj.2017.1402

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *Journal of Business Communication, 45*(4), 363–407. https://doi.org/10.1177/0021943608319388

Höchstädter, A. K., & Scheck, B. (2015). What's in a name: An analysis of impact investing understandings by academics and practitioners. *Journal of Business Ethics, 132*(2), 449–475. https://doi.org/10.1007/s10551-014-2327-0

Hong, H., & Kostovetsky, L. (2012). Red and blue investing: Values and finance. *Journal of Financial Economics, 103*(1), 1–19. https://doi.org/10.1016/j.jfineco.2011.01.006

Hsu, G., & Hannan, M. T. (2005). Identities, genres, and organizational forms. *Organization Science, 16*(5), 474–490. https://doi.org/10.1287/orsc.1050.0151

Huang, L., & Knight, A. P. (2017). Resources and relationships in entrepreneurship: An exchange theory of the development and effects of the entrepreneur-investor relationship. *Academy of Management Review, 42*(1), 80–102. https://doi.org/10.5465/amr.2014.0397

Ioannou, I., & Serafeim, G. (2015). The impact of corporate social responsibility on investment recommendations: Analysts' perceptions and shifting institutional logics. *Strategic Management Journal, 36*(7), 1053–1081. https://doi.org/10.1002/smj.2268

Ireland, R. D., & Hirc, M. A. (1992). Mission statements: Importance, challenge, and recommendations for development. *Business Horizons, 35*(3), 34–42. https://doi.org/10.1016/0007-6813(92)90067-J

Iwashita, H. (2022). Language and identity in the shadow: A multi-case study of a Japanese multinational corporation. *International Business Review, 31*(2), 101913. https://doi.org/10.1016/j.ibusrev.2021.101913

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal, 36*(10), 1435–1457. https://doi.org/10.1002/smj.2294

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

Kent, M. L., & Taylor, M. (1998). Building dialogic relationships through the world wide web. *Public Relations Review, 24*(3), 321–334. https://doi.org/10.1016/S0363-8111(99)80143-X

Kerr, W. R., Lerner, J., & Schoar, A. (2014). The consequences of entrepreneurial finance: Evidence from angel financings. *Review of Financial Studies, 27*(1), 20–55. https://doi.org/10.1093/rfs/hhr098

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics, 133*(1), 237–293. https://doi.org/10.1093/qje/qjx032

Lamont, M., & Molnár, V. (2002). The study of boundaries in the social sciences. *Annual Review of Sociology, 28*(1), 167–195. https://doi.org/10.1146/annurev.soc.28.110601.141107

Lancichinetti, A., Sirer, M. I., Wang, J. X., Acuna, D., Körding, K., & Amaral, L. A. N. (2015). High-reproducibility and high-accuracy method for automated topic classification. *Physical Review X, 5*(1), 011007. https://doi.org/10.1103/PhysRevX.5.011007

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research, 39*(1), 19–37. https://doi.org/10.1080/00909882.2010.536844

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media:

Evidence from Facebook. *Management Science, 64*(11), 5105–5131. https://doi.org/10.1287/mnsc.2017.2902

Lee, M., Adbi, A., & Singh, J. (2020). Categorical cognition and outcome efficiency in impact investing decisions. *Strategic Management Journal, 41*(1), 86–107. https://doi.org/10.1002/smj.3096

Lehner, O. M., & Nicholls, A. (2014). Social finance and crowdfunding for social enterprises: a public–private case study providing legitimacy and leverage. *Venture Capital, 16*(3), 271–286. https://doi.org/10.1080/13691066.2014.925305

Livengood, R. S., & Reger, R. K. (2010). That's our turf! Identity domains and competitive dynamics. *Academy of Management Review, 35*(1), 48–66. https://doi.org/10.5465/amr.35.1.zok48

Lounsbury, M., & Glynn, M. A. (2001). Cultural entrepreneurship: stories, legitimacy, and the acquisition of resources. *Strategic Management Journal, 22*(6–7), 545–564. https://doi.org/10.1002/smj.188

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D.K.-W., Newman, S.-F., Kim, J., & Lee, S.-I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering, 2*(10), 749–760. https://doi.org/10.1038/s41551-018-0304-0

Luo, J., & Kaul, A. (2019). Private action in public interest: The comparative governance of social issues. *Strategic Management Journal, 40*(4), 476–502. https://doi.org/10.1002/smj.2961

Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. *Academy of Management Annals, 8*(1), 57–125. https://doi.org/10.5465/19416520.2014.873177

Manning, S., & Bejarano, T. A. (2017). Convincing the crowd: Entrepreneurial storytelling in crowdfunding campaigns. *Strategic Organization, 15*(2), 194–219. https://doi.org/10.1177/1476127016648500

Martens, M. L., Jennings, J. E., & Jennings, P. D. (2007). Do the stories they tell get them the money they need? The role of entrepreneurial narratives in resource acquisition. *Academy of Management Journal, 50*(5), 1107–1132. https://doi.org/10.5465/amj.2007.27169488

Mcpherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mead, A. (1992). Review of the development of multidimensional scaling methods. *The Statistician, 41*(1), 27. https://doi.org/10.2307/2348634

Miller, C. H., Lane, L. T., Deatrick, L. M., Young, A. M., & Potts, K. A. (2007). Psychological reactance and promotional health messages: The effects of controlling language, lexical concreteness, and the restoration of freedom. *Human Communication Research, 33*(2), 219–240. https://doi.org/10.1111/j.1468-2958.2007.00297.x

Miller, T. L., & Wesley, C. L., II. (2010). Assessing mission and resources for social change: An organizational identity perspective on social venture capitalists' decision criteria. *Entrepreneurship Theory and Practice, 34*(4), 705–733. https://doi.org/10.1111/j.1540-6520.2010.00388.x

Moore, M.-L., Westley, F. R., & Nicholls, A. (2012). The social finance and social innovation nexus 1. *Journal of Social Entrepreneurship, 3*(2), 115–132. https://doi.org/10.1080/19420676.2012.725824

Moss, T. W., Renko, M., Block, E., & Meyskens, M. (2018). Funding the story of hybrid ventures: Crowdfunder lending preferences and linguistic hybridity. *Journal of Business Venturing, 33*(5), 643–659. https://doi.org/10.1016/j.jbusvent.2017.12.004

Nag, R., Corley, K. G., & Gioia, D. A. (2007). The intersection of organizational identity, knowledge, and practice: Attempting strategic change via knowledge grafting. *Academy of Management Journal, 50*(4), 821–847. https://doi.org/10.5465/amj.2007.26279173

Navis, C., & Glynn, M. A. (2010). How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. *Administrative Science Quarterly, 55*(3), 439–471. https://doi.org/10.2189/asqu.2010.55.3.439

Navis, C., & Glynn, M. A. (2011). Legitimate distinctiveness and the entrepreneurial identity: influence on investor judgments of new venture plausibility. *Academy of Management Review, 36*(3), 479–499. https://doi.org/10.5465/amr.2008.0361

Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science, 31*(3), 521–543. https://doi.org/10.1287/mksc.1120.0713

Nicholls, A. (2010). The Institutionalization of Social Investment: The Interplay of Investment Logics and Investor Rationalities. *Journal of Social Entrepreneurship, 1*(1), 70–100. https://doi.org/10.1080/19420671003701257

Nuscheler, D., Engelen, A., & Zahra, S. A. (2019). The role of top management teams in transforming technology-based new ventures' product introductions into growth. *Journal of Business Venturing, 34*(1), 122–140. https://doi.org/10.1016/j.jbusvent.2018.05.009

O'Connor, E. (2002). Storied Business: Typology, Intertextuality, and Traffic in Entrepreneurial Narrative. *Journal of Business Communication, 39*(1), 36–54. https://doi.org/10.1177/002194360203900103

O'Gorman, C., & Doran, R. (1999). Mission statements in small and medium-sized businesses. *Journal of Small Business Management, 37*(4), 59–66.

Ormiston, J., Charlton, K., Donald, M. S., & Seymour, R. G. (2015). Overcoming the challenges of impact investing: Insights from leading investors. *Journal of Social Entrepreneurship, 6*(3), 352–378. https://doi.org/10.1080/19420676.2015.1049285

Pache, A.-C., & Santos, F. (2010). When worlds collide: The internal dynamics of organizational responses to conflicting institutional demands. *Academy of Management Review, 35*(3), 455–476. https://doi.org/10.5465/amr.35.3.zok455

Packard, G., & Berger, J. (2017). How language shapes word of mouth's impact. *Journal of Marketing Research, 54*(4), 572–588. https://doi.org/10.1509/jmr.15.0248

Pan, L., Li, X., Chen, J., & Chen, T. (2020). Sounds novel or familiar? Entrepreneurs' framing strategy in the venture capital market. *Journal of Business Venturing, 35*(2), 105930. https://doi.org/10.1016/j.jbusvent.2019.02.003

Pan, L., McNamara, G., Lee, J. J., Haleblian, J., & (John), and Devers, C. E. (2018). Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions. *Strategic Management Journal, 39*(8), 2204–2225. https://doi.org/10.1002/smj.2733

Parhankangas, A., & Renko, M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing, 32*(2), 215–236. https://doi.org/10.1016/j.jbusvent.2016.11.001

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Périlleux, A. (2015). When social enterprises engage in finance: Agents of change in lending relationships, a Belgian typology. *Strategic Change, 24*(3), 285–300. https://doi.org/10.1002/jsc.2009

Petkova, A. P., Rindova, V. P., & Gupta, A. K. (2013). No news is bad news: Sensegiving activities, media attention, and venture capital funding of new technology organizations. *Organization Science, 24*(3), 865–888. https://doi.org/10.1287/orsc.1120.0759

Pontikes, E. G. (2012). Two sides of the same coin. *Administrative Science Quarterly, 57*(1), 81–118. https://doi.org/10.1177/0001839212446689

Pratt, M. G., & Foreman, P. O. (2000). Classifying managerial responses to multiple organizational identities. *The Academy of Management Review, 25*(1), 18. https://doi.org/10.2307/259261

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 31* (pp. 6638–6648). Neural Information Processing Systems Foundation.

Randjelovic, J., O'Rourke, A. R., & Orsato, R. J. (2003). The emergence ofgreen venture capital. *Business Strategy and the Environment, 12*(4), 240–253. https://doi.org/10.1002/bse.361

Revelli, C., & Viviani, J.-L. (2015). Financial performance of socially responsible investing (SRI): what have we learned? A meta-analysis. *Business Ethics: A European Review, 24*(2), 158–185. https://doi.org/10.1111/beer.12076

Riley, T. J., & Luippold, B. L. (2015). Managing investors' perception through strategic word choices in financial narratives. *Journal of Corporate Accounting & Finance, 26*(5), 57–62. https://doi.org/10.1002/jcaf.22064

Rindova, V. P., Petkova, A. P., & Kotha, S. (2007). Standing out: how new firms in emerging markets build reputation. *Strategic Organization, 5*(1), 31–70. https://doi.org/10.1177/1476127006074389

Rizzi, F., Pellegrini, C., & Battaglia, M. (2018). The structuring of social finance: Emerging approaches for supporting environmentally and socially impactful projects. *Journal of Cleaner Production, 170*, 805–817. https://doi.org/10.1016/j.jclepro.2017.09.167

Rosa, J. A., Porac, J. F., Runser-Spanjol, J., & Saxon, M. S. (1999). Sociocognitive dynamics in a product market. *Journal of Marketing, 63*, 64. https://doi.org/10.2307/1252102

Rosemeyer, M., & Schwenter, S. A. (2019). Entrenchment and persistence in language change: the Spanish past subjunctive. *Corpus Linguistics and Linguistic Theory, 15*(1), 167–204. https://doi.org/10.1515/cllt-2016-0047

Scott, S. G., & Lane, V. R. (2000). A stakeholder approach to organizational identity. *The Academy of Management Review, 25*(1), 43. https://doi.org/10.2307/259262

Short, J. C., & Palmer, T. B. (2008). The application of DICTION to content analysis research in strategic management. *Organizational Research Methods, 11*(4), 727–752. https://doi.org/10.1177/1094428107304534

Smith, B. R., & Bergman, B. J. (2020). The other side of the coin: Investor identity and its role in resource provision. *Journal of Business Venturing Insights, 14*, e00175. https://doi.org/10.1016/j.jbvi.2020.e00175

Sørensen, M. (2007). How smart is smart money? A two-sided matching model of venture capital. *The Journal of Finance, 62*(6), 2725–2762. https://doi.org/10.1111/j.1540-6261.2007.01291.x

Stigliani, I., & Elsbach, K. D. (2018). Identity co-formation in an emerging industry: Forging organizational distinctiveness and industry coherence through sensemaking and sensegiving. *Journal of Management Studies, 55*(8), 1323–1355. https://doi.org/10.1111/joms.12403

Taeuscher, K., Bouncken, R., & Pesch, R. (2021). Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms. *Academy of Management Journal, 64*(1), 149–179. https://doi.org/10.5465/amj.2018.0620

Taeuscher, K., & Rothe, H. (2021). Optimal distinctiveness in platform markets: Leveraging complementors as legitimacy buffers. *Strategic Management Journal, 42*(2), 435–461. https://doi.org/10.1002/smj.3229

Tannen, D. (1995). The power of talk: Who gets heard and why. *Harvard Business Review, 73*(5), 138–148.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Taylor, M., Kent, M. L., & White, W. J. (2001). How activist organizations are using the Internet to build relationships. *Public Relations Review, 27*(3), 263–284. https://doi.org/10.1016/S0363-8111(01)00086-8

ter Doest, L., Semin, G. R., & Sherman, S. J. (2002). Linguistic context and social perception. *Journal of Language and Social Psychology, 21*(3), 195–229. https://doi.org/10.1177/0261927X02021003001

Thakur, A., Sangal, L., & A., & Bindra, H. (2011). Quantitative measurement and comparison of effects of various search engine optimization parameters on Alexa traffic rank. *International Journal of Computer Applications, 26*(5), 15–23. https://doi.org/10.5120/3100-4257

Thibodeau, P. H., & Boroditsky, L. (2013). Natural language metaphors covertly influence reasoning. *PLoS ONE, 8*(1), e52961. https://doi.org/10.1371/journal.pone.0052961

Toma, C. L., & D'Angelo, J. D. (2015). Tell-tale words. *Journal of Language and Social Psychology, 34*(1), 25–45. https://doi.org/10.1177/0261927X14554484

Truong, Y., & Nagy, B. G. (2021). Nascent ventures' green initiatives and angel investor judgments of legitimacy and funding. *Small Business Economics, 57*(4), 1801–1818. https://doi.org/10.1007/s11187-020-00373-5

Vismara, S. (2016). Equity retention and social network theory in equity crowdfunding. *Small Business Economics, 46*(4), 579–590. https://doi.org/10.1007/s11187-016-9710-4

Whetten, D. A. (2006). Albert and whetten revisited: Strengthening the concept of organizational identity. *Journal of Management Inquiry, 15*(3), 219–234. https://doi.org/10.1177/1056492606291200

Wirtz, J. G., & Zimbres, T. M. (2018). A systematic analysis of research applying 'principles of dialogic communication' to organizational websites, blogs, and social media: Implications for theory and practice. *Journal of Public Relations Research, 30*(1–2), 5–34. https://doi.org/10.1080/1062726X.2018.1455146

Wry, T., Lounsbury, M., & Glynn, M. A. (2011). Legitimating nascent collective identities: Coordinating cultural entrepreneurship. *Organization Science, 22*(2), 449–463. https://doi.org/10.1287/orsc.1100.0613

Yan, S., Ferraro, F., Almandoz, J., & (John). (2019). The rise of socially responsible investment funds: The paradoxical role of the financial logic. *Administrative Science Quarterly, 64*(2), 466–501. https://doi.org/10.1177/0001839218773324

Younger, S., & Fisher, G. (2020). The exemplar enigma: New venture image formation in an emergent organizational category. *Journal of Business Venturing, 35*(1), 105897. https://doi.org/10.1016/j.jbusvent.2018.09.002

Zhao, E. Y., Fisher, G., Lounsbury, M., & Miller, D. (2017). Optimal distinctiveness: Broadening the interface between institutional theory and strategic management. *Strategic Management Journal, 38*(1), 93–113. https://doi.org/10.1002/smj.2589

Battilana, J., Besharov, M., & Mitzinneck, B. (2017). On hybrids and hybrid organizing: A review and roadmap for future research. In *The Sage handbook of organizational institutionalism* (pp. 132–169). SAGE Publications.

Baumeister, R. F. (1998). The self. In *Handbook of social psychology* (4th ed., pp. 680–740). McGraw-Hill.

Bishop, D. G., Treviño, L. K., Gioia, D., & Kreiner, G. E. (2019). Leveraging a recessive narrative to transform Joe Paterno's image: Media sensebreaking, sensemaking, and sensegiving during scandal. *Academy of Management Discoveries*, amd.2019.0108. https://doi.org/10.5465/amd.2019.0108

Boni L, Toschi L, Fini R (2021) Investors' aspirations toward social impact: A portfolio-based analysis. *Sustainability 13*(9), 5293. https://doi.org/10.3390/su13095293

Brown, A., & Swersky, A. (2012). *The first billion: A forecast of social investment demand*.

Bugg-Levine, A., & Emerson, J. (2011). Impact investing: Transforming how we make money while making a difference. *Innovations: Technology, Governance, Globalization*, *6*(3), 9–18. https://doi.org/10.1162/INOV_a_00077

Chen, S., & Harrison, R. (2020). Beyond profit vs. purpose: Transactional-relational practices in impact investing. *Journal of Business Venturing Insights*, *14*, e00182. https://doi.org/10.1016/j.jbvi.2020.e00182

Cutolo, D., Ferriani, S., & Cattani, G. (2020). Tell me your story and i will tell your sales: A topic model analysis of narrative style and firm performance on etsy. In *Aesthetics and Style in Strategy* (pp. 103–122). Emerald Group Publishing Limited. https://doi.org/10.1108/S0742-332220200000042005

De Fina, A. (2019). Discourse and identity. In *The Encyclopedia of Applied Linguistics* (pp. 1–8). Wiley. https://doi.org/10.1002/9781405198431.wbeal0326.pub2

Du, Q., Fan, W., Qiao, Z., Wang, G., Zhang, X., & Zhou, M. (2015). Money talks: A predictive model on crowdfunding success using project description. *Proceedings of Americas Conference on Information Systems*, 37.

Dubitzky, W., Granzow, M., & Berrar, D. (2007). Fundamentals of data mining in genomics and proteomics. In *Fundamentals of Data Mining in Genomics and Proteomics*. Springer Science + Business Media. https://doi.org/10.1007/978-0-387-47509-7

Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in belten high*. Wiley-Blackwell.

Ellis, N. C. (2017). Salience in language usage, learning and change. In M. Hundt, S. Mollin, & S. E. Pfenninger (Eds.), *The changing english language* (pp. 71–92). Cambridge University Press. https://doi.org/10.1017/9781316091746.004

Fink, L. (2018). *A sense of purpose*. Harward Law School Forum on Corporate Governance. https://corpgov.law.harvard.edu/2018/01/17/a-sense-of-purpose/

Fink, L. (2020). *A fundamental reshaping of finance*. Blackrock. https://www.blackrock.com/corporate/investor-relations/larry-fink-ceo-letter

Fisher, G. (2012). *Identity and resource mobilization in new ventures: a model of venture identification* [University of Washington]. https://digital.lib.washington.edu/researchworks/handle/1773/20622

Freireich, J., & Fulton, K. (2009). *Investing for social and environmental impact: A design catalysing an emerging industry*.

Fronzetti Colladon, A., & Grippa, F. (2020). Brand intelligence analytics. In A. Przegalinska, F. Grippa, & P. A. Gloor (Eds.), *Digital transformation of collaboration* (pp. 125–141). Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-48993-9_10

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*, *4*(7), eaaq1360. https://doi.org/10.1126/sciadv.aaq1360

Gloor, P. (2017). *Sociometrics and human relationships: Analyzing social networks to manage brands, predict trends, and improve organizational performance*. Emerald Publishing Limited.

Glynn, M. A., & Watkiss, L. (2012). Exploring cultural mechanisms of organizational identity construction. In M. Schultz, S. Maguire, A. Langley, & H. Tsoukas (Eds.),

Constructing Identity in and around Organizations (pp. 63–88). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199640997.001.0001

Guo, W., Sengul, M., & Yu, T. (2020). The impact of executive verbal communication on the convergence of investors' opinions. *Academy of Management Journal*, amj.2019.0711. https://doi.org/10.5465/amj.2019.0711

Hand, D., Dithrich, H., Sunderji, S., & Nova, N. (2020). *2020 annual impact investor survey*.

Harji, K., & Jackson, E. T. (2012). *Accelerating impact: Achievements, challenges and what's next in building the impact investing industry*.

Hart. (2000). *Diction 5.0, the text-analysis program. User's manual*. Sage.

Hilpert, M. (2017). Frequencies in diachronic corpora and knowledge of language. In M. Hundt, S. Mollin, & S. E. Pfenninger (Eds.), *The changing english language* (pp. 49–68). Cambridge University Press. https://doi.org/10.1017/9781316091746.003

Howard, E. (2012). *Challenges and opportunities in social finance in the UK*.

Huang, A. (2008). Similarity measures for text document clustering. *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008)*, 49–56.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 216–225.

Jensen, M. (2010). Legitimizing illegitimacy: How creating market identity legitimizes illegitimate products. In G. Hsu, O. Kocak, & G. Negro (Eds.), *Categories in markets: Origins and evolution* (pp. 39–80). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2010)0000031004

Jivani, A. G. (2011). A comparative study of stemming algorithms. *International Journal of Computer Technology and Applications*, *2*(6), 1930–1938. 10.1.1.642.7100

Jurafsky, D., & Martin, J. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Prentice Hall.

Khor, S. Y. (2020). Language and identity. In *The International Encyclopedia of Linguistic Anthropology* (pp. 1–26). Wiley. https://doi.org/10.1002/9781118786093.iela0164

Lundberg, S. M., Erion, G. G., & Lee, S. I. (2019). *Consistent feature attribution for tree ensembles* (1802.03888). http://arxiv.org/abs/1802.03888

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st Conference on Neural Information Processing System*, 1–10.

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404–411.

Nicholls, A., & Daggers, J. (2016). *The landscape of social impact investment research: Trends and opportunities*.

Nielsen, J. (1997). *How users read on the web*. https://www.nngroup.com/articles/how-users-read-on-the-web/

Nielsen, J. (2008). *How little do users read?* Nielsen Norman Group. https://www.nngroup.com/articles/how-little-do-users-read/

Nielsen, J., & Loranger, H. (2006). Prioritizing web usability. In *Prioritizing web usability*. new riders.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Press.

Perkins, J. (2014). Python 3 text processing with NLTK 3 cookbook. In *Python 3 text processing with NLTK 3 cookbook*. Packt Publishing.

Pharoah, C., Scott, D., & Fisher, A. (2004). *Social enterprise in the balance: Challenges for the voluntary sector*. Charities Aid Foundation.

Reijden, P. V.D., & Koppius, O. (2010). The value of online product buzz in sales forecasting. *Proceedings of the International Conference on Information Systems ICIS*, 1–17.

So, I., & Staskevicius, A. (2015). *Measuring the 'impact' in impact investing*.

Sustainable investing joins the mainstream. (2017). *The Economist*. https://www.economist.com/finance-and-economics/2017/11/25/sustainable-investment-joins-the-mainstream

Tietze, S., Cohen, L., & Musson, G. (2003). Understanding organizations through language. In *Understanding Organizations Through Language*https://doi.org/10.4135/9781446219997

Tomlin, R. S., & Myachykov, A. (2015). Attention and salience. In E. Dabrowska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 31–52). De Gruyter Mouton. https://doi.org/10.1515/9783110292022-003

Wales, K. (2006). Stylistics. In *Encyclopedia of Language & Linguistics* (pp. 213–217). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00548-4

Winkler, H.-J., Rieger, V., & Engelen, A. (2019). Does the CMO's personality matter for web traffic? Evidence from technology-based new ventures. *Journal of the Academy of Marketing Science*, in press. https://doi.org/10.1007/s11747-019-00671-9

Zenker, O. (2018). Language and identity. In *The International Encyclopedia of Anthropology* (pp. 1–7). Wiley. https://doi.org/10.1002/9781118924396.wbiea2271

Zingales, L. (2018). Public companies should prioritise shareholder welfare, not value. *Financial Times*. https://www.ft.com/content/52240e18-e412-11e8-a8a0-99b2e340ffeb. Accessed 30 June 2022.

CRediT author statement **Laura Toschi**: Conceptualization; Writing - Original Draft; Writing - Review & Editing; Funding acquisition. **Elisa Ughetto**: Conceptualization; Writing - Original Draft; Writing - Review & Editing; Data Curation; Investigation; Funding acquisition. **Andrea Fronzetti Colladon**: Conceptualization; Writing - Original Draft; Writing - Review & Editing; Methodology; Software; Formal analysis; Investigation; Visualization.